Recht und Digitalisierung | Digitization and the Law 1

Bayer | Holznagel | Korpisaari | Woods (eds.)

Perspectives on Platform Regulation

Concepts and Models of Social Media Governance Across the Globe

Nomos

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Recht und Digitalisierung | Digitization and the Law

Herausgegeben von | Edited by

Prof. Dr. Roland Broemel
Prof. Dr. Jörn Lüdemann
Prof. Dr. Rupprecht Podszun
Prof. Dr. Heike Schweitzer, LL.M.

Band 1 | Volume 1

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - CC BY-SA

Judit Bayer | Bernd Holznagel
Päivi Korpisaari | Lorna Woods (eds.)

Perspectives on Platform Regulation

Concepts and Models of Social Media Governance Across the Globe

Assistant Editor Jan Kalbhenn

Nomos

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de

ISBN 978-3-8487-8557-5 (Print)
978-3-7489-2978-9 (ePDF)

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN 978-3-8487-8557-5 (Print)
978-3-7489-2978-9 (ePDF)

Library of Congress Cataloging-in-Publication Data
Bayer, Judit | Holznagel, Bernd | Korpisaari, Päivi | Woods, Lorna
Perspectives on Platform Regulation
Concepts and Models of Social Media Governance Across the Globe
Judit Bayer | Bernd Holznagel | Päivi Korpisaari | Lorna Woods (eds.)
601 pp.
Includes bibliographic references and index.

ISBN 978-3-8487-8557-5 (Print)
978-3-7489-2978-9 (ePDF)

1st Edition 2021

© The Authors

Published by
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Production of the printed version:
Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN 978-3-8487-8557-5 (Print)
ISBN 978-3-7489-2978-9 (ePDF)

DOI https://doi.org/10.5771/9783748929789

Onlineversion
Nomos eLibrary

This work is licensed under the Creative Commons Attribution – ShareAlike 4.0 International License.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Foreword:
Perspectives on platform regulation: models and limits

Monroe E. Price

These are startling times in the history of media and information regulation. Existing frameworks fray as disruption becomes the rule. Societies dispute the way to define freedom of expression and, in fear of disappearing stability, emphasize the establishment of order. Authoritarian tendencies capture what were often invented as technologies of freedom. In this environment, governments, the tech companies, and civil society all are in search of redesigning and thereby guiding basic organizing principles. This book excavates, develops, examines and tests a basic concept – the platform as a central mode for classifying thought about this century's experiments in regulating speech and information flows.

The very idea of "the platform" is intriguing. Platforms are a metaphor, and a powerful one. The image can be of a performer-athlete ready to make a perfect dive. Platforms can be sites for exclusive opportunities to demonstrate and frequently, platforms can be defined through issues of access. Platforms can be seized, hijacked and controlled or they can be virtual common carriers. Often it appears as a locus that is neutral and necessary for commerce in the commodity for which the platform accommodates trade. "Platform" has become a weighted term, an opportunity for a wide variety of distinct approaches to regulation to be articulated, legislated and implemented.

The concept of "platform" is appealing because it creates a category distinction (or the illusion of such distinction), one between content production and distribution facilitator. Having and cultivating such a distinction opens the opportunity – so welcome – for creative regulatory choices. The distinction is necessary so as to allow zones of immunity from liability, said to be critical in the development of social media and the Internet. Distinguishing the platform from its users has had complex implications for regulation of ownership in successive iterations of media and society.

The editors of this volume have, in fact, themselves created a platform – a platform for competing designers of regulatory architecture in the field of information and media to describe their findings and arguments. The authors use debates about hate speech and its regulation as a broad case

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

study of the variety of models and the omnipresence of limits on finding a model that can operate in a variety of contexts. Providing a taxonomy of possible regulatory choices, surveying conceptual models, is an important contribution. The editors recognize the significance of observing models as they operate in context. The volume takes the quite difficult step of including descriptions of how various conceptual models fare in an array of geographically distinct environments.

Implicit in the work that characterizes these pages is the recognition of what might be called a "regulatory deficit." In my view a regulatory deficit exists where there is a well-founded societal desire for governmental response to a social need, as yet unsatisfied, coupled with an appropriate understanding of fundamental (including constitutional) limitations. The treatment of hate speech is a useful example, of an area of regulatory deficit as exemplified in this book. The problem of regulatory deficit exists with respect to many chronic areas of crisis: terrorism, harsh political polarization, disinformation and even the general issues of identity and society. In each case, an often desperate search for government response becomes an insistent demand for which a supply of near formulaic remedies is produced. Much of the discourse here identified with platform regulation deals with this problem of regulatory deficit. Of course, not all such demand is owed respect and authors in this book often take a dim view of asserted deficits. The challenge exists of refining the category to measure a demand for regulation that is consistent with international human rights norms and laws. But even this is problematic because it does not necessarily recognize that those long established norms and laws might themselves change and reflect newly deemed necessities for control. Even the immutable sometimes mutates.

In all of this, in the intense culture of debate, collaboration, and experimentation, new patterns of global engagement in the construction of changing regulatory paradigms are striking. Relevant is the relatively plastic, yet liberating idea of the epistemic community: "a network of professionals with recognised expertise and competence in a particular domain and an authoritative claim to policy relevant knowledge within that domain or issue-area."[1] Over time, the potential of such a community has grown as a concept. What one might search for and cherish in epistemic communities is a psycho-social surplus, a quality beyond scholars demonstrating a common view of a way of organizing knowledge. An

1 Peter M. Haas, *Introduction: epistemic communities and international policy coordination* (*International Organization*, 1992), 1.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

epistemic community becomes one that has developed shared views and, among contests for primacy, advances them to realize further a common goal or improve operation of an institution. These characteristics can be seen among scholars working together to improve the understanding of hate speech and the role of platforms. Peter Haas identified typical features of such communities: a *shared set of normative and principled beliefs*; *shared causal beliefs between policy actions and desired outcomes*; *shared criteria for validating knowledge*; and a *common enterprise, presumable out of the conviction that human welfare will be enhanced as a consequence*.[2]

Epistemic communities celebrate the coming together of scholars across disciplines. The volume is the product of the Institute for Telecommunications and Media Law at the University of Muenster cooperating with scholars at the University of Essex and the University of Helsinki. The processes by which the volume was produced demonstrate what is required for a modern epistemic community and the essays in this book exemplify how emerging institutions benefit from the attendant interchange. The (Facebook) Oversight Board grows and changes, often, in response to the active sphere of experts engaged in blogging, writing, zooming, in short bringing insights, viewpoints and expertise to a significant and jurisprudentially challenging project. All this cross-border discussion takes place in a world still defining state sovereignty in an environment where technologies disrupt and industries transcend borders. It is an era of change, radical system-wide change. And it is an era where effort is needed to retain basic values of free expression in the face of geopolitical, technological, and economic transformations. It is a time of extraordinary anxiety about the project of regulation. And therefore it is a time where studies like those provided here are so important.

2 Haas, *Introduction*, 3.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Content

Introduction 13
Judit Bayer, Lorna Woods, Bernd Holznagel

Models of Platform Regulation

Rights and Duties of Online Platforms 25
Judit Bayer

European Legislative Initiative for Very Large Communication Platforms 47
Jan Christopher Kalbhenn

Introducing the Systems Approach and the Statutory Duty of Care 77
Lorna Woods

Policy Developments in the USA to Address Platform Information Disorders 99
Sarah Hartmann

Interoperability of Messenger Services. Possibilities for a Consumer-Friendly Approach 119
Jörg Becker, Bernd Holznagel, Kilian Müller

Six Problems with Facebook's Oversight Board. Not enough contract law, too much human rights. 145
Mårten Schultz

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Screenshots: A Glance beyond the Transatlantic

„Open with Caution". How Taiwan Approaches Platform Governance in the Global Market and Geopolitics 167
Kuo-Wei Wu, Shun-Ling Chen, Poren Chiang

Digital Platform Regulation in Japan – does the soft approach work? 187
Izumi Aizu

Social Media Platform Regulation in India – A Special Reference to The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 215
Siwal Ashwini

Thoughts on the Regulation of Content on Social Media in Latin America: Authors' Rights, Limitations and Content Filtering 233
Maria L. Vazquez, Maria Carolina Herrera Rubio, Alejandro Aréchiga Morales

Topic-based Regulation: Media Law and Data Protection

Media Law Regulation of Social Networks - Country Report: Germany 263
Bernd Holznagel, Jan Christopher Kalbhenn

Reshaping Canada's Broadcasting Act: Solutions in Search of a Problem? 291
Michael Geist

The UK's Approach to Regulation of Digital Platforms 329
Lorna Woods

Social Media Users Data Access: Russian Legal Approach 351
Juliya Kharitonova, Larissa Sannikova

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Hate Speech on Platforms

Protecting Democratic Expression Online: Canada's Work in Progress 367
Richard Janda

Lessons learned from the first years with the NetzDG 415
Maximilian Hemmert-Halswick

Platform Governance at the Periphery: Moderation, Shutdowns and Intervention 433
Giovanni De Gregorio, Nicole Stremlau

Protecting the Freedom of Expression in an Era of "Platformization:" Paving a Road to Censorship? 451
Jacob Mchangama, Natalie Alkiviadou

Online Shaming - a New Challenge for Criminal Justice 473
Kristiina Koivukari, Päivi Korpisaari

The Role of Occupational Safety and Health Legislation in Hate Speech Regulation. Employers' responsibility to prevent and respond to the risk of hate speech at work – the Finnish perspective 489
Enni Ala-Mikkula

Combating Disinformation

Platform (un)accountability. Reviewing Platform Responses to the Global Disinfodemic One Year Onward 509
Trisha Meyer, Alexandre Alaphilippe

Disinformation in the Perspective of Media Pluralism in Europe – the role of platforms 531
Elda Brogi, Konrad Bleyer-Simon

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Regulation of Online Disinformation in Singapore 549
Peng Hwa Ang, Gerard Goggin

Conclusions: Regulatory Responses to Communication Platforms: Models and Limits 565
Judit Bayer, Bernd Holznagel, Päivi Korpisaari, Lorna Woods

The Authors and Editors 585

Index 595

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Introduction

Judit Bayer, Lorna Woods, Bernd Holznagel

1989 – 2021

Online communication has developed tremendously over past decades. In 1989, two innovations created the World Wide Web: HTML, a hypertext markup language, and HTTP, hypertext transfer protocol. These tools, and the related user-friendly browsers provided easy access to the internet for the general public. A rapidly growing offer of websites and services also gave floor to the first legal disputes in a number of jurisdictions, which raised the question: can intermediaries be liable for criminal content, or content that is contrary to private or administrative law?

Notably, many of those landmark cases were related to early forms of social media, such as Usenet (Godfrey v Demon[1]) and bulletin boards (Stratton Oakmont v Prodigy[2]), or otherwise questioning whether the website host takes responsibility for commissioned publications (Blumenthal v Drudge[3]).

The first legal rule applying to internet content was the Communications Decency Act of 1996, Section 230 of which is still subject to discussion. The aim of the act was to regulate indecency on the Internet. While those parts of that Act were struck down by the US Supreme Court in a landmark ruling, Section 230 – which provides for intermediary immunity in relation to content hosted - remained.[4]

The US legislation was relatively active in the period between 1996 and 2000, passing several laws for the protection of children, giving rise to repeated constitutional rulings which annulled the whole or part of some of these for violating the First Amendment. The Digital Millenium Copyright Act introduced the notice-and-takedown regime as a method to deal with copyright infringement. Similarly, the European Union passed

1 Godfrey v. Demon Internet Service (2001) QB 201.
2 Stratton Oakmont, Inc. v. Prodigy Services Co., 23 Media L. Rep. 1794 (N.Y. Sup. Ct. 1995).
3 Blumenthal v. Drudge, 992 F. Supp 44 (D.D.C. 1998).
4 Reno v. American Civil Liberties Union, 521 U.S. 844 (1997).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the E-Commerce Directive in 2000 according to which intermediaries enjoy immunity from suit provided they either did not know about the content complained of or acted expeditiously once on notice. While there are differences between the approach in the EU and the US (and again in other jurisdictions), with the assumption of some form of immunity, it looked like the responsibility of online service providers had been settled in a satisfactory way, giving room for development, but also providing for removal of content where the law provided so.

The mentioned laws are still in effect, even though the development of technology has long overhauled the structure of the 1990's for which they were tailored. The first social network sites were already there from 1996 on, and gained popularity as broadband connection penetrated households after the millenium: Six Degrees, 1996, Wikipedia in 2001, Friendster in 2002, MySpace, LinkedIn, Hi5 in 2003 and Facebook, 2004. During these years, the first attempts with mobile internet were also traceable, but they spread relatively slowly, due to the unattractive user interfaces in the first internet-enabled mobile phones. Meanwhile, the 3G network which enabled faster mobile internet connection, got launched in 2001 in Japan, 2002 in the US and 2003 in EU. The breakthrough happened in Japan around 2004, when software, user interface and other consumer-friendly features were combined to enable rapid access to the open internet. Mobile internet rapidly spread on the heavily regulated Japanese market, which was, however, isolated from the global trend. The international debut of mobile internet as we know it today, came when Apple's iPhone was released in 2007 (on June 29 in the US and on November 9 in the EU). The real „smartphone revolution" was enabled by other producers that produced cheaper hardware and software.[5] The penetration of smartphones and mobile internet opened a new era of how people used the internet.

These landmark steps from several areas were needed to get from the early internet to today's smart-phone dominated platform-based communication culture. Parallel innovations contributed to the accelerated development that occurred in telecommunication technology, hardware and software technology, and online services. Broadband enabled the use of images and sound. Platforms made publishing content a convenience to any lay person even without literacy. And mobile internet put the whole world into the pocket of every teenager – and made online presence a

5 Bloomberg. "The Smartphone Revolution Was the Android Revolution". Aug 6, 2019. https://www.bloomberg.com/graphics/2019-android-global-smartphone-growth/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

uniquely personal, even intimate experience, a place where the social, personal and business life of the individual are blended.

This change occurred in little more than a decade, and legal regulation did not follow through. The E-Commerce Directive's logic reflected the pre-platform age, where providing access, hosting and content could be clearly separated. The new service package provided by platforms did not fit. While the Directive seemed to provide immunity for third party content, provided that it was removed upon notice, courts did not apply this rule on platform intermediaries like eBay or a newspaper's comment section.[6]

Platforms grew and proliferated, to become dominant actors which connect and aggregate supply and demand in all areas of economy and society, from sale and tourism, to dating sites. The mediating role that they do is comparable to a traditional agency, but incomparable in the volume and speed with which the third party information is aggregated, categorised, and ordered to generate a personal offer for the other party. Platforms got access to all-inclusive information about their users: not only their social network, or shopping habits, but their business decisions, fear-generated searches, their whereabouts and many more became accessible information for personalised advertising and content offer.

This mind-boggling system operates on a legislative framework that has responded to the needs of the word-wide-web, the pre-broadband and pre-smartphone age.

In 2016, the potential of social media as an instrument has been demonstrated globally, and it became widely accepted that social media is able to make a global impact on real-life social processes, like elections. As it was later revealed, the US election campaign was infiltrated by disinformation actions and intentional manipulation.[7] The same was exposed regarding the political campaign preceding the Brexit referendum.[8] Both democratic decision-making events were regarded as a rupture to the „genuine" democratic processes and have been heavily investigated. Large research and

---

6 Judgment of the CJEU Loréal v. eBay, C324/09, EctHR judgment Delfi v. Estonia, App.No. 64569/09, June 16, 2015.

7 116th Congress Senate Report of the Select Committee on Intelligence US Senate on Russian Active Measures Campaigns and Interference in the 2016 US Election. Volume 2. Russia's Use of Social Media With Additional Views. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

8 House of Commons, Digital, Culture, Media and Sport Committee. Disinformation and 'fake news': Final Report. 18 February 2019. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

policy efforts have been made to reveal who was responsible, and how to find an appropriate solution to online harms. Facebook's responsibility was also raised by the United Nation for enabling incitement to hatred, in regard of the regrettable Rohingya genocide by Myanmar.[9] The COVID-19 related surge of mis- and disinformation gave yet another impetus to the research and policy initiatives of social media responsibility.[10]

What happened on 6 January 2021 may be regarded as another landmark event. The leaving incumbent US President used his social media channel to express his sympathy towards a violent movement attacking the Capitol. After years of exceptional treatment, his account was suspended at Twitter, Facebook and Instagram for violation of the Community Standards. The event demonstrated that social media communication can contribute to accelerate violence, and gave new impetus to the debate on the boundaries of online free speech, as well as the role of social media platforms.

In view of these impacts of social media on society, no surprise that in the past years, instruments to counteract these possible undesirable effects have been considered around the globe. Hardly a week goes by without reports about the introduction of new measures whether it addresses an ancillary copyright (Australia), anti-trust measures including unbundling (USA) or effective measures against disinformation and hate speech (Canada). Against this background, researchers obviously take up the development and sense (global) trends of legal development in this area.

In December 2020, the European Commission issued two draft laws: the Digital Market Act, and the Digital Services Act. These aim to provide a basic legal framework for platform economy and the platform communications environment. Prior to this, the European Commission has fought hate speech and disinformation with various soft instruments, in particular with induced self-regulation, where the European Commission set the goals, convened the industry actors and let them draw up and sign their Code of Practice against Disinformation. The self-assessments

9 UN Human Rights Council Report of the independent international fact-finding mission on Myanmar. A/HRC/39/64. https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf. „Facebook has been a useful instrument for those seeking to spread hate."

10 Wunderman Thompson, The University of Melbourne and Pollfish, World Health Organization (WHO). "Social Media & COVID-19: A Global Study of Digital Crisis Interaction among Gen Z and Millennials. Key Insights." https://covid19-infodemic.com/assets/download/Social_Media_COVID19_Key_Insights_Document.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of the Code's implementation have been published by the Commission and evaluated by the European Regulators' Group for Audiovisual Media (ERGA).[11] Even though self-regulation proved less effective than hoped, there seems to be no room for strict legislative intervention because of the complexity of these areas. The goal is to design a stricter cooperation between the industry and the Commission as well as national authorities, amounting to co-regulation. This would include that the Commission will facilitate the drafting of the Code, and regularly monitor and evaluate the achievements and its objectives (read more on this in 1.2. by Jan Kalbhenn).

This volume collects a variety of perspectives, representing a geographical diversity, and drawing inspiration from various sectoral approaches. The editors believe that such a discussion can provide an advantage in the drafting process, which may prove to be a long road.

The structure of this book

The idea of this book developed gradually. The first idea emerged in a café in Münster, whose name preserves the memory of the Westphalian Peace Treaty (1648). The idea has further developed and expanded as the second and third wave of the global pandemic limited all contact to online conferences. This ironically allowed us to widen the planned scope of the workshop series, and integrate researchers from other continents as well. Papers which report about the specific perspectives of the regulatory needs in Japan, Taiwan, Russia and the African continent provide an invaluable insight to understand global processes. The first chapter of the volume includes papers which discuss the regulation of online platforms from wide, systemic perspectives. The first paper attempts to shed light on how the extent of platforms' freedom and competence in defining their own rules and deciding about content moderation is perceived, through court decisions and legal instruments. It argues that it would be of primary importance to define platforms' role and responsibility in the communication chain, realising their unique role in aggregating and ranking content. The following papers discuss and analyse the legislative initiatives from three large jurisdictions. Jan Kalbhenn's writing analyses the European

11 ERGA Report on disinformation: Assessment of the implementation of the Code of Practice. https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Union's draft Digital Services Act, with special regard to its rules regarding very large online platforms. Lorna Woods describes the systems approach and the idea of „Statutory Duty of Care", and Sarah Hartmann introduces the debates and policy developments in the US around a reform of the existing legislative framework. The final two papers address innovative approaches to social communication: the writing of Jörg Becker, Bernd Holznagel and Kilian Müller discusses the interoperability of messenger services. This might be a step as decisive as the milestones listed in the first part of this introduction were. The final paper of the first chapter by Mårten Schultz critically explores the Facebook Oversight Board.

The second chapter departs from the usual transatlantic perspective, by including four reports on platform regulation from Taiwan, Japan, India and Latin-America. Taiwan keeps its eyes on the transatlantic legal development and is a favoured hub for the online industry. China's proximity adds a special flavour to its democratic regulatory intentions. The paper by Kuo Wei Wu, Shun-ling Chen and Poren Chiang provides the reader an understanding of this complexity. The Japanese regulatory approach takes the multi-stakeholder view, relying on self- and co-regulation. An overview along with a historical context is provided by Izumi Aizu. India has passed a new regulatory regime in 2021, addressing ethical guidelines for intermediaries and the digital media. This, in the context of freedom of expression is introduced by Siwal Ashwini. The chapter is closed with samples of platform regulation from Latin-American states, with a special focus on copyright by Maria L. Vazquez, with co-authors Maria Carolina Herrera Rubio and Alejandro Aréchiga Morales.

The third chapter examines theme-based regulation of certain aspects of online platform communication. The first and the second paper both explore the media law approach. Bernd Holznagel and Jan Kalbhenn introduce and analyse the amended German Media State Treaty which – as a first in the globe – provided for pluralism measures also for social media platforms. This media regulation takes a comprehensive view to sustain a diverse media sphere, with a special place in it for public broadcasters. Canada, at the time of writing this book, was discussing a new broadcasting legislation, which addressed the streaming services, among others. Michael Geist writes about the bill and the relating controversies in the legal discourse, in particular regarding issues with competition and freedom of expression. The UK, beyond a developing systemic regulation of platforms that has been discussed in chapter 1, also addresses new media with a variety of sectoral laws. These legal concepts, such as data protection, with its implications in advertising law and child protection; competition and consumer protection are elaborated by Lorna Woods.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The topical discourse in Russia is concerned about finding a balance between the protection of personal data and allowing commercial use of big data, as written by Juliya Kharitonova and Larissa Sannikova.

Hate speech and disinformation have been the major triggers for policymakers' reaction in the past decade. In comparison to previous concerns like pornography, copyright and terrorism, this was more difficult to compartmentalise. Hate speech and disinformation have infiltrated the political discourse and impacted social harmony. The basic structure of societies' and of democratic operation are now at stake. Hate speech and disinformation share the feature that they are at the verge of legality. They are often context-dependent and cannot easily be judged. Some states are more tolerant in dealing with these than others. According to the case law of the European Court of Human Rights, falsity alone is not a sufficient reason to restrict freedom of expression, unless there is a legitimate aim, such as the reputation of others. Restriction of commercial content was found more acceptable by the Court, however, the Court also held that it would be unreasonable to restrict freedom of expression only to generally accepted ideas in a sphere in which uncertainty reigns, which is also the case in relation to the COVID-19 infodemics. The regulation of hate speech shows perhaps the largest divergence around jurisdictions among other types of content. In the past five years, both phenomena entered loudly the highest political circles. This prominence enables a more intense impact and reduces the chances for successful regulation. Chapter 4 addresses hate speech, and Chapter 5 disinformation in various states.

Canadian regulation is discussed in the first chapter: in the fourth, its hate speech legislative process is introduced. It is proving harder than anticipated to strike the balance between freedom of expression and the protection of minorities. Richard Janda's article introduces the existing legal framework, the various policy options and recommends ways to depart from a platform business model that serves to amplify extreme content.

Germany has pioneered the fight of illegal hate speech with its Network Enforcement Act. Despite initial criticism, the law is operative and has been amended twice to extend user rights and enable a tighter regulatory control. Maximilian Hemmert-Halswick provides a thorough description and analysis of the law's operation, relating controversies and amendments.

In the global south, hate speech and its suppression both can cause troubling consequences. Giovanni di Gregorio and Nicole Stremlau discuss with a fresh look how internet shut-downs are employed for censorship,

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

and take on the perspective of international law and the humanitarian doctrine to frame information interventions.

International human rights law is explored also by Jacob Mchangama in his essay on over-censorship under the pretext to fight hate speech, with particular focus on South-Africa. With a big geographical leap, we get to Finnish online hate speech. Discriminative online harassment is becoming a social problem that chills the freedom of expression of its victims. A close scrutiny by Päivi Korpisaari and Kristiina Koivukari of the possibilities of further criminalisation concludes that the principles of freedom of expression and of criminal legal guarantees do not leave room for further restriction. Enni Ala-Mikkula examines whether the Finnish labour rules provide guidance to employers to protect their employees from online hate speech.

The fifth chapter discusses the measures in the fight against online disinformation. Trisha Meyer and Alexandre Alaphilippe provide an invaluable account and overview of the self-regulatory responses applied by platforms as a response to the global infodemic. Elda Brogi and Konrad Bleyer-Simon examine disinformation in the light of media pluralism. They introduce the results of the Media Pluralism Monitor in this area, describe the European Digital Media Observatory's activity in relation to disinformation, and discuss European policy solutions. As the last episode in the volume, Ang Peng Hwa and his co-author Gerard Goggin present the counter-disinformation regulation of Singapore and its application.

Acknowledgements

This book would not have come into life without the support of the Karina and Erich Schumann Foundation which supported a research project and the stay of Judit Bayer in Münster. We are thankful for the University of Münster, the University of Essex, and the University of Helsinki, for also having supported this project. The papers have been presented in the workshop series 'Hate speech and Platform Regulation' organised in cooperation of these universities, and hosted by the Institute for Telecommunication and Media Law (ITM) at the University of Münster. Some of the presentations are accessible in podcast or video format from the website of ITM. We also owe credit to those participants of our workshop who did not find time to contribute to this book, like Olga Batura, Meg Chang, Elena Dodonova, Nikolaus Forgó, Ellen P. Goodwill, Faith Gordon, Elfa Yr Gylfadóttir, Irini Katsirea, Matthias C. Kettemann, and Marianna Muravyeva, Sofia Ranchordas and Krisztina Rozgonyi. We thank the Carnegie

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Foundation for enabling that the book can be published openly accessible. We are grateful to Jan Kalbhenn, managing director and senior research fellow of ITM who was a key contributor and organiser of the project. We also thank the junior research fellows at ITM, in particular Derman Aktas-Paszkiet, and Florian Flamme, and the student team of ITM, especially Fee Hinkel for providing constant assistance with the workshop's website, and Anna Laura Askanazy, Felizitas Heet, Benedikt Freese, Olivia Sun, and Hannah Waegner, who contributed to finalising the manuscripts. Many special thanks also to Karmen Stürznickel and Christian Schepers, who keep the Institute in Münster running and are thus cornerstones of the work on the book. We are also thankful for Christopher Goddard for proofreading a part of the manuscripts.

Bibliography

Ovide, Shira. "The Smartphone Revolution Was the Android Revolution." Bloomberg, August 6, 2019. https://www.bloomberg.com/graphics/2019-android-global-smartphone-growth/.

US Senate. "116th Congress Senate Report of the Select Committee on Intelligence US Senate on Russian Active Measures Campaigns and Interference in the 2016 US Election." Volume 2. Russia's Use of Social Media With Additional Views. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

House of Commons, Digital, Culture, Media and Sport Committee. „Disinformation and 'fake news': Final Report." 18 February 2019. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.

UN Human Rights Council. „Report of the independent international fact-finding mission on Myanmar." A/HRC/39/64. https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf. „Facebook has been a useful instrument for those seeking to spread hate."

Wunderman Thompson, The University of Melbourne and Pollfish, World Health Organization (WHO). "Social Media & COVID-19: A Global Study of Digital Crisis Interaction among Gen Z and Millennials. Key Insights." https://covid19-infodemic.com/assets/download/Social_Media_COVID19_Key_Insights_Document.pdf.

ERGA. "Report on disinformation: Assessment of the implementation of the Code of Practice." https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Models of Platform Regulation

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Rights and Duties of Online Platforms

Judit Bayer

Abstract: One of the two extreme ends of regulatory approaches to online platforms treats platforms as independent governors of speech, the other treats platforms as mere conveyors of third-party content. This paper highlights regulatory provisions and court cases that represent one or the other extreme. However, it ultimately found that the approaches are mixed and some instruments, like the draft Digital Services Act, combine both approaches consciously. While the different approaches may not be reconcilable in all cases, umbrella approaches, such as competition law and international human rights law, may set a higher-level framework to bring more consistency.

Keywords: online platforms, human rights, content governance, moderation, Digital Services Act, horizontal effect of human rights.

Chapter 1. Introduction

In the recent decade, social media platforms have gained influence over the public discourse across the globe. Their operation impacts various human rights, primarily freedom of expression and the right to information, but also others like privacy, dignity, the right to free elections, and potentially more.

These private actors do more than just transmit content; with their moderating, ranking, prioritising, and targeting actions, they govern and tailor the public discourse.[1] This activity is built into their design, and they could not operate without performing some form of selection and ranking. In addition to the strictly necessary moderation, further 'optimising'

1 Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media (New Haven: Yale University Press, 2018).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

is carried out to maximise advertising revenues, increase user engagement,[2] and maintain a civilised communicative environment. Thus, they govern content through their infrastructural design on the one hand and their moderation choices on the other.

They do so without being bound by human rights safeguards or accountable for their tailoring actions.[3] The largest social media platforms make considerable efforts to increase their transparency, cooperate with policymakers, and publicly impress that their content moderation choices are governed by moral values. However, when it comes to conflicting human rights, deciding whether content is legal or not becomes more complex. Often, this question can be answered relatively easily (copyright, terrorism, child abuse), although there are borderline cases and controversies even in these fields. One of the most cited examples of social media censorship concerned the photograph that became known as the 'Napalm girl', showing desperate people running from obvious traces of a (Napalm) bomb attack, among them a naked female child. The removal of this picture attracted considerable public outcry and closer scrutiny of the moderation principles.[4] Other types of illegal content cannot be interpreted without knowing the context, such as violation of reputation or certain forms of hate speech, and are more difficult to judge.

There is a variety of approaches to liability for third-party content around the globe, depending partly on the subject matter of the content or on the legal branch, but all provide a certain level of immunity. The American approach provides platforms with immunity for third-party content without conditions[5] (except if the subject matter is copyrighted

2 Hannah Schwär and Qayyah Moynihan, „Instagram and Facebook are intentionally conditioning you to treat your phone like a drug", Business Insider, 5 April 2020, https://www.businessinsider.com/facebook-has-been-deliberately-designed-to-mimic-addictive-painkillers-2018-12.

3 Rikke Frank Jørgensen and Lumi Zuleta, "Private governance of freedom of expression on social media platforms: EU content regulation through the lens of human rights standards," Nordicom Review 41 no. 1 (2020): 51-67, https://doi.org/10.2478/nor-2020-0003.

4 Tarleton Gillespie, Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media (New Haven: Yale University Press, 2018), 7. See also: Kate Klonick, "The Most Important Lesson from the Leaked Facebook Content Moderation Documents," Slate.com, June 29, 2017, https://slate.com/technology/2017/06/the-most-important-lesson-to-learn-about-facebook-content-moderation.html.

5 Communications Decency Act (CDA) 1996, § 230.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

content when the notice-and-takedown regime applies).[6] The European E-Commerce Directive[7] provides conditional exemptions from liability. The Digital Services Act (DSA)[8] has followed this approach, requiring the removal of illegal content. It has also developed procedural safeguards partly following the example of the German Network Enforcement Act (NetzDG).[9]

However, the real question, and the focus of this article, is the extent of platforms' freedoms regarding lawful content. What do they really do and is that activity subject to any legal regulation? The draft Digital Service Act defines 'online platforms' as hosting providers which also disseminate content (Article 2.h) DSA). The word 'disseminate', however, does not accurately reflect the content organising activity that platforms do; they rank, prioritise, deprioritise, and label content. Ironically, this organising activity is the main service platforms provide, beyond mere hosting of content, and it is precisely that which makes them so unique. Unfortunately, this activity is currently not transparent and there is no accountability for platforms.[10] The draft Digital Services Act does not seem to change this. It merely provides for compulsory self-regulation in the field of lawful but harmful content and other risks. Similarly, the Audiovisual Media Services Directive has provided that video-sharing platforms should adopt and apply pro-active self- and co-regulatory schemes to tackle harmful content (Article 28b AVMS Directive).[11]

Deprioritising or labelling and other forms of moderation are based on platforms' community guidelines. While these softer methods interfere less with the individual human right to free expression, they equally interfere with the public discourse. There is "a right to speech, but no right to reach", meaning the freedom is no guarantee that content reaches a high number of users. This catchy phrase disguises a critical aspect of social media platforms' power. First, if a dominant market player chooses to

6 Digital Millenium Copyright Act (DMCA) 1998.

7 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

8 Digital Services Act amending Directive 2000/31/EC COM/2020/825 final.

9 Netzwerkdurchsetzungsgesetz - NetzDG (2017), https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html.

10 (Commercial platforms like eBay etc., provide more transparent ranking criteria to their users than social media platforms.)

11 Audiovisual Media Services Directive, Directive 2010/13/EU.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

deprioritise an item of content, it effectively suppresses it.[12] With this, the platform will have interfered with the right to freedom of expression of the individual speaker (whether this is relevant in the light of the horizontal effect of human rights will be discussed in Chapter 3.b). Second, when such deprioritising is done on a large scale and/or over a long period of time, its accumulative effect has a potential to damage public discourse which impacts societies' democratic processes.

Whether and how the community standards and algorithmic moderation of giant social media platforms influence the public discourse – for example, by pushing some items onto the agenda and suppressing others – is not subject to supervision or accountability. The draft Digital Services Act envisages a co-regulatory scheme to provide for, at a minimum, consultation in setting the goals (Article 35 DSA). Whether the declared goals are fulfilled would be the subject of transparency requirements, but without legal consequences.

Chapter 2. The regulatory frames of platforms' powers

To what extent should platforms independently decide on content standards, including what should remain and receive attention online and what should be suppressed or removed? Should it be a platform's privilege to define content standards and the agenda, and govern the public discourse, similarly to traditional media companies? We are witnessing this happening; it has organically developed this way. The comparison with traditional media companies is tempting but inaccurate in several aspects. First, social media platforms do not publish their own content, and their users are not paid journalists representing the media companies' agenda. Still, with the help of algorithms, companies can prioritise those views they would like to promote. Second, the largest online platform companies reach and engage massively more people than traditional newspapers or broadcasters.[13] The largest newspaper company in the United States (US), based on circulation, reached just over 8.59 million persons

12 Molly K. Land. "Toward an international law of the internet", Harvard International Law Journal 54, no. 2 (2013): 393.

13 "Top 10 U.S. Newspapers by Circulation", Agility PR, last modified January 2021, https://www.agilitypr.com/resources/top-media-outlets/top-10-daily-american-newspapers/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

in 2020,[14] slightly less than the largest single newspaper in the world, Yomiuri Shimbun, with 9.1 million subscribers in the same year.[15] There is no aggregated data on the reach of international newspaper corporations, such as the Murdoch empire. In any case, it is hard to compete with Facebook's 190 million users in the US and 2.7 billion active users globally.[16]

As a consequence of a series of policy decisions, or more likely of their absence, social media lacks accountability. In contrast, traditional media, particularly broadcasting, is subject to significant restrictions regarding content, advertising, and in several countries, ownership. Current regulatory attempts in the United Kingdom (UK) and the European Union (EU) seek to find the middle road and acquire a certain level of supervision over content regulation decisions without making platforms accountable for individual content items. However, advertising and ownership regulation is not currently on the legislative agenda.

Online platforms might be further compared to cable or satellite companies (distributors) which are also subject to legal restrictions in selecting content to be transmitted, as well as their contracting conditions with the end-users. Differences again lie in the providers of content (media companies as responsible publishers in the case of distributors, and lay persons in the case of social media) and the volume of content. Moreover, platforms have a greater potential to govern the display of content than distributors.

This paper examines the relationship between social media platforms' freedom to govern content and the state's regulatory intervention into this freedom. From a comparative perspective, I set the hypothesis that two schools of thoughts (and policy approaches) exist, which represent the two ends of a spectrum:

a) Less freedom to platforms: they are supposed to convey content and only remove what they are obliged to by law, i.e., illegal content. They must respect procedural rights and – in an extreme interpretation of the limits – do not enjoy unlimited freedom in defining their Terms of Services, which must respect consumer protection principles, if not

14 "Leading newspaper companies in the United States in 2020, by total circulation", Statista, June 2020, https://www.statista.com/statistics/234685/leading-newspaper-companies-in-the-us-by-total-weekday-circulation/.

15 "Top Daily Newspapers in the World", Infoplease, last modified April 16 2020, https://www.infoplease.com/culture-entertainment/journalism-literature/top-ten-top-daily-newspapers-world.

16 "Facebook by the Numbers: Stats, Demographics & Fun Facts", Omnicore, last modified January 6 2021, https://www.omnicoreagency.com/facebook-statistics/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

fundamental rights. In other words, they might be obliged to *carry* certain content and *be prohibited from removing* it. This almost treats platforms as common carriers of content that is protected by the right to freedom of expression.

b) Wider freedom to platforms: they enjoy unconditional immunity for third party content and freedom to govern their premises, and can thereby practically regulate users' speech.

During my research, I found that these two categories are not entirely distinct. Further, some court decisions or policy instruments carry elements of both schools. Analysis of these might contribute to a crystallisation of platforms' rights and scope of competence in the formation of public discourse.

Ultimately, the investigation boils down to two simple questions. Who has the upper hand in forming the informational environment: platforms, users, or governments? And what needs to be done to create a balanced division of power, bearing in mind that the rights of one platform user often conflict with those of another user?

To shed light on the underlying legal concepts that may inform this debate, I will explore the developing discussion about the horizontal effect of human rights on private enterprises. There is agreement that states are obliged to ensure the enjoyment of human rights, but this agreement does not include private enterprises. However, an emerging debate can be observed among academic authors and international bodies in this respect, advocating for a more inclusive interpretation of the human rights obligations of private enterprises. This debate will be examined below.

The paper primarily focuses on the European Union with a comparative analysis of relevant case law and legislation, most notably from Germany and the United States. International and self-regulative norms are also drawn into the analysis.

Chapter 2.a. Stricter interpretation of platforms' roles and responsibilities

According to my hypothesis, a stricter interpretation of platforms' freedoms sees platforms' competences limited to the deletion of illegal content. This section of the paper will discuss a collection of laws and decisions representing this strict approach towards platforms' roles.

According to this approach, legislative instruments may limit platforms' freedom in defining which content to carry and which to remove or deprioritise. A typical manifestation of the strict policy approach towards

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

platforms' responsibility is the German Network Enforcement Act (NetzDG). This orders online platforms to remove, upon notification, content that violates the Criminal Code's listed hate speech prohibitions within a short deadline. Large online platforms are obliged to create a procedure for removal which respects users' procedural rights and are subject to transparency obligations, including reporting on their activities.[17] (More on this law can be seen in this volume by Hemmert-Halswick).

The other side of the coin is to oblige platforms to also *carry* certain content. For example, the German new media law provision in the German Media Treaty (MStV) prohibits platforms from discriminating against journalistic content.[18] Furthermore, the draft DSA provides for crisis protocols to be created by very large online platforms and facilitated by the European Commission (Article 37). These would include, among others, "displaying prominent information on the crisis situation provided by Member States' authorities or at Union level". Currently, there are other crisis communication measures within the European body of laws in the realm of cybersecurity incidents[19] and food safety.[20] However, even taken together, these measures fall short of a legal obligation for any provider to carry messages or to prioritise them.

Another element of the strict approach to regulation would be that platforms should carry *all* lawful content without discretion, as held by the Higher State Court (Oberlandesgericht, OLG) München and confirmed by the OLG Berlin.[21] The Court held that Facebook was not allowed to apply a stricter standard than the state; therefore, comments that were not illegal were not to be deleted. This was considered an obligation arising from Facebook's Terms of Service (TOS) as opposed to the Constitution. The TOS violated the principle of good faith when it stated that the platform may remove any content. Additionally, the fact that Facebook alone decided whether a post violated its guidelines was contrary to the

17 NetzDG (2017), https://www.gesetze-im-internet.de/netzdg/BJNR335210017.html.

18 German Media State Treaty (MStV), § 94.

19 Commission Recommendation (EU) 2017/1584, 22-23.

20 Commission Implementing Decision (EU) 2019/300 of 19 February 2019 establishing a general plan for crisis management in the field of the safety of food and feed, Annex I, (Title 2, paragraph 5) "Dissemination of key messages via social media and other tools (specific webpage for example) including, when necessary, the EFSA Communication Experts Network)".

21 OLG München, 24.08.2018 - 18 W 1294/18, NJW 2018, 3115; LG Berlin, 16.01.2018 - 16 O 341/15, GRUR-RR 2018, 372.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Civil Code, which provided for equal rights of the contracting parties.[22] Blocking the user account was interpreted as a unilateral termination or suspension of the contract, which is generally unlawful.[23] In a similar decision against Twitter, the OLG Dresden Court found that Twitter's TOS, which said that they might revise their TOS from time to time, was unlawful.[24] In the Court's view, this could mean that they can change any rule, even the free nature or provision of their services. Importantly, the German Civil Code includes clear limitations on the content of General Terms and Conditions,[25] among which unilateral amendment of the terms is invalid.[26]

Besides, the content in the Twitter case was not illegal; it was satirical. Therefore, even if it violated the TOS, it was covered by freedom of expression. The OLG Dresden Court later held that the 'indirect third party effect' or indirect horizontal effect of fundamental rights, an established principle in German constitutional law (see more on this below), should ensure that satirical expressions do not result in a deletion of the account. Although this horizontal effect does not directly oblige private entities to ensure fundamental rights in relation to other private entities, it should ensure a certain level of respect in civil law relationships, particularly regarding the general terms and the ambiguous legal terms of civil law.[27] With this argumentation, the OLG Dresden went further than the OLG München, which established its verdict on the Civil Code's provisions on equal rights of the parties and limitations of the General Terms and Conditions.

In another case, the Regional Court of Frankfurt held that the blocking and deletion of a statement is not justified if the statement is covered by freedom of expression.[28] The court referred to the indirect third-party effect of fundamental rights. In this case, Facebook had removed a political

22 BGB [German Civil Code] (87th edition, 2021), § 241 para. 2.
23 R. Schwartmann and R. L. Mühlenbeck, „NetzDG und das virtuelle Hausrecht sozialer Netzwerke" (2020) ZRP, 170.
24 LG Dresden, 12. 11. 2019 – 1a O 1056/19, MMR 2020, 247; OLG Dresden, 07.04.2020 - 4 U 2805/19, MMR 2020, 626.
25 BGB, § 305-310.
26 BGB, § 308, no. 4-5.
27 J. Merck, "OLG Dresden: Twitter darf Accounts nicht ohne ausreichenden Grund sperren", LHR, June 29 2020, https://www.lhr-law.de/magazin/social-media-recht/olg-dresden-twitter-darf-accounts-nicht-ohne-ausreichenden-grund-sperren/.
28 LG Frankfurt am Main, 14.05.2018 - 2-03 O 182/18, MMR 2018, 545.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

opinion that did not amount to hate speech and suspended the user's account for 30 days.[29]

In other judgments, the German courts found that Facebook's community guidelines adequately respected human rights principles.[30] Despite the positive findings in favour of the platform, this signals an anticipation that if platforms fail to adequately respect human rights, their decisions will be invalidated. Therefore, these cases are also relevant to the "strict" approach, albeit they represent a more relaxed expectation than that permitting the removal of illegal content only: if there is general respect for human rights, then even lawful content may be removable.

However, German jurisprudence regarding the human rights obligations of platforms is not consistent, as demonstrated by a 2021 case decided in Braunschweig at first and second instances.[31] The court of first instance declared that as an operator of a social network with considerable market power, Facebook owed an enhanced duty to respect fundamental rights. It held that the basic legal content of the fundamental rights should also prevail in private law, particularly the general clauses and other terms that need to be interpreted in light of the fundamental rights. Therefore, the terms of the contract should be interpreted in an opinion-friendly manner. At the same time, it also recognised Facebook's fundamental rights to pursue business and to property (Articles 12 and 14 of the German Basic Law) and held that Facebook was not obliged to publish all expressions of opinion without discretion, even if they were protected by freedom of expression. However, the content in question in the said case did not amount to hate speech, and the removal was therefore unjustified. Yet the appeal court disagreed; it denied that Facebook has a heightened obligation to respect fundamental rights or that its guidelines would need to be interpreted in an opinion-friendly manner. Moreover, it held that even state authorities are not required to provide a means for expressing and disseminating opinions. Certainly, there is no such obligation for pri-

---

29 The translation of the removed opinion is: "The pseudo-left T is a warmonger first class! Wasn't it this hate speech that recently whistled that you were about to go bankrupt? NO LOSS! is my opinion!"

30 OLG Karlsruhe, 25.06.2018 - 15 W 86/18, NJW 2018, 3110; LG Heidelberg, 28.8.2018 – 1 O 71/18, MMR 2018, 773.

31 OLG Braunschweig, 05.02.2021 - 1 U 9/20, decision of second instance court, preceded by the first instance decision of LG Braunschweig, 11.12.2019 - 9 O 4199/18. The statement in question was: "Den Schrott versenken, das ist ein illegales Schlepperschiff!" translated as "Sink the scrap, this is an illegal tugboat!" in response to the news headline: "Private rescue ship "Aquarius" returns to the Mediterranean off Libya."

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

vate companies. The Appeal Court statement that Facebook does not even have a dominant position in the dissemination of opinions demonstrates the level of controversy. The Appeal Court explained that the basic rights are not directly applicable between private parties but only have indirect third-party effect in private law. Finally, it found that the incriminating expression amounted to hate speech, and the removal was justified.

This leads us to the second chapter which examines the more relaxed approach towards platforms' responsibility, allowing them more freedom to decide.

Chapter 2.b. Wider freedom to platforms

From this angle, state interference is undesirable and private governance more trustworthy. Social media platforms are regarded as legitimate governors of their premises and users' expressions. The clearest manifestation of this is Section 230 of the Communications Decency Act of the United States or, more specifically, its "Good Samaritan" provision. The rule provides immunity to any actor for the speech of third persons, even if they moderate the content for reasons of decency.[32] Subsection (c) (2) explicitly says "whether or not such material is constitutionally protected", by which it presumes that constitutionally protected material may also be removed or restricted. Platforms are free to carry illegal content without risk of being liable (until a court order or a specific act[33] obliges them to remove it), and they are free to remove lawful content, similarly. This freedom is even more robust in light of the state action doctrine[34] according to which private institutions do not have constitutional obligations, only the

32 CDA § 230. (c) (1) "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." (c)(2) "any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected."

33 The Digital Millenium Copyright Act provides for the takedown of copyrighted content upon notice.

34 Stephan Jaggi, "State Action Doctrine", Oxford Constitutional Law, last modified October 2017, https://oxcon.ouplaw.com/view/10.1093/law-mpeccol/law-mpeccol-e473; see also: "State Action Requirement", LLI, https://www.law.cornell.edu/wex/state_action_requirement.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

state does.[35] Some US policy experts question this convenience in the hope of gaining more control over platforms.[36] The debate encompasses the two competing views discussed in this paper. One argument is that with freedom should come responsibility,[37] however, control would furnish the government with power over speech, which is another cause for concern and contrary to American First Amendment tradition.

Under this more liberal approach, it is clear that platforms have the freedom to decide about content removal, content prioritising, deprioritising, and labelling according to their own standards (whether transparently or not is another question). However, it is still unknown whether this competence would also include curating content or generating their own content. 'Curated' content presents walled gardens meant to provide controlled, trustworthy information to the public. This was used by Twitter, Facebook, Mozilla and TikTok in the fight against the COVID-19 infodemic to present authentic scientific information to the public. This curated content – which has features of a digest or a magazine – represents a service different from the usual activity of ranking and prioritising. Selecting and presenting the content in one bundle includes editorial decisions. As a response to the pandemic, these can be regarded as extraordinary, crisis-related content offers.[38] The question is, does this practice have a place

35 Amelie Heldt, "The President and Free Speech: Consequences of Twitter's Fact-Checking Indication", Internet Policy Review, June 4, 2020, https://policyreview.info/articles/news/president-and-free-speech-consequences-twitters-fact-checking-indication/1483.

36 Ilya Banares, Rebecca Kern and Naomi Nix, "Facebook, Twitter, Google CEOs Split Over Social Media's Shield", Bloomberg, March 24 2021, https://www.bloomberg.com/news/articles/2021-03-24/zuckerberg-supports-section-230-reform-ahead-of-house-hearing. Among others, the conservative Chairman of the Federal Communications Commission, Ajit Pai – the same person responsible for erasing the rule on network neutrality in the US – supports the plan to limit Section 230's scope. Jessica Guynn, "Trump vs. Big Tech: Everything you need to know about Section 230 and why everyone hates it", USA Today Tech, https://eu.usatoday.com/story/tech/2020/10/15/trump-section-230-facebook-twitter-google-conservative-bias/3670858001/.

37 Spelled out by Nancy Pelosi, Speaker of the House in an interview: "But I do think that for the privilege of 230, there has to be a bigger sense of responsibility on it". https://www.vox.com/2019/4/12/18307957/nancy-pelosi-donald-trump-twitter-tweet-cheap-freak-presidency-kara-swisher-decode-podcast-interview.

38 See also in: Judit Bayer, Bernd Holznagel, Katarzyna Lubianiec, et al., "Disinformation and propaganda: impact on the functioning of the rule of law and democratic processes in the EU and its Member States , 2021 update". http://www.europarl.europa.eu/RegData/etudes/STUD/2021/653633/EXPO_STU(2021)653633_EN.pdf

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

outside (the pandemic) crisis? If yes, this would bring online platforms' services a big step closer to that of media providers. Facebook News services are, similarly, a type of content aggregation that has been selected and promoted by the platform.[39] Questions of responsibility and accountability for these remain.

An extreme interpretation of this liberal approach has been taken regarding search engines in the US.[40] It has been argued that Baidu, or Google, have First Amendment rights to select and edit the search results of their users.

This selection and sorting is "a mix of science and art" and a way of "how each search engine company tries to keep users coming back to it rather than to its competitors".[41] In this logic, it is entirely users' risk whether the search results are trustworthy. The monopolistic status of search engines may provide a new perspective. Liability for generating own content is less ambiguous; platforms would bear content providers' liability (rather than hosting providers' only). Proposed measures under the draft Digital Markets Act (DMA)[42] would prohibit gatekeepers from giving their own content priority in the ranking (Article 6.1.d. DMA), but gatekeepers would nevertheless still be allowed to provide such services.

German case law also provides examples for this more liberal approach. Their main line of argument is that platforms' TOS may set the "house rules" of the company as a result of their freedom of entrepreneurship (Article 12 of the German Basic Law). Those rules may depart from the Constitution and may restrict content that would otherwise be protected by the right to freedom of expression.[43] These rules should, however,

---

39 Facebook News, 'Introducing Facebook News'.

40 Eric Goldman, "Of Course The First Amendment Protects Baidu's Search Engine, Even When it Censors Pro-Democracy Results", Forbes Cross-Post (blog), Technology and Marketing Law Blog, March 28, 2014, https://www.forbes.com/sites/ericgoldman/2014/03/28/of-course-the-first-amendment-protects-baidus-search-engine-even-when-it-censors-pro-democracy-results/?sh=1d21a62b4ec8.

41 Eugene Volokh and Donald Falk, "Google – First Amendment Protection for Search Engine Search Results", UCLA School of Law Research Paper No. 12-22, April 10, 2012, http://dx.doi.org/10.2139/ssrn.2055364.

42 Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) (2020), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A842%3AFIN.

43 OLG Karlsruhe, 28.02.2019 - 6 W 81/18, NJW-RR 2019, 1006; LG Frankfurt/Main, 10.09.2018 - 2-03 O 310/18, MMR 2018, 770; See also: Daniel Holznagel, "Put-back- Ansprüche gegen soziale Netzwerke: Quo vadis?", (2019) 8 CR

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

still be subject to the German Civil Code, which provides for principles of fairness concerning general TOS (see above). Their respective market power impacts the evaluation of the TOS, as monopolistic companies owe a higher level of responsibility to provide fair conditions. This brings us to the enhanced responsibility of those companies whose services are comparable to a public function (see below).

Chapter 3. The bigger picture

As mentioned, the two schools of interpretation are not strictly separate in reality. Systemic-level regulatory approaches would be able to connect them, acting as an umbrella. One umbrella approach is infrastructural regulation (3a). The other is the emerging interpretation of the direct applicability of international human rights (3b). Both perspectives understand online platforms to be uniquely powerful actors of the global market and are therefore expected to apply primarily to very large market players.

Chapter 3.a. Infrastructural regulatory approach

Infrastructural regulation may serve as a bridge between the two schools of interpretation. Legal acknowledgement of some platforms' dominant status on the market leads to passing rules on interoperability and regulating the contracting terms of these actors. In the European Union, the Digital Markets Act has gone this direction by defining 'gatekeepers' and imposing on them the obligation to apply fair contractual terms with their business users (Article 5-6 DMA). There is discussion of treating platforms as public utilities in the US, comparing them to a range of industries, from railroads to certain media outlets, in the position of a gatekeeper.[44] This perspective may lead to antitrust considerations and rules of interoperability.

This approach may not appear to relate directly to content regulation and users' rights; however, the search for the appropriate role of online

---

35, no. 8 (2019): 518-526; Matthias Friehe, "Löschen und Sperren in sozialen Netzwerken", NJW 73, no. 24 (2020): 1697-1702.

44 Nikolas Guggenberger, "Essential Platforms", Yale Law & Economics Research Paper 24, no. 2 (2020): 237-343, https://ssrn.com/abstract=3703361 or http://dx.doi.org/10.2139/ssrn.3703361.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

platforms is a search for the appropriate power balance in a market where private corporations control access to services that are becoming vital to society.[45] Not only are broadband internet, finances, and e-commerce vital, but so is participation in online communities. The market power and monopoly status of a service provider have a crucial impact on users not only as consumers but also as citizens. It directly affects their fundamental right to receive and impart information (Article 10 ECHR, Article 11 Charter of Fundamental Rights of the European Union, Article 19 ICCPR).

Chapter 3.b. Horizontal effect of human rights

The analogy to public utilities also raises questions about contracting obligations. For example, are online platforms entitled to ban anyone permanently from their services? A German court assessed this question and found that Facebook has no obligation to conclude a contract, even if they are in a monopolistic position.[46] However, their dominance may impact how the Terms of Services are judged (see above). In another case, the Constitutional Court found that where excluding a user from services would significantly influence that user's social participation, the service provider may only do so under certain conditions and when respecting safeguards. Among these, the service provider must respect the right to a fair trial, allow a hearing and give reasons for decisions. This ruling related to a ban from sports establishments for extremist behaviour, and it is undecided whether it applies to platform media as well.[47]

Suspension of a user account has become a central issue after Facebook and Twitter suspended the account of US President Donald Trump for posts that were regarded as inciting violence during an attack on the Capi-

45 K. Sabeel Rahman, "The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept", Cardozo Law Review 39, no. 5 (2018): 1621-1692.

46 LG Görlitz, 29.11.2019 - 1 O 295/19 EV, MMR 2020, 196; OLG Dresden, 16.06.2020 - 4 U 2890/19, MMR 2021, 58.

47 Judgment of the German Constitutional Court, BVerfG, 11.04.2018 - 1 BvR 3080/09, Stadionverbot, NJW 2018, 1667.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tol.[48] The much-debated decision was referred to the Facebook Oversight Board for a decision on its lawfulness.

The Facebook Oversight Board was established by the largest social media platform to interpret and decide standards for the platform. The platform commissions the Board members, but its organisation is independent. The Charter of the Board stipulates its competences and defines the extent of Facebook's obligation to follow its decisions.[49] Thus, the quasi-authoritative body gives the impression of independent oversight, supported by the diversity and competence of its members, but it is in fact part of the platform's voluntary self-regulation. (See a critical analysis of the construction by Mårten Schultz in this volume).

In its decision about Donald Trump,[50] the Board found that the decision to suspend his account was justified. However, the terms of contract and Community Standards of the platform provided for either definite-period suspension or ultimate exclusion from the platform. Suspension for an indefinite period, in the absence of criteria defining whether and when the account will be reinstated, violated these terms and standards. The Board did not overrule Facebook's decision on the merits of suspension but instead referred the case back for review and gave principles to guide the new decision. [51]

When discussing the roles and obligations of platforms to their users, the question of whether platforms are subject to human rights obligations inevitably emerges. The Facebook Oversight Board relies on principles of public international law in its decision-making. Facebook asserted it is bound by the UN Guiding Principles on Business and Human Rights (UNGPs) in March 2021. Additionally, the Board also referred to the Rabat Plan of Action, General Comment No. 34 of the Human Rights Committee (2011), and the UN Special Rapporteur's report on freedom of opinion and expression A/HRC/38/35 (2018).

---

48 "The Capitol Attack Was the Most Documented Crime in History. Will That Ensure Justice?", Time, 9 April 2021, https://time.com/5953486/january-capitol-attack-investigation/.

49 Oversight Board Charter, https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf.

50 Decision 2021-001 FB-FBR, https://www.oversightboard.com/decision/FB-691QAMHJ.

51 Judit Bayer, "The Power of Softness, The Trump Decision of the Facebook Oversight Board", Inforrm's Blog, May 11, 2021, https://inforrm.org/2021/05/11/the-power-of-softness-the-trump-decision-of-the-facebook-oversight-board-judit-bayer/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

German jurisprudence has a clear stance on this issue. Since the *Lüth* case,[52] German Basic Law is held to have an indirect effect on individuals as third parties in relation to private entities (indirect third-party effect). This has been reinforced by several decisions, as cited above, which declared that online platforms, although not directly bound by the Basic Law, should respect its principles on fundamental rights.[53] However, the exact extent of this legal requirement has not yet been conclusively discussed.[54] Hungarian constitutional case law also holds that the state has a positive obligation to ensure the necessary conditions for democratic public opinion to remain operative,[55] for example, through public service media.[56]

In contrast to the European approach, the US posits that private entities are not bound by the Constitution as a result of the state action doctrine.[57] With a few exceptions,[58] the US courts generally reject the idea that private entities would be bound to respect human rights.[59]

International human rights bodies take the view that states are obliged to ensure the protection of human rights even vis-a-vis private entities. This means that individuals are entitled to seek redress against perceived violations by private entities. Therefore, states owe a responsibility under international law to prevent, punish and remediate human rights violations by private entities.[60] Jørgensen and Zuleta argue that the UN appears

---

52 BVerfG, 15.01.1958 - 1 BvR 400/51.

53 LG Frankfurt/Main, 10.09.2018 - 2-03 O 310/18, MMR 2018, 770; , LG Frankfurt/Main, Beschluss vom 14.05.2018 - 2-03 O 182/18, MMR 2018, 545; see also BVerfG Lüth-Urteil, 15.01.1958 - 1 BvR 400/51, NJW 1958, 257.

54 Jörn Reinhardt and Melisa Yazicioglu, "Grundrechtsbindung Und Transparenzpflichten Sozialer Netzwerke", Den Wandel Begleiten - IT-Rechtliche Herausforderungen Der Digitalisierung, 2020, 819.

55 Hungarian Constitutional Court, 30/1992. (V. 26.).

56 László Majtényi, Máté Szabó, Alkotmányjog (Eötvös Károly Közpolitikai Intézet, 2005). https://regi.tankonyvtar.hu/hu/tartalom/tkt/alkotmanyjog/index.html

57 Amélie Heldt, "Trump's Very Own Platform? Two Scenarios and Their Legal Implications", JuWissBlog, January 11, 2021, https://www.juwiss.de/03-2021/.

58 Marsh v. Alabama, 326 U.S. 501 (1946), https://supreme.justia.com/cases/federal/us/326/501/; PruneYard Shopping Center v. Robins, 447 U.S. 74 (1980), https://supreme.justia.com/cases/federal/us/447/74/.

59 See this in detail by: Amélie Heldt, "Merging the Social and the Public: How Social Media Platforms Could be a New Public Forum" Mitchell Hamline Law Review 46, no. 5 (2020): https://ssrn.com/abstract=3460067.

60 UNHR Committee, General Comment no. 31. The nature of the general legal obligation imposed on state parties to the Covenant, (CCCPR/C/21/Rev.1./Add.13) 2004, para. 8 (p.54-55).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to foster the view that human rights standards apply to companies. Rather than owing direct responsibility, however, their obligation is akin to the "risk assessment" method (see below). [61]

The Council of Europe takes a pro-active attitude in this respect. Under the European Convention on Human Rights, states are obliged to prevent, protect, and remediate human rights violations by private entities. Moreover, the Committee of Ministers is occupied with the issue of the human rights responsibilities of private corporations. In its 2012 Recommendation on the Protection of Human Rights with Regard to Social Networking Services, the Committee called upon online intermediaries to "respect human rights and the rule of law" by implementing self- and co-regulatory mechanisms, including procedural safeguards and accessible, effective remedies.[62] Further, it explicitly referred to the UN Guiding Principles in its 2014 Recommendation as a guide to human rights for Internet users, and suggested that platforms should respect the standards of the European Convention on Human Rights (ECHR) in their content removal, deletions and suspensions of user accounts.[63] The EU Charter of Fundamental Rights also seems to have horizontal effect, as shown by a decision of the Court of Justice of the European Union (CJEU)[64] and academic authors.[65]

Under the European Court of Human Rights (ECtHR) case law, states have a positive obligation to actively promote pluralism in society and the

61 Rikke Frank Jørgensen and Lumi Zuleta, "Private Governance of Freedom of Expression on Social Media Platforms", Nordicom Review 41, no. 1 (2020): 51–67, https://doi.org/10.2478/nor-2020-0003.

62 Recommendation CM/Rec (2012)4 of the Committee of Ministers on the Protection of Human Rights with Regard to Social Networking Services.

63 Recommendation CM/Rec (2014)6 of the Committee of Ministers on a guide to human rights for Internet users suggests that platforms should respect the standards of the ECHR in their content removal and account for removal decisions, at 53.

64 Joined cases C-569/16 and C-570/16 Stadt Wuppertal v. Maria Elisabeth Bauer and Volker Willmeroth v. Martina Broßonn, Judgment of 6 November 2018, discussed by Dorota Leczykiewicz, "The Judgment in Bauer and the Effect of the EU Charter of Fundamental Rights in Horizontal Situations", European Review of Contract Law 16, no. 2 (2020): 323–333, https://doi.org/10.1515/ercl-2020-0017.

65 Eleni Frantziou, "The Horizontal Effect of the Charter of Fundamental Rights of the EU: Rediscovering the Reasons for Horizontality", European Law Journal 21, no. 5 (2015): 657–679, https://fra.europa.eu/en/node/35696.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

media.[66] This positive obligation extends to ensuring an environment that is favourable to freedom of expression.[67]

States also have a positive obligation to ensure respect for private life (Article 8 ECHR).[68] In the context of social media, privacy includes autonomy in developing one's social life and online persona, in being seen by others as one chooses to be seen.[69] However, not all interferences with individual human rights involving online intermediaries would trigger states' positive obligations.[70]

In sum, there is growing academic literature and court practice concerning the horizontal effect of human rights owed by companies, including to respect the rights of individuals. However, its exact interpretation is still in development.[71]

Chapter 4. Conclusion

Online platforms fulfil a new role in e-business and public communication with significant new characteristics that differentiate them from previously known industry actors. The content ranking, recommending, prioritising, and deprioritising choices of these platforms are currently not addressed by legal rules, even though these decisions have a major impact on users' online experiences. Commercial platforms' activity affects economic pro-

66 Tarlach McGonagle, "The Council of Europe and Internet Intermediaries: A Case Study of Tentative Posturing", in Human Rights in the Age of Platforms, 242. Edited by Rikke Frank Jørgensen and David Kaye. Cambridge, MA: The MIT Press, 2019.

67 McGonagle, (2019) cites: Dink v. Turkey, nos. 2668/07 and 4 others, September 14, 2010.

68 Marckx v Belgium App, no. 6833/74, S. A No 31 [31] (1979), Đorđević v Croatia App. No. 41526/10 ECHR 2012-V [87]–[88] (2012).

69 See more in: Lorna Woods, "Social media: it is not just about Article 10" in: The Legal Challenges of Social Media, edited by David Mangan, Department of Law, Maynooth University and Lorna E. Gillies, Edinburgh Napier University, UK, Elgar Law, Technology and Society series, 2017.

70 McGonagle, (2019) cites: ECtHR, 2017. Tamiz v. the United Kingdom, No. 3877/14 (2017), para. 82-84. and Pihl v. Sweden, No. 74742/14 (2017).

71 See more on this: McGonagle, (2019), Agnès Callamard, "The Human Rights Obligations of Non-State Actors" in Human Rights in the Age of Platforms, 191, edited by Rikke Frank Jørgensen and David Kaye. Cambridge, MA: The MIT Press, 2019; see also: Gunther Teubner, "Horizontal Effects of Constitutional Rights on the Internet: A Legal Case on the Digital Constitution", The Italian Law Journal 3, no. 1 (2017): 193-205.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

cesses, whereas social media platforms affect communicative processes. The latter directly impacts the public discourse and, therefore, the democratic processes.

This paper has compared two regulatory approaches. One leaves decisions regarding content governance entirely to the platform. At its extreme, platforms are free to moderate content and remove lawful or carry unlawful content without governmental supervision or interference (notwithstanding judicial orders) (US, CDA 230). In its more moderated form, platforms owe a duty of care but are free to decide how they fulfil this duty of a well-maintained platform (UK, Statutory Duty of Care, see more in this volume by Lorna Woods).

The other approach would define rather precisely what type of content is to be removed or moderated and, in its extreme, would not tolerate the removal of lawful content. However, this extreme version is seen only sporadically. In reality, the approaches are mixed. For example, the EU's Digital Services Act provides for the removal of illegal content upon notice and sets out obligations to respect procedural rights in the notice and removal process. It orders platforms to carry out risk assessments and mitigate risks in a co-regulatory framework (EU, DSA).

Viewed critically, platforms act either as regulators themselves or as vectors of state regulation. The first case raises the suspicion of private censorship, whereas the second attracts the criticism of states' outsourcing censorship.[72]

Finally, the paper examined how private entities can become directly responsible for human rights: by the horizontal effect of human rights and an enhanced responsibility due to their market dominance or, perhaps, by obtaining a public utility status.

In a search to find the best option to ensure the – sometimes conflicting – human rights of users are respected, we find ourselves between a rock and a hard place, having to decide whether we prefer regulation by the state or by private actors.

With political accountability in a democratic system, a state would be better equipped to regulate in a field interwoven with fundamental rights sensitivities. However, this is unpractical in many ways due to the vast amount of content, cultural diversity of users, and fast development of

---

72 Marc J. Bossuyt, Guide to the "Travaux préparatoires" of the International Covenant on Civil and Political Rights, Leiden, Dordrecht: Martinus Nijhoff, 1987, 385. See also: Molly K. Land (2013) "Toward an International Law of the Internet", Harvard Law Review 54, no. 2 (2013): 393, 445; see also: Callamard (2019).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

technology. Further, in many authoritarian states, online platforms bring a fresh breeze of liberalism and ensure freedoms that could not otherwise be exercised.

Online social participation has become an indispensable necessity for many. Like so many achievements of civilisation, from clean water to education, it is possible but not desirable or acceptable for one to live without access to social platforms. However, the unregulated and unaccountable power of online platforms may lead to arbitrary decisions affecting citizens in ways that are seen as disproportionate.

There is one agreeable point between the various approaches: the standards pledged by online platforms themselves are contractual terms, or "house rules", and should be abided by as a minimum.

Bibliography

Bayer, Judit, Natalija Bitiukova, Petra Bard, Judit Szakács, Alberto Alemanno, and Erik Uszkiewicz. "Disinformation and Propaganda – Impact on the Functioning of the Rule of Law in the EU and Its Member States." *HEC Paris Research Paper No. LAW-2019-1341*, Update 2020.

Bayer, Judit. "The Power of Softness, The Trump Decision of the Facebook Oversight Board." *Inforrm's Blog*, May 11, 2021.

Bossuyt, Marc J. *Guide to the "Travaux préparatoires" of the International Covenant on Civil and Political Rights*, Leiden, Dordrecht: Martinus Nijhoff, 1987.

Callamard, Agnès. "The Human Rights Obligations of Non-State Actors." In *Human Rights in the Age of Platforms*, 191. Edited by Rikke Frank Jørgensen and David Kaye. Cambridge, MA: The MIT Press, 2019.

Frantziou, Eleni. "The Horizontal Effect of the Charter of Fundamental Rights of the EU: Rediscovering the Reasons for Horizontality." *European Law Journal* 21, no. 5 (2015): 657–679.

Friehe, Matthias. "Löschen und Sperren in sozialen Netzwerken", *NJW* 73, no. 24 (2020): 1697-1702.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press, 2018.

Goldman, Eric. "Of Course The First Amendment Protects Baidu's Search Engine, Even When it Censors Pro-Democracy Results", *Forbes Cross-Post* (blog), *Technology and Marketing Law Blog*, March 28, 2014. https://www.forbes.com/sites/ericgoldman/2014/03/28/of-course-the-first-amendment-protects-baidus-search-engine-even-when-it-censors-pro-democracy-results/?sh=1d21a62b4ec8.

Guggenberger, Nikolas. "Essential Platforms." *Yale Law & Economics Research Paper* 24, no. 2 (2020): 237-343.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Heldt, Amelie. "The President and Free Speech: Consequences of Twitter's Fact-Checking Indication." *Internet Policy Review*, June 4, 2020.

Heldt, Amélie. "Trump's Very Own Platform? Two Scenarios and Their Legal Implications." *JuWissBlog*, January 11, 2021.

Heldt, Amélie. "Merging the Social and the Public: How Social Media Platforms Could Be a New Public Forum." *Mitchell Hamline Law Review* 46, no. 5 (2020).

Holznagel, Daniel. "Put-back- Ansprüche gegen soziale Netzwerke: Quo vadis?", (2019) 8 CR 35, no. 8 (2019): 518-526.

Jørgensen, Rikke Frank, and Lumi Zuleta. "Private Governance of Freedom of Expression on Social Media Platforms." *Nordicom Review* 41, no. 1 (2020): 51–67.

Klonick, Kate. "The Most Important Lesson from the Leaked Facebook Content Moderation Documents." *Slate.com*, 29.6.2017.

Land, Molly. "Toward an International Law of the Internet." *Harvard Law Review* 54, no. 2 (2013): 393-458.

Leczykiewicz, Dorota. "The Judgment in Bauer and the Effect of the EU Charter of Fundamental Rights in Horizontal Situations." *European Review of Contract Law* 16, no. 2 (2020): 323–333.

McGonagle, Tarlach. "The Council of Europe and Internet Intermediaries: A Case Study of Tentative Posturing." In *Human Rights in the Age of Platforms*, 242. Edited by Rikke Frank Jørgensen and David Kaye. Cambridge, MA: The MIT Press, 2019.

Rahman, K. Sabeel. "The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept." *Cardozo Law Review* 39, no. 5 (2018): 1621-1692.

Reinhardt, Jörn, and Melisa Yazicioglu. "Grundrechtsbindung Und Transparenzpflichten Sozialer Netzwerke." In *Den Wandel Begleiten - IT-Rechtliche Herausforderungen Der Digitalisierung*, 819. Jürgen Taeger, 2020.

Schwär, Hannah and Moynihan, Qayyah. "Instagram and Facebook are intentionally conditioning you to treat your phone like a drug." *Business Insider*, o5.07.2020.

Schwartmann, Rolf, and Robin Mühlenbeck. "NetzDG Und Das Virtuelle Hausrecht Sozialer Netzwerke." *ZRP*, (2020): 170–172.

Teubner, Gunther. "Horizontal Effects of Constitutional Rights on the Internet: A Legal Case on the Digital Constitution." *The Italian Law Journal* 3, no. 1 (2017): 193-205.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

European Legislative Initiative for Very Large Communication Platforms

Jan Christopher Kalbhenn

Abstract: In December 2020, the European Commission published its drafts for a Digital Services Act and a Digital Markets Act. With this legislative project the Commission introduces new regulations for the content moderation and market behaviours of very large online platforms, especially social networks. In addition to fixed requirements for all online platforms, due diligence requirements are also introduced for very large online platforms. This is intended to protect a wide range of legal interests, including public health, civil society discourse, or effects in connection with elections. This would also allow the Commission to push for further targeted measures in relation to hate speech, as well as disinformation under certain conditions and in the event of non-compliance with the rules of the Digital Services Act. It is possible that specifications on the interface design and algorithm architecture of the platform could be tailored to individual platforms.

Keywords: online platform; Digital Services Act; Digital Markets Act; content moderation; due diligence; media law; disinformation; hate speech; risk assessment and risk mitigation; design specifications; recommender system; social media; advertisement.

Chapter 1. Europe-wide regulation of digital platforms

The effects of the internet and platform economy were recently analysed by the media scientist and philosopher Joseph Vogl. His verdict is trenchant and drastic. From the rule of the financial markets to the new network giants to the dynamized opinion industry, lies a trail of destruction. Democracy, freedom and social responsibility are being damaged. In the digital age, new forms of entrepreneurial power have emerged that overwrite democracy with their own evaluation logic. Tech companies would intervene ever more massively in the decision-making of governments, so-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

cieties and economies across national borders.[1] The European Commission has also taken a look at the impact of the platform economy and the dominance of individual tech companies. Following the 2018 General Data Protection Regulation and the 2019 Copyright Directive, the Commission presented another legislative package for the internet in December 2020.[2] The draft Digital Markets Act contains competition rules for gatekeepers. The draft Digital Services Act contains media law requirements for platforms to protect fundamental rights on online platforms. Both sets of rules set particularly far-reaching specifications for especially large platforms. The Commission is thus also addressing the problem of hate speech and disinformation, not least in response to national go-it-alone measures such as the German Network Enforcement Act and the State Media Treaty.[3] Decision-making practice on abuse of dominant market positions by dominant platforms is also given legal form.

This article shows how the Commission intends to ensure protection of fundamental rights on large platforms and guarantee fair competition by holding very large platforms in particular to account and in doing so also imposing requirements on the architecture of the algorithms and design of platform interfaces.

Chapter 2. Digital Services Act and Digital Markets Act

Chapter 2.a. Background

In December 2020, the European Commission presented the European Action Plan for Democracy.[4] This is a catalogue of measures to be implemented over the entire term of the current Commission. The Commission's overarching goal is to empower citizens and build more resilient

1 Joseph Vogl, *Kapital und Ressentiment*, 2021.
2 List of EU Regulatory Instruments on Digital Platforms see Annex to this Article.
3 Another law with references to media law platform regulation is the Commission's proposed AI Act, See Kalbhenn, Jan „Designvorgaben für Chatbots, Deepfakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung", *ZUM – Zeitschrift für Urheber- und Medienrecht*, No. 8/9 (2021).
4 European Commission, Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, on the European democracy action plan, Brussels, 3.12.2020.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

democracies across the EU. Specifically, free and fair elections are to be promoted, media freedom expanded, and disinformation combated. In it, the European Commission states that the 'digital revolution' has changed democracy. In the digital realm, it is fundamentally challenging to enforce the law, and there are concerns about the transparency and accountability of online platforms. As concrete measures, the Commission announced uniform legislation on these issues across Europe. Many of the issues raised have so far been addressed through non-binding voluntary commitments and codes of conduct. These measures, for example in the area of hate speech and disinformation, were generally viewed positively. However, not least because of national solo efforts in regulation of online platforms such as social networks, the Commission has also recognized the need to achieve EU-wide harmonization of application of the law. For example, Germany, France and Austria already have or are planning initial laws to combat hate speech on social networks.[5] Germany has also already enacted the first media law regulations for communication platforms.[6]

A similar picture emerges in competition law. In recent years, the European Commission has increasingly conducted proceedings against the major platform companies and has regularly found abuse of market power.[7] National antitrust authorities in the Member States have also made high-profile decisions in this area, such as the German Federal Cartel Office prohibiting Facebook from combining user data from its Facebook, WhatsApp and Instagram services.[8]

With both draft regulations – the Digital Markets Act and the Digital Services Act – the Commission has initiated the legislative process. The EU

---

5 Maximilian-Hemmert-Halswick "Lessons learned from the first years with the NetzDG" (chapter in this book); these laws are also criticized for violating the principle of origin laid down in Art. 3 E-Commerce Directive. According to this, the place of establishment is decisive for an online company in legal terms and the respective member state is responsible for enforcing the law. The EU was forced to react to these developments and national advances with the Digital Services Act and to bring order to the legal system.

6 Bernd Holznagel and Jan Kalbhenn "Media law regulation of social networks" (chapter in this book).

7 Andreas Grünwald, "Big Tech-Regulierung zwischen GWB-Novelle und Digital Markets Act", *MMR – Zeitschrift für IT-Recht und Recht der Digitalisierung*, No. 12 (2020).

8 German Federal Cartel Authority, Case Summary, Facebook, Exploitative business terms pursuant to Section 19(1) GWB for inadequate data processing, 15 February 2019, https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=4.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

has decided to propose the legislative acts in the form of regulations. These laws would apply directly in all Member States of the European Union after a transition period, as also applied to the General Data Protection Regulation (GDPR). As a result, the Digital Services Act and the Digital Markets Act would supersede the previously applicable law in their areas of application in favour of uniform regulation. However, there is still a long way to go before the final text of the regulation is adopted.

Chapter 2.b. Regulatory targets

The Digital Service Act (DSA) has two main purposes. On the one hand, creation of uniform rules for all Member States is intended to promote the – digital – single market.[9] Another objective is to ensure protection of EU citizens' fundamental rights on the internet.[10] This primarily involves protection of freedom of expression, protection of the personal rights of those affected by hate speech, and protection of freedom of information.

The Digital Markets Act (DMA) is also intended to impose harmonised rules on central platform services throughout Europe by way of a regulation, thus ensuring competition and fair digital markets throughout the Union in which gatekeepers operate. [11]

Chapter 2.c. Focus on very large platforms

To achieve these goals, the Digital Services Act creates a comprehensive set of regulations for the online economy and addresses intermediaries. Media law regulations are also created or supplemented in the process. The draft follows the principle of graduated responsibility. The decisive factor is initially how "close" the intermediary is to the content and to which group the content is made accessible. Only rudimentary obligations apply to companies that are solely responsible for infrastructure or temporary intermediate storage, such as internet access providers. Extended obligations apply to hosting services such as cloud and web hosting providers. The Digital Services Act imposes strict requirements on online platforms. These are defined very broadly as hosting service providers that allow

9 Art. 1 sec. 1 DSA.
10 Art. 1 sec. 1 DSA.
11 Art. 1 sec. 1 DMA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

users to store and share information with the public.[12] The size of online platforms also plays a role. Small platforms are excluded from the scope of specific obligations and are spared in favour of innovativeness.[13] Very large online platforms, on the other hand, are subject to significant obligations. These are online platforms that have an average of 45 million active users in the EU.[14] Very large online platforms include Facebook, Twitter, YouTube, Twitch, Instagram, and TikTok.

The Digital Markets Act imposes further binding obligations on these digital companies. It focuses on 'central platform services'. These are a series of services that are listed exhaustively. They include online brokerage services such as AirBnB, online search engines such as Google Search, social networks such as Instagram and TikTok, video sharing platform services such as YouTube, messenger services such as WhatsApp, operating systems, cloud computing services, and advertising services, including advertising networks and advertising exchanges. The obligations of the Digital Markets Act only apply to operators of central platforms if they are designated as gatekeepers pursuant to Art. 3 DMA. The prerequisite for this designation is that the platform service has a significant impact on the internal market, and operates a central platform service that serves commercial users as an important gateway to end users. With regard to its activities, it must hold a consolidated and permanent position. However, it is also sufficient if it is foreseeable that it will attain such a position in the near future.[15] Art. 3 DMA regulates the procedure to ensure that the Commission becomes aware of the fact that a company's thresholds have been reached. Gatekeeper status will be reviewed on a regular basis, and the designation may be changed or revoked.[16] Thus, the Digital Markets Act basically covers such platforms that are addressed in the Digital Services Act as very large platforms – including TikTok, Instagram, Twitter, and so on.

Chapter 3. The new ABC of European platform regulation

The Digital Services Act sets out to make the internet a secure, predictable and trustworthy environment in the age of the platform economy and social networks. The fundamental rights enshrined in the European Char-

12 Art. 2 lit. h DSA.
13 Art. 16 DSA.
14 Art. 25 DSA.
15 Art. 3 sec. 1 DMA.
16 Art. 4 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ter of Fundamental Rights are to be effectively protected. The definition catalogue in Article 2 of the Digital Services Act already sets out the field for this. The dangers to certain legal interests posed by platforms come primarily from the content disseminated there and the way content is presented and weighted. [17] It is therefore not surprising that the definition catalogue contains many key terms that relate to certain categories of content (advertising, illegal content) or their mediation (content moderation, recommendation system). In some cases, these terms are now being defined for the first time.

Chapter 3.a. Content moderation

The term 'content moderation' is central to the goals and objectives of the Digital Services Act. This is understood by the draft to mean the activities of providers of intermediary services to identify, determine and combat illegal content or information provided by users that is incompatible with the provider's general terms and conditions. This includes measures related to the availability, visibility and accessibility of illegal content or information.[18] Downgrading, blocking access or removal are given as examples. Also included are measures that restrict the ability of users to provide information. This also includes closure or temporary suspension of a user account for content moderation. This definition is very broad. Thus, the Digital Services Act affects all means available to platforms to manage content.

Chapter 3.b. Illegal content

Illegal content is a special category of content to which the Digital Services Act attaches certain legal consequences. The Digital Services Act defines this as all information that does not comply with EU law or the law of a Member State.[19] This can also include content that violates the law by referring to an activity. It also covers sale of products or provision of services. This very broad definition and the equally broad definition

17 Sinan Aral, *The Hype Machine*, London, 202; Maik Fielitz and Holger Marcks, *Digitaler Faschismus*, Berlin 2020.
18 Art. 2 lit. p DSA.
19 Art. 2 lit. g DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of online platforms result in a wide scope of application of the Digital Services Act. Even trading platforms such as Amazon and eBay are subject to the regulations on content moderation of illegal content.

Chapter 3.c. Advertising

Advertising is central to the business model of many platforms.[20] Even the Amazon trading platform is increasingly generating revenue from advertising. Advertising is a special content category to which both the Digital Service Act and the Digital Markets Act attach certain legal obligations. For both sets of regulations, the Digital Services Act defines what is meant by advertising. According to this definition, it is information intended to disseminate the message of a legal or natural person that is displayed by an online platform for publicity in return for payment.[21] Advertising for non-commercial purposes is also included. In terms of legal consequences, the Digital Services Act differentiates between general advertising and advertising 'delivered' by micro-targeting.

Chapter 3.d. Recommendation systems

Not least to deliver money-making content, advertising, to the user, recommendation systems are essential components of the architecture of online platforms. Without algorithmic moderation, organisation of the mass of content would not be possible. At the same time, the personalization they enable is a central component of (advertising) business models. The Digital Services Act defines this as a fully or partially automated system used by an online platform to suggest specific information to users.[22] This can be triggered either by a search or by other means. This must determine the relative order or prominence of the information displayed.

---

20 Tim Hwang, *Subprime Attention Crisis*, New York, 2020. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: PublicAffairs, 2019.

21 Art. 2 lit. n DSA.

22 Art. 2 lit. o DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3.e. General terms and conditions

The legal relationship between online platforms and their users is initially governed by civil law. This is usually done by means of general terms and conditions. What is meant by this is defined uniformly for all Member States by the Digital Services Act. They are any terms, conditions or specifications, regardless of their name or form, that govern the contractual relationship between the provider of intermediary services and users.[23] Behind this are also the community standards that have reached a high level of detail on communication platforms such as Facebook, for example, and according to which content is deleted or blocked millions of times. The Digital Services Act does not shy away from intervening in the contractual relationship between platforms and users and prescribing minimum requirements.

Chapter 4. Rigid requirements for content moderation in the Digital Services Act.

Overview of new obligations[24]

	Intermediary services *(cumulative obligations)*	Hosting services *(cumulative obligations)*	Online platforms *(cumulative obligations)*	Very large platforms *(cumulative obligations)*
Transparency reporting	■	■	■	■
Requirements on terms of service due on account of fundamental rights	■	■	■	■
Cooperation with national authorities following orders	■	■	■	■

23 Art. 2 lit. q DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

	Intermediary services *(cumulative obligations)*	Hosting services *(cumulative obligations)*	Online platforms *(cumulative obligations)*	Very large platforms *(cumulative obligations)*
Points of contact and, where necessary, legal representative	■	■	■	■
Notice and action and obligation to provide information to users		■	■	■
Complaint and redress mechanism and out of court dispute settlement			■	■
Trusted flaggers			■	■
Measures against abusive notices and counter-notices			■	■
Vetting credentials of third- party suppliers ("KYBC")			■	■
User-facing transparency of online advertising			■	■
Reporting criminal offences			■	■
Risk management obligations and compliance officer				■
External risk auditing and public accountability				■
Transparency of recommender systems and user choice for access to information				■
Data sharing with authorities and researchers				■
Codes of conduct				■

24 https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

	Intermediary services *(cumulative obligations)*	Hosting services *(cumulative obligations)*	Online platforms *(cumulative obligations)*	Very large platforms *(cumulative obligations)*
Crisis response cooperation				■

Chapter 4.a. Transparency as a basic rule of content moderation

With the central provision in Article 12 Digital Services Act, the legislator intervenes in the contractual relationship between platform and user. The Digital Services Act supplements contract law in the area of platform general terms and conditions (GTCs) and community standards. The content of GTCs is not specified, for example by model GTCs. However, certain information must be provided. For example, information must be provided on any restrictions on the information provided by users that they impose in connection with use of their service. Disclosures must include information about any policies, procedures, measures, and tools used to moderate content, including algorithmic decision making and human review. This is appropriate since content moderation is now heavily processed algorithmically. [25] Information must also be understandable and made publicly available in an easily accessible form. If these rules are part of the contract, users can also take legal action to enforce them.

Online platforms must also clearly state in their terms and conditions how they handle account suspensions.[26] The Digital Services Act stipulates those accounts of users who frequently provide obviously illegal content must be blocked. The Digital Services Act thus defines a minimum standard of protection. However, platform providers can also[27] set a higher standard of protection as long as fundamental rights are respected. This is because, according to Art. 12(2) Digital Services Act, when applying and enforcing the restrictions designated in their community standards, they

25 Kate Klonick, "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression", *The Yale Law Journal*, 2021.
26 Art. 20 sec. 4 DSA.
27 Art. 20 sec. 1 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

must do so carefully, objectively and proportionately, taking into account the rights of all stakeholders, as well as the applicable fundamental rights of users. This makes the fundamental rights of users the benchmark for content moderation on online platforms.

In their general terms and conditions, online platforms must also present the key parameters of recommendation systems.

Chapter 4.b. Account suspensions in case of abusive behaviour

For the first time, a regulation uniform for all online platforms is envisaged, which would set the conditions under which accounts on communication platforms are to be blocked. The standard formulates a minimum standard that does not prevent online platforms from providing stricter regulations in their community standards.[28] Online platforms are to suspend user accounts at least temporarily in the event of abusive behaviour – if a user frequently posts obviously illegal content. In this context, that is the case if a layperson recognizes it as evidently unlawful without closer examination. [29]

Chapter 4.c. Recommendation systems

With the design of user interfaces, online platforms can strongly influence users' decisions. Selection behaviour by users depends on how highlighted or hidden, understandable or incomprehensible are certain functions offered.[30] If legislators are concerned that a function is not hidden from users by platform services, they can use design specifications to ensure that a particular option is present in the interface design. The Commission has opted for such a requirement in the area of algorithmic recommendation systems for content moderation, to which the Commission rightly attaches central importance in dissemination of content.[31] In the recitals, the Commission refers to the considerable potential of systems to spread certain messages virally. The Digital Services Act initially aims to counter these

28 Recital 47 DSA.
29 Recital 47 DSA.
30 Cliff Kuang and Robert Fabricant, *User Friendly*, London 2019.
31 Natali Helberger, "On the Democratic Role of News Recommenders", 2019, *Digital Journalism*, 993-1012.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

risks through transparency. Very large online platforms must therefore present the most important parameters of recommendation systems in an accessible and easily understandable way in their general terms and conditions. All options with which the most important parameters can be changed or influenced are to be pointed out. User autonomy is to be strengthened by providing at least one profiling-free (as defined by the GDPR) option.[32] The Digital Services Act makes a design specification in the event that several such options are provided. In that case, the design of the user interface must provide an 'easily accessible function' for the user to select the recommendation system.

Chapter 4.d. Complaint management for illegal content

The Digital Services Act provides a differentiated regime for dealing with illegal content. The principle of 'notice-and-takedown' continues to apply. The new requirements for complaint management aim to make it as easy as possible for platform users or civil society organizations to give notice. By imposing organisational requirements on network operators, they are to be given opportunities to have illegal content removed from online platforms. The Digital Services Act does not contain details on takedown contrary to the German Netzwerkdurchsetzungesgesetz (NetzDG) that sets time limits for deletion or blocking of content. Again, stricter requirements are placed on online platforms and very large online platforms than on hosting services.

a) Upward compatible ground rules for all hosting services

The basic rules for hosting providers are upwardly compatible and apply to all online platforms. All hosting services must set up an easy-to-use complaints system.[33] This is intended to allow users to submit complaints that enable providers to make a qualified decision on the illegality of the content. Consistently, certain requirements must be met. To be included: Reasons for the illegality, exact location (URL), name and e-mail address of the complainant included. In addition, the complainant should receive an acknowledgement of receipt and is entitled to a speedy decision. If

32 Art. 29 sec. 1 DSA.
33 Art. 14 sec. 1 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the decision is based on artificial intelligence or automation, this must be made transparent.

If content is removed or blocked, the person concerned should be fully informed of the reasons.[34] The legal standard violated must be stated, as well as the circumstances on which the decision is based. Reasons must also be given for violations of community standards.

b) Special regulations for online platforms

The rights of users are to be protected by differentiated procedural requirements. Online platforms should set up an internal complaints management system enabling checks on whether content has been deleted or blocked. Temporary suspension from platform use or deletion of the user account should also be handled via this.[35] The review must be free of charge and easily accessible. Complaints must be made available for violations of legal regulations but also of community standards. The decision on the complaint should also be made expeditiously and the complainant must be informed of the decision. The decision in the complaint procedure must not be based exclusively on an automated procedure.[36] In the initial complaint procedure, on the other hand, a fully automated decision may be issued.[37] A human being must be involved in renewed control ("*human in the loop*"). Providers must draw the attention of the data subject to the possible alternative procedure in the decision.

c) Low-threshold out-of-court alternative procedure

Users whose content has been deleted or blocked should be able to challenge the decisions from the online platform complaints procedure in an out-of-court procedure.[38] For this purpose, out-of-court dispute resolution bodies are to be established, which in turn require recognition and

34 Art. 15 DSA.
35 For details on the NetzDG amendment 2021 see Hemmert-Halswick "Lessons learned from the first years with the NetzDG" (Chapter in this book).
36 Art. 17 sec. 5 DSA.
37 Kalbhenn and Hemmert-Halswick, „EU-weite Vorgaben für die Content-Moderation in sozialen Netzwerken", *ZUM – Zeitschrift für Urheber- und Medienrecht,* No. 3 (2021).
38 Art. 18 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

must first meet certain conditions – prove that they are impartial and independent of online platforms and users, have the necessary expertise, maintain clear and fair rules of procedure, and are easily accessible by electronic communication (18 (2) DSA). Member States are allowed to set up arbitration bodies themselves.[39] This offers civil society organizations an opportunity to help shape the legal framework for content moderation. There is also the option of seeking legal protection in court. [40]

d) Trusted flaggers

Another gateway for civil society to help shape content moderation is hidden in the regulation on trusted flags. This status can be granted to public bodies or non-governmental organizations and 'semi-public' bodies, for example organizations that report illegal, racist and xenophobic statements on the internet.[41] In content moderation, some platforms already rely on trusted flaggers. YouTube traditionally uses trusted partners in the area of copyright to feed the Content ID system.[42] In the area of other content control, YouTube also grants this status to individual organisations and confers on their reports increased trustworthiness. Such reports are processed more quickly. In the future, the Digital Services Act will shape this practice, which has so far been purely a matter of private law, into law.[43] Online platforms will then be obligated to ensure technically and organisationally that reports from trusted flaggers are processed with priority and without delay. In that way, the speed of measures against illegal content can be increased.

Trusted Flaggers may only be institutions but not individuals. They must prove that they have special expertise and competence in combating illegal content. It is also a prerequisite that they represent collective interests. They must work carefully and objectively.

The rule guarantees a legally secure status for Trusted Flagger from erratic platform decisions by providing legal certainty. YouTube currently reserves the right to change the eligibility requirements for the Trusted Flag-

39 Art. 18 sec. 4 DSA.

40 In Germany, there is already much case law on content moderation, *see* Holznagel and Kalbhenn "Media law regulation of social networks" (chapter in this book).

41 Recital 46 DSA.

42 Robert Gorwa et al., "Algorithmic content moderation: Technical and political challenges in the automation of platform governance", *Big Data & Society*, 2020.

43 Art. 19 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ger programme or suspend the programme at its discretion. This would be unlawful under the Digital Services Act. The complete opposite of a Trusted Flagger is regulated in Art. 20 (2) DSA, namely users who frequently submit notices or complaints that are manifestly unfounded. In the future, online platforms are to block these users from reporting further content.

Chapter 4.e. Serious crimes

Online platforms are to be obliged to inform the danger prevention or law enforcement authorities in the event of a suspected serious crime.[44] This is about protecting the life or safety of persons. The recitals make it clear that this requirement does not legitimize profiling or similar planned observations.[45]

Chapter 4.f. Advertising

One content category that is particularly valuable for platforms is advertising. The Digital Services Act distinguishes between advertising that is displayed equally to all users (standard advertising) and advertising that is displayed individually to users via micro-targeting.[46] Online platforms must make standard advertising clearly recognisable as advertising and allow the advertiser to be identified.[47] Advertising using micro-targeting should contain meaningful information about the key addressing parameters. The logic used should be explained in a meaningful way.[48]

Very large platforms are subject to even more stringent transparency requirements. They pose an increased risk due to their reach. They also have more data at their disposal to perfect behavioural analysis for targeted advertising, with the associated increased risks. Very large online platforms must now store the content of the ad, the advertiser, the period of the ad, the specification of recipient groups and important parameters for targeting, and the total number of recipients reached one year after the

44 Art. 21 DSA.
45 Recital 48 DSA.
46 On the human rights impact of microtargeting ads see Judit Bayer, "Double harm to voters: data-driven micro-targeting and democratic public discourse", *Internet Policy Review*, *9*(1) 2020.
47 Art. 24 DSA.
48 Recital 52 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

last insertion in a publicly accessible database.[49] Industry standards are intended to make advertising databases interoperable.[50] This should make it easier to analyse the risks associated with the spread of advertising. The Recitals of the Digital Services Act refer to unlawful advertising or manipulative techniques and disinformation that have a negative impact on public health, public safety, civil discourse, political participation and equality.[51]

For political advertising, the Commission has announced a legislative act in the Action Plan for Democracy.

Chapter 4.g. Official announcements

Very large online platforms also play a central role in informing citizens in crisis situations. Situations where public safety or public health are at risk – such as the Corona pandemic or attacks – misinformation spreads particularly quickly via online platforms and can lead to further damage. For such situations, the Commission is to develop crisis protocols for content moderation with Member State authorities.[52] For example, it may be regulated that information from national authorities is displayed prominently. Some platforms have implemented such measures voluntarily so far. Facebook prioritized displaying information from the World Health Organization during the Corona pandemic and enabled a missing-persons-search-feature during the attacks on the Bataclan theatre in Paris. This far-reaching regulation appears appropriate in view of the high reach of the platforms and their partial monopoly position. In European telecommunications law, it is still possible to set up public warning systems via messenger services.

Chapter 4.h. Interim summary

In the systematics of the Digital Services Act, the completed catalogue of rigid rules for content moderation represents a minimum standard

49 Art. 30 DSA. This rule builds on the Code of Conduct and has already been implemented by some platforms - not to the full satisfaction of critics - on a voluntary basis.
50 Art. 34 sec. 1 lit. b DSA.
51 Recital 63 DSA.
52 Art. 37 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

applicable to all online platforms, regardless of the business model of the platform service, the content distributed there, or the target group. Gradations are only made with regard to the size of online platforms. The rules apply in the same way to platforms as diverse as Airbnb, TikTok, Amazon and Parler. This is not surprising, given that minimum standards for protection of fundamental rights should be ensured by procedural rules on all platforms. It is striking that many of the rules are already in place in German media law, in the shape of the Network Enforcement Act of 2017 and the State Media Treaty.[53]

In order to counter highly complex dangers such as disinformation with targeted regulation, other factors must be taken into account. The business model pursued by the platform service, the media competence of the user community and, last but not least, the precise (algorithm) architecture and the interface design of platforms are all relevant. Architecture and design are significantly tailored to the business model. Only when these and other factors are included a sustainable regulation and a threat mitigation is possible. To contain systemic risks, the Digital Service Act therefore relies on flexible specifications for very large platforms and creates extensive due diligence obligations.

Chapter 5. Flexible specifications for systemic risks of very large platforms

Chapter 5.a. Risk assessment

For very large online platforms such as Facebook, Instagram, TikTok, YouTube, iTunes and Spotify, the Digital Services Act presents a flexible instrument aimed at protecting a wide range of legal interests and taking into account the specifics and business models of the services. Additional obligations are imposed for managing systemic risks. Central to this is a mechanism for assessing and minimizing risks. According to Art. 26 DSA, it is to become mandatory for very large online platforms to identify, analyse and assess all material systemic risks arising from the operation and use of their services once a year. Mandatorily, the risk analysis has to include the following three points:

53 Kalbhenn and Hemmert-Halswick, *"EU-weite Vorgaben für die Content-Moderation in sozialen Netzwerken"*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

- dissemination of illegal content,
- the negative impact on the exercise of fundamental rights (in particular, private and family life, freedom of expression and information, prohibition of discrimination, and rights of the child); and
- intentional manipulation of their service with a negative impact on protection of public health, minors, civil discourse, or impact related to elections and public safety.

Risks in the latter area can arise, for example, from the use of bots or (partially) automated communication.[54] Risk assessment must primarily consider content moderation systems, recommendation systems, and systems for selecting and displaying advertising.

Chapter 5.b. Minimisation of risks

Very large online platforms will be required to minimize the risks thus identified.[55] To this end, they are to take appropriate, proportionate and effective risk mitigation measures tailored to the systemic risks identified. A wide range of possible adjustments is conceivable here. This also applies to the design and architecture of the platforms. The law provides a non-exhaustive catalogue of examples of risk mitigation measures. According to this, risk mitigation can be achieved primarily by adapting content moderation or recommendation systems, decision-making processes, the features or functioning of their services, or their general terms and conditions. Targeted measures to restrict the display of advertising are also mentioned, as well as strengthening internal processes with regard to identifying systemic risks.

Chapter 5.c. Audit, data access law, reporting

It is initially the responsibility of the platforms to analyse and minimise risks. Whether the providers of very large online platforms also comply with these due diligence obligations is the subject of an annual independent audit. Detailed regulations are specified for this purpose. If very large online platforms receive a non-positive audit report, they must give due

54 Recital 68 DSA.
55 Art. 27 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

consideration to all operational recommendations addressed to them and take the necessary measures to implement them. If they do not implement recommendations, they are required to give reasons and outline alternative measures.[56]

Researchers should be given a framework for compelling access to data from very large online platforms.[57] Facebook, YouTube, and the like should provide data to researchers limited to identifying and understanding systematic risks. The Digital Services Coordinator and Commission may also require access to data. For example, to rule on the accuracy and functional specifics of algorithmic systems, or for content moderation, recommendation systems, or advertising systems.

Very large platforms must publish a comprehensive transparency report once a year on risk identification, risk-minimising measures, the audit report and the resulting adjustments. This obligation is in addition to the existing reporting obligation for all intermediaries under Art. 13 DSA.[58]

Chapter 5.d. Design specifications and architecture specifications

In large-scale socio-technical systems, the design (interface) and architecture (algorithms) also play a significant role.[59] These are central elements for influencing user engagement in the sense of the business model and for suggesting or facilitating certain decisions for users.[60] For this and other platform specifics, the Commission can provide guidance under certain conditions as part of its oversight. This is because the Commission has a broad set of tools at its disposal for supervision, investigation and enforcement. This means that the Commission can also intervene in the design and architecture of very large online platforms. For example, if an online platform fails to comply with the provisions of the Digital Services Act, the Commission can take interim measures,[61] declare commitments by very large online platforms to be binding,[62] and issue orders for non-com-

56 Art. 28 DSA.
57 Art. 31 sec. 2 DSA.
58 Art. 33 sec. 2 DSA.
59 Jeffrey Chan, "Ethics in large-scale socio-technical systems", in Laura Scherling and Andrew DeRosa (eds.): *Ethics in Design and Communication*, New York 2020.
60 Nir Eyal, *Hooked*, New York, 2019; Cliff Kuang and Robert Fabricant, *User Friendly*, New York, 2019.
61 Art. 55 DSA.
62 Art. 56 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

pliance.[63] If systemic risks are not effectively minimized, the Commission may, in cases of urgency due to the risk of serious harm to users, issue interim orders based on a *prima facie* finding of non-compliance. Although, these are to be limited in time. They may be extended. As interim injunctions, highly specific risk mitigation requirements can be imposed on platforms. The Commission can thus intervene directly in the (interface) design and (algorithm) architecture of online platforms. If, for example, it turns out that a systemic risk emanates from a certain algorithmic programming and the platform operator cannot get this under control, the Commission can issue concrete architectural specifications in this regard. Then, for example, reprogramming the weighting of algorithms could be specified. If it turns out that functions integrated into the design of the platform – such as an endless scroll – are prone to risk, direct design specifications can be made.

Chapter 5.e. Summary

Management of systemic risks is initially left to platforms through the assessment process with subsequent risk minimisation process. It is up to them to assess the risks in the designated fields and to make proposals as to how they can be minimised. However, the Commission does not have to stand idly by, but can intervene at all stages of this process. In addition, the audit promises to provide insights into the complex world of systemic risks posed by very large online platforms.

If stringent design or architectural requirements are imposed via interim injunctions, such requirements sometimes deeply interfere with the platform business model. However, the legal interests in question are all-important, so that interference with the fundamental economic rights of service providers can be justified. A complete ban on certain designs and architectures is also conceivable. It would not be surprising if technologies such as endless scrolling, auto-play, or other designs discussed under the term 'dark pattern' were prohibited for certain platforms and certain target groups that are particularly worthy of protection (such as children).

63 Art. 58 DSA.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 6. Market conduct rules for gatekeepers in the Digital Markets Act.

The market power of a few large technology groups is considerable. At the same time, platform markets have special features, such as lock-ins and network effects.[64] These first had to be understood by the regulatory authorities. In recent years, the EU Commission as well as national antitrust authorities have conducted several competition law proceedings against companies such as Apple, Microsoft, Google and Facebook. These companies were accused of obstruction and exploitation strategies, and very high fines were not infrequently imposed. The findings of these proceedings are now found as prohibitions and commandments in respect of certain behaviours in the market. The Digital Markets Act relies on *ex ante* regulation for these practices. Further orders are then not necessary for effectiveness. At the heart of the Digital Markets Act are the "obligations" in Art. 5 DMA and "obligations that may be further specified" enumerated in Art. 6 DMA.

Chapter 6.a. Rigid commandments and prohibitions

Art. 5 DMA contains rigid requirements and prohibitions for gatekeepers. There is no need for further concretisation in individual cases by the EU Commission. Accordingly, for gatekeepers the following is prohibited:

- merge personal data of different own services or services of third parties without a compliant consent according to General Data Protection Regulation (lit a),
- prevent commercial users from reporting matters related to gatekeeper practices to a competent authority (lit d),
- to require the use of its own identification service (lit e),
- make granting access dependent on a subscription or registration with another service (lit f).

Mandatory gatekeepers must

- enable commercial users to offer the same products or services to end users at different prices or conditions than through the gatekeeper's online intermediary services (lit b),

64 Philipp Staab, *Digitaler Kapitalismus*, Berlin 2019; Nick Srnicek, *Platform Capitalism*, London 2017.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

- enable commercial users to promote offers to end users acquired through the central platform service (lit c),
- and to conclude contracts with these end users via the gatekeeper's central platform services or by other means (lit c),
- and enable end users to access or use content, subscriptions, features or other elements by using a business user's software application through the gatekeeper's central platform services, if the end user has purchased such elements from the relevant business user without using the gatekeeper's central platform services (lit g).
- advertisers and publishers receive information about publication of a particular advertisement and for each of the gatekeeper's relevant advertising services (lit g).

Chapter 6.b. Other commandments and prohibitions

Article 6 DMA contains further requirements and prohibitions. The law states that these "may contain obligations of gatekeepers that are to be specified in more detail". However, this is not explained further in the Digital Markets Act. The following practices are prohibited for gatekeepers:

- to use non-publicly accessible data generated via the central platform service by commercial users in competition with such commercial users (lit a),
- give preference in ranking to services and products offered by the gatekeeper itself over similar services or products offered by third parties, and must carry out the ranking on the basis of fair and non-discriminatory conditions (lit d),
- refrain from technically limiting the possibilities to switch between different software applications and services (lit e),

In addition, a number of bids are set up. Gatekeepers must:

- enable end users to uninstall software applications preinstalled on its central platform service (lit b),
- enable the installation and effective use of third-party software applications and app stores that use or interoperate with gatekeeper operating systems (lit c), [65]

65 Gatekeeper may take reasonable steps to ensure that third party software applications or third party operated stores for software applications do not compromise the integrity of hardware or operating systems provided by the gatekeeper.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

- provide commercial users and ancillary service providers with access to and interoperability with operating systems, hardware or software functions for the provision of ancillary services (lit f),
- Provide advertisers and publishers, free of charge, with access to performance measurement and information they need to conduct their own independent review of advertising inventory (lit g),
- ensure effective portability of data generated by users and end-users and provide tools to facilitate data transfer and ensure permanent real-time access (lit h),
- provide commercial users, free of charge, with effective, high-quality and permanent real-time access to data provided or generated in connection with use of the relevant central platform services by such commercial users and end-users using the products or services of such commercial users (lit i),
- grant third parties operating online search engines access to ranking, search, click and display data relating to unpaid and paid search results at their request on fair, reasonable and non-discriminatory terms (lit j);
- apply fair and non-discriminatory general terms and conditions for commercial users' access to its app store (lit k).

Chapter 6.c. Enforcement of market rules for gatekeepers

Powers of investigation, enforcement and monitoring are regulated in detail. It is also possible for certain obligations to be suspended upon request or to be exempted from obligations for compelling reasons of public interest. Under Article 22 DMA, in urgent cases where there is a risk of serious and irreparable harm to commercial users or end users of gatekeepers, the Commission may order interim measures against a gatekeeper on the basis of an infringement of Article 5 DMA or Article 6 DMA. Fines are possible in the amount of up to 10% of annual turnover.

Both with the DMA and the DSA, the European Commission proposes to centralize the supervision of digital corporations' cross-border conduct in the Union in its own hands.[66]

66 Torsten Gerpott „Wer reguliert zukünftig Betreiber großer Online-Plattformen?“, Wirtschaft und Wettbewerb, No. 9 (2021).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 7. Conclusion

Joseph Vogl recently recommended a series of measures as a solution to 'infodemias' on the net: "Increase friction, reduce speed, insert cooling periods, extend pauses, increase signal noise, disrupt cycles, interrupt automatisms, shut down."[67] With the Digital Services Act and the Digital Markets Act and other regulations,[68] the European Commission is putting forward comprehensive proposals to regulate the digital economy.[69] In doing so, it is responding comprehensively to the threat to legal assets and fundamental rights posed by online platform business models. The focus is on very large platforms, for which an extensive catalogue of obligations is being drawn up. These must first implement a catalogue of rigid requirements for content moderation that applies regardless of the type of platform or business model. Airbnb, Uber, Facebook, and Amazon must then make the criteria of their content moderation transparent, maintain advertising databases and offer non-personalized recommendation systems. This also interferes with the business models. Users will also be protected by certain procedural rules, such as specific requirements, among them the obligation to provide reasons in the case of content deletion and the possibility to object. Platforms must protect their users from users who regularly disseminate illegal content by temporarily blocking such accounts. These basic rules also address the involvement of artificial intelligence in the process. For the most part, these requirements are formulated as minimum standards, which also allow platforms to apply stricter standards. However, any content moderation measures must respect the fundamental rights of users.

---

67 Julia Encke and Harald Staun "Die Nutzer spielen mit", Frankfurter Allgemeine Sonntagszeitung, March 14, 2021, https://www.faz.net/aktuell/feuilleton/debatten/plattformkapitalismus-joseph-vogl-ueber-kapital-und-ressentiment-17241098.html?printPagedArticle=true#pageIndex_2.

68 List of EU Regulatory Instruments on Digital Platforms see Annex to this Article; for European Artificial Intelligence Act see Jan Kalbhenn „Designvorgaben für Chatbots, Deepfakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung“, *ZUM – Zeitschrift für Urheber- und Medienrecht*, No. 8/9 (2021); for other Digital European Regulation see Boris Paal and Lea Kumkar „Die digitale Zukunft Europas“, *ZfDR – Zeitschrift für Digitalisierung und Recht*, No 2 (2021).

69 Regulation under telecommunications law as services of general interest could go even further, see Christoph Busch, *Regulierung Digitaler Plattformen als Infrastruktur der Daseinsvorsorge*, 2021.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Digital Services Act takes into account that online platforms cannot be lumped together. It makes a difference whether information and opinions are disseminated or goods are offered for sale on a very large online platform. Advertising-driven offerings also regularly pose different risks than those in which the individual conclusion of a contract is settled with commissions. Systemic risks of this kind are a complex matter that must be assessed differently from platform to platform. Correctly, the Digital Services Act relies on due diligence to address these risks.[70] In this regard, it is first in the hands of platforms to procure empiricism and identify risks. The right of initiative to mitigate risks also lies with the platforms themselves. If they fail to do so, the platforms are even given opportunities to make improvements. Only gradually – if the risks are not sufficiently minimized – does the sanctions regime take effect. It is then also possible to give platforms concrete specifications for the design and architecture of their platforms and to prescribe (interface) designs or (algorithm) architectures. The Digital Markets Act goes much further. As an *ultima ratio*, it provides for exclusion of a gatekeeper from the market.

Some commentators see the proposed regulatory regime as borrowing from financial market regulation. There, the listing of securities can be suspended if orderly trading is temporarily jeopardized or if this appears necessary to protect investors. These interventions in the free flow of market activity are known as 'circle breakers'. Such ad hoc interventions are not initially found in the repertoire of the Digital Services Act. Rather, incisive measures are only possible after a chain of misconduct. Like trading in financial products, the marketplace of opinions has become enormously automated and accelerated, especially on social networks.[71] In extreme cases of virally spread hatred, disinformation, and other content dangerous to weighty legal assets, a kind of 'circle breaker' could be considered, so that in extreme situations 'trading' would also have to be suspended on social media. This measure, which fits into the canon of measures recommended by Vogl ("Increase frictions, reduce speed, insert cool-down periods, extend pauses, increase signal noise, disrupt circuits, interrupt automatisms, shut down."), remains the responsibility of individual users and civil society.[72]

---

70 Lorna Woods and Bernd Holznagel, "Rechtsgüterschutz im Internet – Regulierung durch Sorgfaltspflichten in England und Deutschland", *Juristen Zeitung* No. 6 (March 19, 2021).

71 Armin Nassehi. *Muster*, Munich, 2019.

72 James William, *Stand out of our light: Freedom and resistance in the attention economy*, New York 2018; Jenny Odell, *How to do nothing: Resisting the Attention Econo-*

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Bibliography

Aral, Sinan. The Hype Machine: How Social Media Disrupts Our Elections, Our Economy and Our Health - and how we must adapt. London, 2020.

Bayer, Judit. "Double harm to voters: data-driven micro-targeting and democratic public discourse." Internet Policy Review, 9(1) (2020). https://doi.org/10.14763/2020.1.1460.

Bundeskartellamt. Case Summary, Facebook, Exploitative business terms pursuant to Section 19(1) GWB for inadequate data processing (February 15, 2019), https://www.bundeskartellamt.de/SharedDocs/Entscheidung/EN/Fallberichte/Missbrauchsaufsicht/2019/B6-22-16.pdf?__blob=publicationFile&v=4

Busch, Christoph. Regulating Digital Platforms as Infrastructure for Services of General Interest, 2021.

European Commission. Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, On the European democracy action plan. Brussels, 3.12.2020.

European Commission. Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, 2021, Brussels, 15.12.2020.

European Commission. Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act). Brussels, 15.12.2020.

Encke, Julia, Staun, Harald. "Die Nutzer spielen mit." Frankfurter Allgemeine Sonntagszeitung, March 14, 2021. https://www.faz.net/aktuell/feuilleton/debatten/plattformkapitalismus-joseph-vogl-ueber-kapital-und-ressentiment-17241098.html?printPagedArticle=true#pageIndex_2.

Eyal, Nir. Hooked: How to build habit forming products. London 2019.

Fielitz, Maik, Marcks, Hoger. Digitaler Faschismus: Die sozialen Medien als Motor des Rechtsextremismus. Berlin, 2020.

Gerpott, Torsten. „Wer reguliert zukünftig Betreiber großer Online-Plattformen?" Wirtschaft und Wettbewerb, No. 9 (2021): 481-487.

Gorwa, Robert, Binns, Reuben, Katzenbach, Christian. Algorithmic content moderation: technical and political challenges in the automation of platform governance. Big Data & Society, 2020.

Andreas Grünwald. "Big Tech-Regulierung zwischen GWB-Novelle und Digital Markets Act", *MMR – Zeitschrift für IT-Recht und Recht der Digitalisierung.* No. 12 (2020): 822-826.

Helberger, Natali. On the Democratic Role of News Recommenders. Digital Journalism 2019, 993-1012. DOI: 10.1080/21670811.2019.1623700.

---

my, London, 2019; Geert Lovink, *Sad by Design: On Platform Nihilism*, London, 2019.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Hwang, Tim. Subprime Attention Crisis. New York, 2020.

Kalbhenn, Jan and Hemmert-Halswick, Maximilian. „EU-weite Vorgaben für die Content-Moderation in sozialen Netzwerken." *ZUM – Zeitschrift für Urheber- und Medienrecht,* No. 3 (2021): 185-194.

Kalbhenn, Jan „Designvorgaben für Chatbots, Deepfakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung", *ZUM – Zeitschrift für Urheber- und Medienrecht,* No. 8/9 (2021) 663-674.

Klonick, Kate. The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression. The Yale Law Journal, 2021

Kuang, Cliff, Fabricant, Robert. User Friendly - How the hidden rules of design are changing the way we live, work, and play. Penguin 2019.

Lovink, Geert. Sad by Design: On Platform Nihilism. London, 2019.

Scherling, Laura, DeRosa, Andrew, Ethics in Design and Communication. London, New York, 2020.

Nassehi, Armin. Muster: Theorie der digitalen Gesellschaft. Munich, 2019.

Odell, Jenny. How to do nothing: Resisting the Attention Economy: London, 2019.

Paal, Boris, Kumkar, Lea ‚Die digitale Zukunft Europas',Zeitschrift für Digitalisierung und Recht, No 2 (2021): 97-131.

Staab, Philipp. Digitaler Kapitalismus: Markt und Herrschaft in der Ökonomie der Unknappheit. Berlin: Suhrkamp, 2019.

Srnicek, Nick. Platform Capitalsim. Cambridge 2017.

Vogl, Joseph. Kapial und Ressentiment. Berlin 2021.

William, James. Stand out of our light: Freedom and resistance in the attention economy: New York, 2018.

Woods, Lorna, Holznagel, Bernd 'Rechtsgüterschutz im Internet – Regulierung durch Sorgfaltspflichten in England und Deutschland' Juristen Zeitung No. 6 (March 19, 2021): 276-285.

Zuboff, Shoshana. The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs, 2019.

Annex: List of Europe's Digital Regulatory Instruments

- *e-Privacy Directive* (Directive 2002/58/EC) – July 12th, 2002
  - Aims at ensuring an equal level of protection of personal data processing, free movement of such data and of electronic commu-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

nication equipment and services in the community by setting out rules for providers of electronic communication services.[73]

- *General Data Protection Regulation* (GDPR; Regulation (EU) 2016/679) – April 27th, 2016
  - Sets out rules regarding personal data processing according to the principle of graduated regulation to ensure the protection of fundamental rights, in particular their right to protection of personal data.[74]
- *Code of Practice on Disinformation and related documents* – October 2018
  - Voluntary agreement signed by online platforms and advertisers as well as parts of the advertising industry that sets out self-regulatory standards to fight disinformation, monitor and improve online policies and ensure greater transparency and accountability.[75]
- *Audiovisual Media Services Directive* (AVMSD/ Directive (EU) 2018/1808) – Nov. 14th 2018
  - Directive amending *Directive 2010/13/EU* extends media law regulation to video-on-demand and video-sharing platforms such as YouTube, Netflix or Facebook: Tighter protection of minors, ban on inflammatory, violent and terrorist content, quota for European productions.[76]
- *Directive on Copyright in the Digital Single Market* (Directive (EU) 2019/790) – April 17th 2019
  - Includes new rules for fairer remuneration of creatives and rights holders, press publishers and journalists, especially when their works are used online, and increases transparency in their relationships with online platforms.[77]
- *Platform to Business Regulation* (P2B Regulation; Regulation (EU) 2019/1150) – June 20th, 2019
  - Aims at increasing fairness and transparency to business users of online intermediation services and corporate websites in relation to online search engines by imposing transparency requirements on

73 https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058
74 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434
75 https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/european-democracy-action-plan/strengthening-eu-code-practice-disinformation_en
76 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L1808&rid=9
77 https://eur-lex.europa.eu/eli/dir/2019/790/oj

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

those providers that are established or reside in the EU and offer goods or services to consumers located in the EU.[78]

- *Open Data Directive* (Directive (EU) 2019/1024) – June 20th 2019
  - Aims at making public sector and publicly funded data re-usable and introducing the concept of high-value dataset and applies to content held by museums, libraries and archives (written texts, databases, audio files and film fragments); not: educational, scientific and Open Data Directive.[79]
- *European strategy for data* (COM/2020/66 final) – February 19th, 2020
  - Aims at creating a single market for data allowing data sharing within the EU and across sectors benefiting businesses, researchers and public administrations.[80]
- *Data-governance Act* (COM/2020/767) – Nov. 25th 2020
  - Legislative proposal aiming at creating a framework that facilitates data-sharing and re-using of data laying down a voluntary registration framework for entities that collect and process data made available for altruistic purposes.[81]
- *European Democracy Action Plan* (COM/2020/790) – December 3rd, 2020
  - Aims at promoting democratic participation in free and fair elections, strengthen media freedom/pluralism and counter disinformation, foreign interference and information influence operations through legislative and non-legislative measures.[82]
- *Digital Services Act* (DSA; COM/2020/825 final) – Dec. 15th, 2020
  - Sets an accountability framework for online intermediary services/platforms to promote transparency, protect consumers and their online rights, and improve content moderation. Imposes different obligations for different categories of online intermediaries according to their role, size and impact online.[83]
  - Amendment to the e-Commerce Directive adopted in 2000.
- *Digital Markets Act* (DMA; COM/2020/842 final)– Dec. 15th, 2020

78 https://eur-lex.europa.eu/eli/dir/2019/790/oj
79 https://eur-lex.europa.eu/eli/dir/2019/1024/oj
80 https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
81 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767
82 https://ec.europa.eu/info/sites/default/files/edap_factsheet8.pdf
83 https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/689357/EPRS_BRI(2021)689357_EN.pdf

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

  - Sets criteria defining and prohibiting unfair practices by platforms that act as digital "gatekeepers" to the single market and provides market investigation-based enforcement mechanisms.[84]
- *Artificial Intelligence Act* (AI Regulation; COM/2021/206) – April 21st, 2021
  - Regulatory framework on the development, marketing and use of Artificial Intelligence that applies to providers of AI systems in the Union, users of AI systems located within the Union and providers and users of AI systems that are in a third country, where the output produced by the system is used in the Union.[85]

84 https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2347
85 https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Introducing the Systems Approach and the Statutory Duty of Care

Lorna Woods

Abstract: Early policy in relation to the internet framed questions from the perspective of liability for individual items of content. With the growth of social media, the approach struggles to deal with the scale of material as well as the contextual subjectivity of the acceptability of some types of content. This chapter explains a different approach, based on the work of Carnegie UK Trust, that moves away from direct content regulation to look at the services on which that content is created and disseminated. It argues that those services are not neutral as to that content, and that design choices can operate to create or exacerbate problems. The proposal is that of a risk managed approach to service development, aiming to achieve 'safety by design'. Although the orginal Carnegie proposal was based in English law, it is argued that the esssential elements of this approach could be deployed in other legal systems.

Keywords: duty of care – risk assessment – safety – choice architecture – design – online harms

Chapter 1. Introduction

Early policy-making in the context of the Internet saw the positives of the 'information society' and sought to minimise roadblocks on the 'information superhighway'. The legal framework dealing with 'intermediaries', which remains in place more than two decades later, aimed at removing disincentives to innovation in the sector.[1] A commonality between the EU and American approach was to protect intermediaries from exposure

1 Concerns about innovation remain – see e.g. D. Geradin, "Online Intermediation Platforms and Free Trade Principles: Some Reflections on the Uber Preliminary Ruling Case" in Ortiz (ed), *Internet: Competition and Regulation of Online Platforms*, (Competition Policy International, 2016).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to legal liability in respect of user content hosted or disseminated across their respective services, though the two regimes nonetheless differed in the scope of protection offered. Even by the early 2000's, when fewer people were online and less frequently so, concerns about abuse of the internet were starting to arise. Twenty years on, a wider range of threats are perceived, some arising from specific types of content for example hate speech, others from behaviours, including addiction. Pressure for regulatory action has grown, but much has focussed on dealing with individual items of content and the possibility of removing intermediaries' immunity. This chapter challenges that approach and proposes an alternative approach, based on work done under the aegis of the Carnegie UK Trust, what might be termed a systems-based approach and implemented – in the UK context – by a 'statutory duty of care'.[2] The elaboration of this approach, and the assumptions underpinning it, has the objective of identifying the key elements that could be deployed elsewhere, whether using the same or different implementing mechanisms.

Chapter 2. A Traditional Approach to Liability for Content

Policy in the field of communications, including the mass media, accepted a basic distinction between content creator (including publisher and curator) and those whose role was dissemination – for example, a telecommunications operator. This distinction can be seen, for example, in the development of the EU communications package,[3] though of course there have always been connections between content and network (see e.g. the position of electronic programme guides and the discussion of net neutral-

2 W. Perrin and L. Woods, *'Duty of Care' – Full Report*, April 2019, https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/ developing earlier work in support of a private members bill: https://bills.parliament.uk/bills/1877.

3 This distinction was also present in EU regulation on this issue and can now be found in the European Electronic Communications Code, Directive 2018/1972 , [2018] OJ l 321/36, rec 7; see also views of Court of Justice in Case C-518/11 *UPC Nederland*, judgment 7 November 2013, EU:C:2013:709, para 41; Case C-475/12 *UPC DTH*, judgment 30 April 2014, EU:C:2014:285, para 43; Case C-142/18 *Skype Communications Sarl* v *Institut belge des services postaux et des telecommunications (IBPT)*, judgment 5 June 2019, EU:C:2019:460, para 28. Helberger et al. also note this dichotomy in "Governing online platforms: from contested to cooperative responsibility" (2018) 34(1) *The Information Society* 1-14, p. 2.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ity).[4] A similar concern with the boundary between content creation and curation (ranging from commissioning content, via choices about scheduling and prominence through to ex post moderation) and its dissemination and the role of knowledge in determining the boundary between the two can be seen in the immunity provisions for "information society service" providers in the EU,[5] a distinction implemented in the UK and retained post Brexit. Neutral[6] intermediaries[7] (responsible for transmission, caching or hosting[8]) receive immunity on condition such an intermediary acts expeditiously to remove content once aware of its problematic nature under domestic law.[9] While this frame of analysis may seem appropriate for the transmission infrastructure or for other services that play a purely technical role in the dissemination of bits and bytes, it does not fit so well for some of the online platforms (a term which is only just recently beginning to be defined in legal terms), especially social media platforms which structure to a marked degree the content to which users are exposed. The extent to

---

4 The development of "information society services" (ISS) as a regulatory category blurs this boundary somewhat as they can be content services or more related to transmission; the regulatory response was to carve out some types of ISS from the general regime and treat them as similar to broadcast services: E. Dommering, "General Introduction", in Castendyk, Dommering and Scheuer (eds) *European Media Law* (Alphena/d Rijn: Kluwer Law International, 2008), para 10. See also text attached to n 5 et seq below.

5 Articles 12-14 e-Commerce Directive, Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market [2000] OJ L178/1

6 Case C-324/09 *L'Oreal v eBay*, [2011] ECR-I 6011 (Grand Chanber), para 124 and see para 122 for examples of activity that a diligent economic operator may engage in; the test of 'diligent economic operator' was applied by the Northern Irish Court of Appeal in *C.G.* v *Facebook Ireland Ltd* [2016] NICA 54, para 72.

7 This has been described as a 'catch-all term': J. Weaver 'Google IP Infringements: No results found?' (2018) 40 EIPR 759; see also M. Husovec, *Injunctions Against Intermediaries in the European Union: Accountable but not liable?* (Cambridge: Cambridge University Press, 2017), 16-17.

8 Originally these phrased were included in Articles 12-14 e-Commerce Directive, but definitions have been expanded in the Proposal for a Regulation on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC (COM/2020/825 final), 15 December 2020; overview of services in scope provided by, for example, A. Vijay, "Liability of internet service providers – a review study from the European perspective" (2019) 41 *EIPR* 451; D. Fernàndez, "ISP Liability Between EU and USA" (2016) 17 *Computer Law Review International* 36.

9 What this means has not yet been fully harmonised: see e.g. Husovec (n 7), pp 52-57.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

which those platforms could be said to have knowledge of this content though is open to debate; while the platform processes influence what users see, much of this process is automated.[10]

There has been increasing concern about the availability and prevalence of certain types of content on the Internet, specifically on social media platforms. Concerns about child sexual abuse and exploitation material as well as terrorist content have been a subject of concern since the early 2000's but there are now a wider range of concerns.[11] Solutions have considered making the take-down of content more effective (and solutions in this field would clearly be useful); some have suggested that immunity be removed.[12] Focussing a regulatory regime aimed at platforms on the content they host is, however, problematic. While platforms may prompt or promote certain types of content, they do not create it or commission it; they are not responsible for it in the same way as those that create or reuse that content. Moreover, the size of some of the platforms is in itself an issue; so much content is uploaded (which brings issues of speed as well as of scale) that it would be hard to consider items of content individually (and automated techniques bring their own issues). Moreover, the range of types of content and their audiences are wide and diverse with different expectations in relation to those different types of content. The assessment of the acceptability of items of content is to a large degree context specific. While countries will vary as to their tolerance for certain types of content, speech may be understood differently within those countries or by sub-groups within those countries. Ofcom noted some of these problems given that 'the internet is fundamentally different from television and radio in its nature, audience and scale'.[13] Moreover, this is an area in which there is not only variety in service type but also frequent innovation. Any approach

10 See e.g. Vijay (n 8), p.454; T. Gillespie, *Custodians of the Internet* (New Haven/London: Yale University Press, 2018), p. 7.

11 See issues identified in DCMS, Internet Safety Strategy – Green Paper, October 2017, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf.

12 Committee on Standards in Public Life, Intimidation in Public Life: A Review by the Committee on Standards in Public Life (Cm 9543), December 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/666927/6.3637_CO_v6_061217_Web3.1__2_.pdf.

13 Sharon White, "Tackling Online Harm – a regulator's perspective", speech by Sharon White to the Royal Television Society, 18 September 2018, https://www.ofcom.org.uk/about-ofcom/latest/media/speeches/2018/tackling-online-harm (accessed 18 March 2021); see also OFCOM, *Discussion Document: Addressing Harmful Content Online*, p.25.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to regulation would need, therefore, to be to some degree future-proof. As the Interim Report of the DCMS Select Committee fake news inquiry recognised, what is needed is an approach that recognises a 'third way', one that is not dependent on a simplistic division between content on transmission.[14]

Chapter 3. A Different Model

Thinking about social media platforms as quasi-publishers limits the possible policy responses. A different analogy may give rise to different policy options and a return to the language of the 1990's – to "cyberspace"[15] (the virtual world created by the links between computers) – may provide a hint as to where to look for alternative inspiration. The range of services provided across the Internet is wide and may be used differently by different groups; these services provide the place for lots of different activities to happen on-line as take place in a range of spaces off-line. They provide a mechanism for users to engage with one another, to be entertained, to discover information, to advertise and to buy and sell. In the off-line context, providers of spaces are not necessarily regulated in relation to what happens in that place (though some may be – e.g. pubs, casinos, sandwich shops) but they each have some responsibility for the safety of the place, a responsibility which is often dealt with through an assessment of hazards and risks and the likelihoods of harm arising to users of the space. Space management also communicates different expectations as to user behaviour in those spaces. This then leads us to the position that, rather than imposing liability on platforms for individual items of content, they should be expected to assess their respective platforms for safety of their users, and others affected by the service, taking into account how those platforms are used. In moving away from content-focussed regulations, the difficulties in dealing with different understandings about the meaning and acceptability of certain types of content in different jurisdictions, as well as issues arising from scale, may be ameliorated.

---

14 DCMS Select Committee, *Disinformation and 'fake news': Interim Report* (Fifth Report of Session 2017–19), 24 July 2018 (HC 363).

15 The term is derived from William Gibson's novel *Neuromancer* (Victor Gollancz, 1984).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

One might ask, however, what harm may arise from a platform apart from the content itself? The inherent constraints which are found in the physical world do not operate online, and this has allowed the introduction of sophisticated choice architectures aimed at maximising user interaction. This is not necessarily bad, but nor is it neutral – especially when we compare people's interactions online with those offline. It has long been noted that people speaking online experience a disinhibition effect[16] (though the causes are not yet fully understood). Given that users' online experience is mediated by the platforms, the design of the platform could seek to compensate for this; to remind users that others using social media are (in the main) humans too.[17] However, the motivating objective in platform design seems to have been the support of the service providers' bottom line, regardless of consequence. Designing to maximise user engagement for the purpose of acquiring data and delivering adverts, it seems the platforms rather seek to exploit our cognitive weaknesses.[18] So, while a 'like button' can be used as a substitute for nonverbal cues that might be otherwise absent and be seen by the user as a signal of appreciation, for the platforms it is data the accumulation of which can be exploited to understand much more about users than those users may appreciate. A range of adverse consequences has arisen, which some have linked back to design choices, and which risk endangering the well-being of individuals and the functioning of democratic societies: cyber-bullying and hate speech; the polarisation of public debate and the rapid spread of false (and

16 J. Suler, 'The Online disinhibition effect" (2004) 7(3) *Cyberpsychol Behav* 321-6, doi:10.1089/109493104129295.

17 Work on tools and techniques for this is starting in some areas: see e.g. the Prosocial Design Network which lists features and the prosocial consequences they might have and seeks to test them, https://www.prosocialdesign.org/.

18 S. Zuboff, *The Age of Surveillance Capitalism– The Fight for a Human Future at the New Frontier of Power*, (1st ed) (Profile Publishers: London, 2019); in an earlier article she describes "a ubiquitous networked institutional regime that records, modifies, and commodifies everyday experience from toasters to bodies, communication to thought, all with a view to establishing new pathways to monetization and profit" "Big other: surveillance capitalism and the prospects of an information civilization" (2015) 30 *Journal of Information Technology* 75-89, p. 81.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

harmful) information.[19] Commentators have pointed to the dangers of content creators responding to the metrics provided by many platforms, whether to sell products themselves (including influencers) or to chase the feel-good glow and being 'liked' – and users are thereby trained to produce response-creating content[20]. Others note that the tools provided for promoting content, aimed at driving user engagement and in effect operating as a trap,[21] priorise extreme, violent and shocking content – that which engages strong negative emotions – with the risk that, for example, conspiracy theories are promoted.[22] Similarly, lies travel faster than the truth (though whether lies are believed is another question);[23] misinformation may thrive because off-line epistemic cues and gatekeeper controls are absent, or because users are nudged to respond and to share or are distracted from considering accuracy.[24] The way information is presented may affect user behaviour: Facebook ran an experiment on its users' newsfeeds that suggested that including social information in an "I voted" button (in this case, displaying faces of friends who had clicked on the button) affected both click rates and real-world voting.[25] Targeted advertising, based on who knows what grounds, raise questions about not

19 S. Bradshaw and P. N. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation* (Working Paper 2019.2: Project on Computational Propaganda) (Oxford, 2019).

20 W. J. Brady et al., "How Social Learnings Amplifies Moral Outrage Expression in Online Social Networks" (2021) (paper under review, available: https://psyarxiv.com/gf7t5/).

21 Anthropological research suggests that those coding recommender algorithms see their function as 'hooking' users; that these algorithms operate as a trap: N. Seaver, "Captivating algorithms: Recommender systems as traps" (2018) Journal of Material Culture, https://journals.sagepub.com/doi/10.1177/1359183518820366.

22 E. Hussein et al., "Measuring misinformation in video search platforms: An audit study on YouTube" (2020) Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1), Article 48. doi 10.1145/3392854.

23 See J. Allen et al., "Evaluating the fake news problem at the scale of the information ecosystem" (2020) 6(14) *Sci Adv* eaay3539, doi: 10.1126/sciadv.aay3539 ; Kozyreva et al., "Citizens versus the Internet: Confronting Digital Challenges with Cognitive Tools" (2020) 21(3) *Psychol Sci Public Interest*, 103-156, doi: 10.1177/1529100620946707.

24 G. Pennycook et al., "Shifting attention to accuracy can reduce misinformation online" (2021) *Nature*, 17 March 2021, https://doi.org/10.1038/s41586-021-03344-2.

25 Kozyreva (n 23).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

just manipulation[26] but also intrusion into our respective *fora internum*.[27] Concerns have long been raised about 'filter bubbles' but more generally about the range of topics of information users receive.[28] It has also been suggested that the very short-form format of news based on headlines and snippets gives users the illusion of being informed.[29] Targeting may be weaponised by nefarious actors.[30] 'Sock puppet accounts' and networks of coordinated accounts may spread and embed false information and sow discord. While users are not just passive recipients in the online environment,[31] and users may innovate and disrupt at least some of the time, it must be recognised that not everybody has the capability to hack the system. As Leiser notes, some of the theoretical models in this area have fallen into a common trap: that of assuming that all users are rational and fully informed; and underplaying the role of cognitive weaknesses most humans exhibit.[32] Additionally, the tools provided to users to take control

---

26 S. Matz et al., "Psychological targeting as an effective approach to digital mass persuasion" (2017) 114(48) *Proc Natl Acad Sci USA* 12714, doi: 10.1073/pnas.1710966114.

27 S. Alegre, "Rethinking the Right to Freedom of Thought in the 21st Century" (2017) 3 *Eur. Hum. Rights. Rev* 221; S. Zuboff (n. 11); S. Alegre, "Regulating around Freedom I the "forum internum"" (2021) *ERA Forum* 591.

28 C. Sunstein, "Republic.com 2.0", p. 5; in *#Republic: Divided Democracy in the Age of Social Medi*a (Princeton, NJ, USA, and Oxford, UK: Princeton University Press, 2017), Sunstein also notes 'asymmetrical updating', that is a strong tendency to favour evidence that confirms our beliefs and ignore or misread evidence that does not. How to compensate for this does not seem to be a simple matter of ensuring more diverse viewpoints are presented. While some studies (e.g. Bakshy et al., "Exposure to ideologically diverse news and opinion on Facebook" (2015) 348 *Science* 1130, DOI 1-.1126/science.aaa1160) suggest that user choice may be part of this, others have suggested that algorithmic amplification has a role to play through the creation of a variant of feedback loop: A. J. B. Chaney et al., "How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility" (2018) RecSys '18, October 2–7, https://arxiv.org/pdf/1710.11214.pdf.

29 S. Schäfer, "Illusion of knowledge through Facebook news? Effects of snack news in a news feed on perceived knowledge, attitude strength, and willingness for discussions" (2020) 103 *Computers in Human Behavior* 1–12. 10.1016/j.chb.2019.08.031.

30 See concerns expressed by the DCMS Select Committee, *Disinformation and 'fake news': Final Report*, 18 February 2019, https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179102.htm.

31 A. Murray, *Regulation of Cyberspace*, (2007, Oxford University Press).

32 M. Leiser, "The Problem with 'Dots': questioning the role of rationality in the online environment" (2016) 30 *International Review of Law, Computers and Technology* 191.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of their own environment are not extensive and may not be easy to use nor recognise specific risks or problems faced by particular groups.

It is the fact that these choices lie in the hands of the operators meaning that placing responsibility on the operators for the design and operation of their respective platforms is legitimate; they are being held responsible for their own actions, not those of others. The designers are the risk-creators and thus best-place to manage those risks.[33] While not all the possible issues are fully understood, platform operators can still ask themselves the question how this service is working; is there evidence that there might be side effects; what content and safety curation tools can we provide (especially considering some groups may have particular needs); and what the alternative to a given feature is? Perhaps all inventors and designers should ask themselves, 'what happens when this scales and what happens when the bad people get hold of it?' In this, the approach looks at features and user behaviours and their likely impacts at a general level, not assessing individual items of content.

Chapter 5. Risk Assessment: A Model from Work Spaces

If we think of social media platforms as quasi-public spaces, the regulation ensuring those spaces are safe may constitute a model for the implementation of the system-based approach. In the UK, the main mechanism is found in the Health and Safety at Work Act 1974 (HSWA).[34] It provides a statutory duty of care – that is a duty of care similar to that found in the tortious doctrine of negligence – but specified (and possibly amended) by the terms of legislation. Section 2(1) HSWA states:

> It shall be the duty of every employer to ensure, so far as is reasonably practicable, the health, safety and welfare at work of all his employees.

This is a very broad category and s. 3(1) extends the duty beyond the employer's duty to employees to include "persons not in his employment who may be affected" by the business. The Act also imposes reciprocal duties on the employees.

---

33 Robens Report: Safety and Health at Work, July 1972 (Cmnd 5034).

34 For the development of the statutory duty of care and its difference from the duty of care found in the common law doctrine of negligence see L. Woods, "The duty of care in the Online Harms White Paper" (2019) 11(1) *Journal of Media Law* 6.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

While the nature of the obligation is broad – in the case of the duty to employees, it is to prevent harm and as regards others it is avoidance of exposure to risks to their health or safety – the HSWA gives examples of specific issues about which the employee must take action. Examples include: provision of machinery that is safe; the training of relevant individuals; and the maintenance of a safe working environment. This list of actions does not replace the general duty. The HSWA additionally contains an obligation on an employer "to prepare and as often as may be appropriate revise a written statement of his general policy with respect to the health and safety at work": this is the beginnings of formalising a preventative approach, based on an assessment of risks posed.

The regime is enforced by a regulator, the Health and Safety Executive, which has a range of powers including "improvement notices", "prohibition notices" and prosecution. Recourse to the criminal law is a matter of last resort and sentencing guidelines identify factors that influence the heaviness of the penalty. Factors that tend towards high penalties include flagrant disregard of the law, failing to adopt measures that are recognised standards, failing to respond to concerns, or to change/review systems following a prior incident as well as serious or systematic failure within the organisation to address risk. So, while the duty of care is still described as being owed to a certain group of people (employees in s. 2(1) and persons "affected by an undertaking" in s. 3(1)), general enforcement powers lie elsewhere. Individuals suffering injury are not empowered to bring action under this regime; injury suffered is dealt with through traditional negligence claims. This point highlights the difference between individual instances of harm and the environment giving rise to the risk of harm.

There are a number of points which suggest that an over-arching duty such as that found in HSWA is an appropriate model. It applies widely and in a range of different sorts of contexts; it applies to almost all employers and the myriad activities that go on in them. A similar tool could presumably be deployed across social media and the many purposes for and ways in which those platforms are used. A factor in the general duty's usefulness is the fact that, with the exception of a limited number of high risk activities which are controlled by specific regulations[35], it does not set down detailed rules with regards to what must be done in each workplace. It rather sets out some general duties that employers have both as regards their employees and the general public, but leaves the employer

35 For example, see Control of Major Accident Hazards Regulations 1999 (SI 1999/743).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to identify appropriate implementation mechanisms. This allows the employer's obligation to be tailored to the specific risks found in a particular (work) environment, subject to the guidance from the regulator. As well as providing for flexibility within the current range of providers, it allows a certain degree of future-proofing as new features, services or problems are introduced. It also allows for new research on understanding risks and how to mitigate against them to be taken into account as that body of research develops. An outcome orientated approach, which implies that an employer should seek to identify steps that would be reasonably effective in the relevant context, also mitigates the risk of a tick box approach were specific, detailed rules (e.g. ban bots; prohibit anonymous accounts) to be adopted. Finally, the distinction between the environment creating the risk of harm and the individual instances of harm broadly parallels the distinction between the systems constituting the platform/service and the individual instances of content or behaviour.

Chapter 6. The Statutory Duty of Care: A Proposal

This leads us to system-based regulation, where 'system' is understood in two ways:

- the focus of regulation is on the software system (or more broadly the service, including the business model) itself rather than on the content hosted on the service; and
- providers of such services should have a system (understood as a process) in place to risk assess the service and individual features of the service – and to take appropriate steps to address concerns arising.

The operator of the system should be subject to an overarching, general duty of care. The duty of care must set out the persons to whom the duty is owed,[36] the types of harm from which that person should be protected as well as the operators within scope.

As regards the first point, the Carnegie proposal suggested that both users and non-users of a service were owed a duty of care, provided that non-users were affected by the operation of the platform. In this, it followed the model of the HSWA. The reasoning was that persons could be

36 Note the Environmental Protection Act 1990 uses a similar mechanism but does not identify the beneficiary of the duty.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

harmed by behaviours on a platform even if they had not joined it, for example in the case of "revenge porn".

The proposal also noted that it was important that the types of harm be identified in statute,[37] but that the vectors of harm may be elaborated in regulatory guidance (especially in the light of developing research). Although the types of harm need some clarification, these can be reasonably broad categories, as the HSWA demonstrates; regulatory guidance can fill in the details. These categories of harm should be identified by reference to the impact on the victim, not by reference to whether the speech might be considered illegal or not.[38] The criminal law is not always the best proxy for understanding harm and, crucially, also does not focus on the role of the platform itself in encouraging, facilitating or exacerbating the occurrence of harm. As noted above, it is the fact that the platforms are risk creators that justifies the decision to regulate at this point.

The Carnegie proposal sought to define social media, on the basis of the following characteristics – that services:

- have a strong two-way or multiway communications component;
- display user-generated content;[39]
- publicly or to a large member/user audience or group.

This could include some private messaging apps that allowed large groups to communicate. Search engines were excluded because, although they have an effect on the information provided to users, they may give rise to issues surrounding the right to information, prominence and diversity which may necessitate a different response. Also excluded were actors,

37 See similarly Digital, Culture, Media and Sport Select Committee, *Disinformation and 'fake news': Final Report*, Eighth Report of Session 2017-19 (HC 1791), 18th February 2019, paras 31-32; in other sectors, e.g. broadcasting as well as the HSWA, regulators are entrusted with understanding the precise meaning of harm.

38 In this, the proposal differs from the characteristics that Cole, Etteldorf and Ullrich ascribe to duty of care models: Cole, Etteldorf and Ullrich, *Cross-border Dissemination of Online Content* (Baden-Baden: Nomos Verlagsgesellschaft, 2020), p 202 which limits risk assessment to illegal content and behaviours.

39 The Audiovisual Media Services refers to use-generated content and contains a definition of "user generated video"; the UK implementation of this provision does not use the same terminology. For discussion of the difficulties with the definitions in the Audiovisual Media Services Directive see L. Woods, "Video-sharing platforms in the revised Audiovisual Media Services Directive" (2018) 23 (3) *Communications Law* 127.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

essentially the broadcast and print media that are already subject to regulatory or self-regulatory regimes.

The essential element of this model is a risk assessment considering the service, including its individual features, and the business model of the service. The focus of enquiry is the impact the structures and business choices have in creating a risky environment. The system-based approach is neutral as to the topics of content (though part of that system will involve dealing with complaints and with content that is contrary to the law); as such, the system may be less open to the accusation that regulation will result in excessive take-down on unclear bases[40].

Risk assessments require the identification of hazards (that is something that could cause harm) and determine how likely it is that each hazard will occur and how severe the consequences would be. A risk assessment should take into account relevant human rights. Freedom of expression is obviously important but it is not the only right. Moreover, design choices may have discriminatory effects in the enjoyment of rights (the use of AI in content moderation is one example). The assessment of consequences operates at a general level rather than seeking to determine outcomes in particular cases. In this there is a difference from a regime aimed at compensating individual victims. The starting point is the platform and the likely consequences of its use; it is not about starting with an instance of harm or a category of content and trying to work backwards in respect of that particular example. As a final stage, the operator should determine the appropriate mitigating steps – whether this be not to deploy the new feature/change, to amend it, or to bring in some compensating measure. At the least, the operator should perform risk assessment before introducing new processes or activities, before introducing changes to existing processes or activities (such as a significant change to an algorithm), or when the company identifies a new hazard (e.g. becomes aware of research); it should also monitor whether the mitigating steps seem to be effective. This process was described as instituting a harm reduction cycle. We envisaged that a regulator would have some say in identifying what a good risk assessment looks like, but for risky services (including large services), the Carnegie proposal also envisaged some involvement of relevant civil society actors. In this, transparency at some level of granularity and within a framework set by the regulator, is key.

The duty is not focussed on particular technologies or the problems they cause. It allows a platform to take into account the interplay of

40 Cole et al (n 38) note this criticism, p. 204.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

different features in terms of risk assessment and mitigation. It is also not limited to technical specifications, but may take into account when, how and to whom features or services are deployed.[41] As HSWA illustrates, the fact that the statutory duty of care is a general obligation does not mean that statute cannot specify specific obligations within that general duty – for example, the need to have an effective complaints mechanism, obligations of transparency for particular issues, the need to take particular steps with regard to specific types of content (e.g. child sexual abuse and exploitation material).

In carrying out their duty of care, platform operators are not expected to achieve perfection. An appropriate threshold is similar to that found in the doctrine of negligence; it is not a strict liability regime. Rather, an operator should take reasonable steps in relation to foreseeable harms. Whether an operator has satisfied the duty will be determined by the regulator; jurisprudence from the doctrine of negligence is not binding in this regard.[42] "Reasonable" and "foreseeable" should take into account the platform's use, including its user base size and profile, as well as any relevant industry standards. While the service provider may not engage in wilful blindness, nor should they be judged with the benefit of hindsight. "Reasonable steps" do not require a perfectly sanitised environment; rather the requirement aims to consider the role platforms play in creating or exacerbating the problems. Moreover, the mere fact that there problematic content or behaviours may be found on a platform does not in and of itself constitute a violation of the duty of care. Ultimately, while the regime is orientated towards a particular result, the question of whether an operator has satisfied its duty of care is not answered by numbers of take-downs nor numbers of problematic posts/instances of use (though a platform on which there are many instances of problematic content may be less likely to have satisfied the duty of care). Liability is about engagement which the risk assessment and mitigation process; it does not involve liability for content.

As a result of the focus on design, the tools and changes are not limited to ensuring that a take-down regime operates effectively and fairly, though it should do that. There are three main points of influence before we reach the question of whether content should be taken down: the point at which a user engages with the platform (including sign up processes, means of

41 In this it is different from proposals which focus on a specific technology or technical standard, outlined Cole et al (n 38), p. 202.

42 Clerk and Lindsall on Torts (23rd ed) para 8-56.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

finding others in a group, and tools to communicate for example augmented reality filters/overlays); the mechanisms by which content is disseminated (e.g. search engines, hashtags, recommender systems, newsfeeds); and the mechanisms by which recipient users engage with content, including choosing not to engage with it, but also mechanisms such as tools for sharing/forwarding/demonstrating approval or disapproval. Examples of this category include retweeting, liking, forwarding tools, as well as those allowing users to block or mute incoming messages. Each of these points may have an impact on the content available – in terms the content created as well as the way content flows across platforms. Significantly, as many interventions allow speech to continue, they may be less intrusive to users' freedom of expression.[43]

Insofar as platforms operate as advertising services, the duty of care should extend to this aspect of the service too, with regard to protecting users.[44] Questions that might be asked include whether the platform engages in any KYC ("know your client") processes as regards advertisers; and what sorts of ads does it permit – do any require specific safeguards? Further, how are audiences segmented (e.g. what controls are there around permitted groupings/topics – are any segments impermissible or undesirable)? The availability of micro-targeting itself should be assessed for its risks.

The last port of call is take-down. An operator needs to ensure that it has an adequate complaints mechanism that is accessible and easy to use and which operates in a fair, timely and transparent manner.[45] As well as reporting on numbers and speed of take-down, reporting should consider what is being taken down, and why, as well as categories of complainant (with the intention of not only identifying where unforeseen problems arise, but also identifying and mitigating against discrimination in the complaints system).

---

43 For a consideration of the issues and some of the difficulties surrounding this analysis in the context of the Carnegie UK Trust proposal, see L. Woods, "The Carnegie Statutory Duty of Care and Fundamental Freedoms", 2019, https://www.carnegieuktrust.org.uk/publications/doc-fundamental-freedoms/.

44 This viewpoint was adopted by the Centre for Data Ethics and Innovation in its recommendations to Government: CDEI, *Review of Online Targeting*, 4 February 2020, https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations.

45 The House of Lords Communications Committee noted the need for consistent enforcement as well as transparency of complaints handling: *Growing up with the Internet* (2nd Report of Session 2016–17) (HL Paper 130), 21 March 2017, paras 241-2.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Carnegie proposal also envisaged that platforms should develop a triage process for emergent problems; while the detail of the problem may be unknown, it is fairly certain that new problems will arise, as the issue of misinformation and disinformation related to Covid-19 illustrates. The interface with law enforcement and relevant regulatory authorities (e.g. Advertising Standards Authority, Financial Conduct Authority) in the exercise of their powers should also be considered.[46]

The increasingly problematic nature of the social media environment suggests that self-regulation (even self-regulation engaging with voluntary codes of practice) has not worked well. Moreover, reliance on users to take action before the courts is unlikely to constitute a sufficient corrective for a range of reasons but notably because of the asymmetry of resources and knowledge between the major platforms and litigants. A regulator is required, even if the proposed scheme is not a traditional top-down command scheme. It is crucial, especially given the importance of freedom of expression in the functioning of a democracy, that the regulatory be independent from both industry and from government. It must make decisions based on objective evidence (and not under pressure from other interests) and be viewed as a credible regulator by the public. Independence means that it must have sufficient resources, as well as relevant expertise. A completely new regulator created by statute would take some years before it was operational. The Carnegie proposal therefore envisaged extending the powers of the existing telecommunications and media regulator, Ofcom. This approach has a number of advantages. It spreads the regulator's overheads further, draws upon existing expertise within the regulator (both in terms of process and substantive knowledge) and allows a faster start.

The responsibilities of the regulator would include identifying actors in scope; developing good practice and guidance about harms and vectors by which harm could be caused (including where appropriate approving industry codes of practice and standards); monitoring the harm reduction cycle and risk assessment processes; and enforcing the duty of care. The

46 The obligations on platforms to cooperate have arising in the enforcement of intellectual property rights, especially in connection with loss of immunity; see e.g. Husovec (n 7). Cooperation with regulatory authorities and law enforcement has drawn less attention, but see e.g. mechanisms envisaged by the recently agreed Regulation on addressing the dissemination of terrorist content online: Regulation 2021/784 [2021] OJ L172/79. As part of the Carnegie Proposal a model was proposed: see W. Perrin and L. Woods, "Online Harms – Interlocking Regulation" (Blog), 11 September 2020, https://www.carnegieuktrust.org.uk/blog/online-harms-interlocking-regulation/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Carnegie proposal also included information gathering powers for the regulator.[47] As in many other regulatory fields, failure to comply should be a violation of the regime in and of itself.

Finally, the regime must have sanctions, though any enforcement action should be context specific and proportionate, especially given the fundamental rights in play (including but not limited to freedom of expression). The range of mechanisms available within the HSWA are interesting because they allow the regulator to try improve conditions rather than just punish the operator; to some extent the GDPR and the Data Protection Act 2018 have a similar approach. Other options include adverse publicity orders where the operator is required to display a message on its screen most visible to all users detailing its offence which could result in reputational losses.[48] Another possibility, albeit one that would require some thought in terms of implementation, is borrowing techniques from restorative justice.[49] For those that will not comply, the regulator should be empowered to impose fines, including GDPR or competition policy magnitude fines. The more difficult questions relate to what to do in extreme cases. Should there be a power to send a social media services company director to prison (as in the HSWA) or to turn off the service? The Digital Economy Act 2017 (DEA) contains power[50] (which was never brought into force) for the age verification regulator to issue a notice to internet service providers to block a website in the UK. Blocking orders, even if technically effective, raise concerns about 'collateral censorship' – where a platform is blocked the speech rights of the platform's users are affected. This is particularly the case where there are large platforms carrying many different types of content (most of which would be unproblematic). These sorts of mechanisms – as well as criminal sanctions for speech – raise questions about their proportionality from a freedom of expression perspective. The DEA provided what could be a middle ground, though again this provision has not been brought into force. Section 21 empowers

47 On the importance of evidence gathering powers, see the evidence of Sharon White to the DCMS Select Committee, *Disinformation and 'fake news': Final Report*, (Eighth Report of Session 2017-19) (HC 1791), 18 February 2019, para 33.

48 On the effectiveness of mechanisms leading to reputational loss see e.g. Armour et al., "Regulatory Sanctions and Reputational Damage in Financial Markets" (2017) 52(4) *Journal and Financial and Quantitative Analysis* 1429 – 1448.

49 Restorative justice is used in the context of criminal justice in England and Wales; see here for CPS guidance: https://www.cps.gov.uk/legal-guidance/restorative-justice.

50 Section 23 Digital Economy Act 2017.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the regulator to issue notices to others who are dealing with the non-complying operator, such as credit card or other payment services. According to the Explanatory Memorandum to the DEA, the purpose of such a notice is to bring the problem to the attention of these ancillary service providers so as "to enable them to consider whether to withdraw services",[51] thus disrupting the provision of the service. This approach might be deemed problematic in that it uses private actors as enforcement mechanisms,[52] though it should be noted that similar techniques have been used in other regulatory contexts (e.g. cinemas were used as enforcement mechanisms for age ratings for films).

Chapter 7. Conclusion

This paper has sought to distinguish between two models of regulation in respect of social media: that aimed at content, which has been traditionally used in the context of speech concerns and specifically in relation to the mass media; and systemic regulation, which takes a process-based risk assessment approach to regulation used in many industrial sectors. Drawing on insights about the impact of design and choice architecture on user freedom and behaviour, and based on the work of Carnegie UK Trust, it has argued for the target of regulation to be the software and business systems that make up social media services. Not only do these systems have an impact on user behaviour but choices about the design and deployment of such systems are under control of the relevant companies. Looking to the UK legal environment, Carnegie UK Trust proposed a particular vehicle by which systemic regulation could be deployed: the statutory duty of care to create a general obligation enforced by a regulator rather than ex post individual litigation. While the statutory duty of care as a vehicle

51 Explanatory Memorandum to the Draft Online Pornography (Commercial Basis) Regulations 2018, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/749750/Explanatory_Memorandum_to_the_Draft_Online_Pornography__Commercial_Basis__Regulations_2018.pdf

52 The DEA did not impose penalties on those which did not cooperate; in this it might be different from the context of intermediaries in intellectual property. More generally see M. MacCarthy, "What Payment Intermediaries are Doing about Online Liability and Why it Matters" (2010) 25 *Berkley Technology Law Journal* 1037, especially p 1056.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to implement this model may be particular to the UK, the underlying regulatory model could be deployed in other jurisdictions.

Bibliography

Alegre, S. "Rethinking the Right to Freedom of Thought in the 21st Century". (2017) 3 *Eur. Hum. Rights. Rev* 221.

Alegre, S. "Regulating around Freedom I the 'forum internum'". (2021) *ERA Forum* 591.

Allen, J. et al. "Evaluating the fake news problem at the scale of the information ecosystem". (2020) 6(14) *Sci Adv* eaay3539, doi: 10.1126/sciadv.aay3539A.

Armour et al. "Regulatory Sanctions and Reputational Damage in Financial Markets." (2017) 52(4) *Journal and Financial and Quantitative Analysis* 1429 – 1448.

Bakshy et al. "Exposure to ideologically diverse news and opinion on Facebook". (2015) 348 *Science* 1130, DOI 1-.1126/science.aaa1160.

Bradshaw, S and Howard, P. N. *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation.* (Working Paper 2019.2: Project on Computational Propaganda) (Oxford, 2019).

Brady, W. J. et al. "How Social Learnings Amplifies Moral Outrage Expression in Online Social Networks" (2021). (paper under review, https://psyarxiv.com/gf7t5/).

CDEI. *Review of Online Targeting*. 4 February 2020. https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations.

Chaney, A. J. B. et al. "How Algorithmic Confounding in Recommendation Systems Increases Homogeneity and Decreases Utility." (2018) RecSys '18, October 2–7. https://arxiv.org/pdf/1710.11214.pdf.

Cole, Etteldorf and Ullrich. *Cross-border Dissemination of Online Content.* Baden-Baden: Nomos Verlagsgesellschaft, 2020.

Committee on Standards in Public Life. Intimidation in Public Life: A Review by the Committee on Standards in Public Life (Cm 9543). December 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/666927/6.3637_CO_v6_061217_Web3.1__2_.pdf.

DCMS. Internet Safety Strategy – Green Paper. October 2017. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf.

DCMS Select Committee. *Disinformation and 'fake news': Interim Report* (Fifth Report of Session 2017-19). 24 July 2018 (HC 363).

DCMS Select Committee. *Disinformation and 'fake news': Final Report***.** Eighth Report of Session 2017- 19 (HC 1791). 18 February 2019. https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-report-published-17-19/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Dommering, E. "General Introduction". In: Castendyk, Dommering and Scheuer (eds). *European Media Law.* (Alphena/d Rijn: Kluwer Law International, 2008).

Fernàndez, D. "ISP Liability Between EU and USA". (2016) 17 *Computer Law Review International* 36.

Geradin, D. "Online Intermediation Plattforms and Free Trade Principles: Some Reflections on the Uber Preliminary Ruling Case." In: Ortiz (ed). *Internet: Competition and Regulation of Online Platforms*. (Competition Policy International, 2016).

Gillespie, T. *Custodians of the Internet.* (New Haven/London: Yale University Press, 2018).

Helberger et al. "Governing online platforms: from contested to cooperative responsibility". 2018.

House of Lords Communications Committee. *Growing up with the Internet* (2nd Report of Session 2016–17) (HL Paper 130). 21 March 2017.

Husovec, M. *Injunctions Against Intermediaries in the European Union: Accountable but not liable?.* (Cambridge: Cambridge University Press, 2017).

Hussein, E. et al. "Measuring misinformation in video search platforms: An audit study on YouTube". (2020) Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1), Article 48. doi 10.1145/3392854.

Kozyreva et al. "Citizens versus the Internet: Confronting Digital Challenges with Cognitive Tools". (2020) 21(3) *Psychol Sci Public Interest* 103-156, doi: 10.1177/1529100620946707.

Leiser, M. "The Problem with 'Dots': questioning the role of rationality in the online environment." (2016) 30 *International Review of Law, Computers and Technology* 191.

MacCarthy, M. "What Payment Intermediaries are Doing about Online Liability and Why it Matters." (2010) 25 B*erkley Technology Law Journal* 1037.

Matz, S. et al. "Psychological targeting as an effective approach to digital mass persuasion". (2017) 114(48) *Proc Natl Acad Sci USA* 12714, doi: 10.1073/pnas.1710966114.

Murray, A. *Regulation of Cyberspace*. (2007, Oxford University Press).

OFCOM. *Discussion Document:Addressing Harmful Content Online.*

Pennycook, G. et al. "Shifting attention to accuracy can reduce misinformation online". (2021) *Nature*, 17 March 2021. https://doi.org/10.1038/s41586-021-03344-2.

Perrin, W and Woods, Lorna. *"Duty of Care" – Full Report*. April 2019. https://www.carnegieuktrust.org.uk/project/harm-reduction-in-social-media/.

Perrin, W and Woods, Lorna. Online Harms – Interlocking Regulation (Blog). 11 September 2020. https://www.carnegieuktrust.org.uk/blog/online-harms-interlocking-regulation/.

Robens Report: Safety and Health at Work, July 1972 (Cmnd 5034).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Schäfer, S. "Illusion of knowledge through Facebook news? Effects of snack news in a news feed on perceived knowledge, attitude strength, and willingness for discussions." (2020) 103 *Computers in Human Behavior* 1–12. 10.1016/j.chb.2019.08.031.

Seaver, N. "Captivating algorithms: Recommender systems as traps". (2018) Journal of Material Culture. https://journals.sagepub.com/doi/10.1177/1359183518820366.

Suler, J. "The Online disinhibition effect". 2004 7(3) *Cyberpsychol Behav* 321-6, doi: 10.1089/1094931041291295.

Sunstein, C. *Republic.com 2.0* (Princeton, NJ Princeton University Press, 2007).

Sunstein, C. *#Republic: Divided Democracy in the Age of Social Media* (Princeton, NJ, USA, and Oxford, UK: Princeton University Press, 2017).

Vijay, A. "Liability of internet service providers – a review study from the European perspective". (2019).

Woods, Lorna. "Video-sharing platforms in the revised Audiovisual Media Services Directive." (2018) 23 (3) *Communications Law* 127.

Woods, Lorna. The Carnegie Statutory Duty of Care and Fundamental Freedoms. 2019. https://www.carnegieuktrust.org.uk/publications/doc-fundamental-freedoms/.

Woods, Lorna. "The duty of care in the Online Harms White Paper." (2019) 11(1) *Journal of Media Law* 6.

Zuboff, S. "Big other: surveillance capitalism and the prospects of an information civilization." (2015) 30 *Journal of Information Technology* 75-89.

Zuboff, S. *The Age of Surveillance Capitalism– The Fight for a Human Future at the New Frontier of Power* (1st ed). (Profile Publishers: London, 2019).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Policy Developments in the USA to Address Platform Information Disorders*

Sarah Hartmann

Abstract: This chapter focuses on three factors contributing to the larger problem of information disorders in online platform environments – lack of reliable sources, lack of platform accountability, and lack of competition. By addressing these root causes, legislators can try to reshape the current communication environment in order to make it less vulnerable to information disorders. This chapter highlights current policy proposals and discussions on promoting trustworthy local news, incentivizing platforms to decrease the circulation of harmful speech through reform of Section 230, and increasing competition by mandating data portability and interoperability.

Keywords: Platform Regulation; Disinformation; Section 230 Reform; Intermediary Liability; Local News Subsidies; Data Portability; Interoperability

Chapter 1. Introduction and Overview

Online platforms are intrinsically linked to information disorders as a petri dish that allows extreme content, conspiracy theories and false information to multiply.[1] The term "information disorder" refers to content with different levels and combinations of falseness and intent to harm.[2]

---

* The chapter is based on Prof. Ellen P. Goodman's presentation during the workshop "Platform and Media Regulation – New Trends in Western Democracies" in February 2021. The author would like to thank Prof. Goodman for her helpful and valuable advice and comments.

1 Hunt Allcott and Matthew Gentzkow, "Social media and fake news in the 2016 election", *Stanford University, Journal of Economic Perspectives* 31 no. 2 (2017): 221.

2 Unknowingly incorrect representations (mis-information), intentionally manipulating or fabricated content (dis-information) and factual information and speech meant to attack or cause harm, such as hate speech or publication of private in-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Information disorders include many buzzword phenomena such as "fake news" and "hate speech", but are not limited to these vague terms.

A couple of decades ago, conspiracy theorists did not have the means to reach large audiences, let alone specifically target those they deemed like-minded or receptive to their message. Access to multipliers, such as broadcasting and print media, was controlled by professional journalistic institutions that acted as a filter for extremist or factually false content to protect themselves from liability. At the dawn of the internet age, individual messages could be published through private websites to a potentially unlimited audience. In practice, most private websites remained the online equivalent of soapbox speeches and never attracted wide public attention. Only the emergence of social media platforms introduced the element of amplification to an instant and expanding audience. Unlike legacy media outlets, platforms in their function as intermediaries do not filter content according to journalistic standards[3] and apply little to no upfront restriction, protected from liability for third party content as "neutral" intermediaries.[4] Meanwhile, platforms have drained advertising revenues of other media providers,[5] especially on the local level,[6] and effectively immunized themselves against potential competitors by holding their user's data hostage.

---

formation (mal-information), see Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward and interdisciplinary framework for research and policy making*, (Council of Europe report DGI(2017)09, 2017), 21, https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77.

3 Ellen P. Goodman, "Digital Information Fidelity and Friction", *Knight First Amendment Institute at Columbia University*, February 26, 2020, https://knightcolumbia.org/content/digital-fidelity-and-friction.

4 Guy Rolnick et al., *Protecting Journalism in the Age of Digital Platforms* (Stigler Center for the Study of the Economy and the State, University of Chicago Booth School of Business, July 1, 2019), 190, http://www.columbia.edu/~ap3116/papers/MediaReportFinal.pdf.

5 Jerrold Nadler, and David N. Cicilline, *Investigation of Competition in Digital Markets – majority staff report and recommendations*, (Subcommitee on antitrust, commercial and administrative law of the committee on the judiciary, 2020), 57 f., https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf?utm_campaign=4493-519.

6 Penelope Muse Abernathy, *News Deserts and Ghost Newspapers – Will Local News Survive*? (The Center for Innovation and Sustainability in Local media, Hussmann School of Journalism and Media, University of North Carolina at Chapel Hill, 2020), 8, https://www.usnewsdeserts.com/wp-content/uploads/2020/06/2020_News_Deserts_and_Ghost_Newspapers.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The effects of this media environment and the consequences of information disorders became especially evident in the United States in 2020 and 2021: from widespread misinformation about COVID-19, such as the alleged inefficacy of wearing face masks,[7] to allegations of election fraud culminating in the unprecedented capitol riots of January 6th 2021.[8] Discussions on the fallout inevitably zeroed in on the role of online platforms[9] and future preventive measures, with the US Congress holding a hearing[10] on the role of social media platforms in promoting misinformation and extremist content in late March 2021.

Across-the-board consensus maintains the need for measures against information disorders. This consensus is deceptive, however, as little common ground exists on the issues to be addressed or suitable countermeasures. Therefore, current policy proposals cover several fields and present a wide array of approaches. The following overview focuses on three factors contributing to the larger problem of information disorders –lack of reliable sources, lack of platform accountability, and lack of competition. This overview is not meant to be exhaustive, but instead aims to show the diversity of proposals and highlight the most promising or most prolific current policy approaches. Where appropriate, proposals are put into context with

7 See Richard A. Stein et al., "Conspiracy theories in the era of COVID-19: A tale of two pandemics", *The International Journal of Clinical Practice* 75 no. 2 (2021), 1, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7995222/pdf/IJCP-75-e13778.pdf .

8 See Timothy W. Luke, "Democracy under threat after 2020 national elections in the USA: 'stop the steal' or 'give more to the grifter-in-chief?'", *Educational Philosophy and Theory* (2021), https://www.tandfonline.com/doi/pdf/10.1080/00131857.2021.1889327?needAccess=true.

9 See Facebook's internal Report "Stop the Steal and Patriot Party: the Growth and Mitigation of an Adversarial Harmful Movement, available through *buzzfeednews*, April 26, 2021, https://www.buzzfeednews.com/article/ryanmac/full-facebook-stop-the-steal-internal-report?origin=tuh.

10 See H.R. Committee on Energy and Commerce, Memorandum on joint hearing "Disinformation Nation: Social Media's Role in Promoting Extremism and Disinformation", March 22, 2021, https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-20210325-SD002.pdf

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

recent EU initiatives such as the Digital Services Act[11] and Digital Markets Act.[12]

Chapter 2. Lack of Reliable Sources – Measures against the Decline of Local News

One factor contributing to the spread of mis- and disinformation is a lack of trusted reporting and distrust in available reporting.[13] Users are less likely to believe and perpetuate falsehoods if these are presented alongside reliable news on the same topics. An abundance of quality journalistic content in users' timelines might not directly counteract intentional communication of factually incorrect or misleading content, but it would immunize many of its recipients, enabling them to identify information as false.[14] In essence, enough "good" speech could go a long way towards countering "bad" speech.[15]

Unfortunately, traditional news outlets as a source of "good" speech have for years been suffering from declining revenues and competition with online media. The economic crisis of 2009 and, more recently, the effects[16] of the COVID-19 pandemic have in particular taken their toll

11 European Commission, "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC", COM(2020) 825 final, December 15, 2020.

12 European Commission, "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on contestable and fair markets in the digital sector (Digital Markets Act)", COM/2020/842 final, December 15, 2020.

13 Rasmus Kleis Nielsen and Lucas Grave, '*News you don't believe': Audience perspectives on fake news* (Oxford: Reuters Institute for the Study of Journalism, 2017), 7, https://ora.ox.ac.uk/objects/uuid:6eff4d14-bc72-404d-b78a-4c2573459ab8/download_file?file_format=pdf&safe_filename=Nielsen%2B-%2BAudience%2Bperspectives%2Bon%2Bfake%2Bnews.pdf&type_of_work=Report.

14 Nielsen and Grave, *News you don't believe,* 5.

15 Marko Milanovic, "Viral Misinformation and the Freedom of Expression: Part I", *EJIL:Talk!, Blog of the European Journal of International Law*, April 13, 2020, https://www.ejiltalk.org/viral-misinformation-and-the-freedom-of-expression-part-i/.

16 Anya Schiffrin, Hannah Clifford, and Kylie Tumiatti, *Saving Journalism: A Vision for the Post-Covid World* (Konrad Adenauer Stiftung, January 2021), 3 f., https://www.kas.de/documents/283221/283270/KAS_Saving+Journalism.pdf/8ee31596-7166-30b4-551f-c442686f91ae?version=1.4&t=1611338643015.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

on local newspapers and local broadcasters, the main and most trusted[17] source of news throughout the country. The system of decentralized and small private news providers was often unable to offer resistance to volatile market conditions. The resulting "news desert"[18] areas without access to local news providers are more vulnerable to unchecked information or misrepresentations that fill the void left behind.[19]

A recent report[20] by Senator Maria Cantwell identified the market behaviour of dominant online platforms as one of two major reasons for the struggling local news sector. Besides the general loss of ad business to online media,[21] news outlets suffer from "hijacking" of their content by news aggregators, especially by Google and Facebook, with little to no compensation.[22] Her findings are in line with the conclusions of a House investigation of competition in digital markets,[23] which also pointed to the dependency of news outlets on large platforms to disseminate their content.[24] On the one hand, news aggregation services and platforms are important points of entry to direct users to news sites and generate traffic.[25] On the other hand, news sites often compete with their own content excerpts and headlines presented by aggregators, rendering a visit to the source webpage unnecessary.[26] Overall, news content providers lack the bargaining power to determine the conditions of access to their content on platforms.[27] Changes in the platforms' recommender algorithms, such as Facebook's adjustment to its News Feed in 2018, have had major (negative) financial impacts on news sites and remain completely beyond their control.[28] Platforms may even place one-sided restrictions on content providers' ability to monetize content on their own sites through ad placement or paywalls, as was recently the case with Google's Accelerated

17 Maria Cantwell, *Local Journalism – America's Most Trusted News Sources Threatened* (U.S. Senate Committee on Commerce, Science, and Transportation, October 2020), 7 f., https://www.cantwell.senate.gov/imo/media/doc/Local%20Journalism%20Report%2010.26.20_430pm.pdf.
18 Abernathy, *News Deserts and Ghost Newspapers*, 8.
19 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 62.
20 Cantwell, *Local Journalism*.
21 Cantwell, *Local Journalism*, 14 f.
22 Cantwell, *Local Journalism*, 28 f.
23 Nadler and Cicilline, *Investigation of Competition in Digital Markets.*
24 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 63.
25 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 63.
26 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 59.
27 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 64
28 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 63.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Mobile Pages feature for news.[29] The program requires news web pages to be hosted on Google's infrastructure with a limited number of ads to allow for faster loading times and features no flexibility for paywalls.[30]

Both the recommendations of the House Investigation and the findings in Senator Cantwell's report suggest an antitrust approach, targeting certain platform business practices as abusive.[31] Senator Cantwell especially points out the need to address retaliatory practices, like hiding or removing local news content.[32] In order to improve the disparity between the bargaining power of local news providers and platforms, both reports suggest introducing a (temporary) safe harbour for news publishers and broadcasters to collectively bargain with news aggregators.[33] The House Investigation references[34] a draft bill[35] by Representative Cicilline, who also co-authored the Investigation, which sought to establish a limitation of liability under antitrust law for news content creators. The exemption would apply to negotiations among news content creators to collectively withhold content from online content distributors or collectively negotiate the terms for content distribution, given that the negotiations are non-discriminatory to other news providers and the agreed terms would be available to all news content creators.[36]

The Local Journalism Sustainability Act,[37] proposed in July 2020 by Representative Kirkpatrick, chooses a different approach, not relying on antitrust law but rather creating tax incentives in order to support local media. According to the draft bill, individuals are allowed tax credits of up to 250 USD for subscriptions to local newspapers[38] and small businesses[39] are granted tax credits up to 5.000 USD for advertising in local newspa-

29 Cantwell, *Local Journalism*, 31 f.
30 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 65.
31 Cantwell, *Local Journalism*, 56; Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 389 ff.
32 Cantwell, *Local Journalism*, 56.
33 Cantwell, *Local Journalism*, 55; Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 388.
34 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 389.
35 Journalism Competition and Preservation Act of 2019, H.R. 2054, 116th Cong. (2019); see also Abernathy, *News Deserts and Ghost Newspapers*, 83 f.
36 See Sec. 2 (b) Journalism Competition and Preservation Act of 2019, H.R. 2054, 116th Cong. (2019).
37 Local Journalism Sustainability Act, H.R. 7640, 116th Cong. (2020).
38 The tax credit covers 80% of the subscription costs for the first year and 50% for the following years, see Sec. 2 (c) Local Journalism Sustainability Act, H.R. 7640, 116th Cong. (2020).
39 Businesses with less than 1.000 employees.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

pers, radio or television.[40] Local newspapers are also given direct tax credit for 50% of their journalistically qualified employees' salaries.[41]

There appears to be hesitation to provide direct state subsidies to local news providers[42] outside of minor COVID pension relief.[43] In 2020, Members of the House of Representatives suggested allocating a portion of the government's ad budget to local media.[44] Civil society proposals[45] have meanwhile established the idea of cross-financing journalism through taxes on platform ad or other revenue.[46] On the state level, Maryland has already introduced a scale tax on revenue from digital ads displayed to citizens of Maryland.[47] While the tax is not tied to promotion of local journalism and has a strong likelihood of being struck down, it could still serve as a case study for other states in their efforts to fund local news. New Jersey, on the other hand, does not currently tax digital advertising, but has provided funds for the "Civic Information Consortium",[48] which will distribute grants to projects reviving local media.[49]

---

40 Sec. 2, 4 Local Journalism Sustainability Act, H.R. 7640, 116th Cong. (2020).

41 Up to 12.500 USD per quarter and 30% from the fifth quarter, see Sec. 3 (b) (1), (c) Local Journalism Sustainability Act, H.E. 7640, 116th Cong. (2020).

42 Schiffrin, Clifford, and Tumiatti, *Saving Journalism*, 12.

43 Craig Forman, "Covid Relief Bill Throws Lifeline to Transform Local news", *NiemanReports*, March 10, 2021, https://niemanreports.org/articles/covid-relief-bill-throws-lifeline-to-transform-local-news/; see also Abernathy, *News Deserts and Ghost Newspapers*, 80.

44 See the statement of Debbie Dingell et al. of April 20, 2021, https://debbiedingell.house.gov/uploadedfiles/200420supportlocalbroadcasters.pdf; a very similar proposal was brought forward in Rep. Ryan's Protect Local Media Act, H.R. 6913, 116th Cong. (2020).

45 See Schiffrin, Clifford, and Tumiatti, *Saving Journalism*, 24 f.; see also Guy Rolnick et al., *Protecting Journalism*, 34 ff. with a 'Media-Voucher' proposal; David Ardia et al., "Addressing the decline of local news, rise of platforms, and spread of mis- and disinformation online – A summary of current research and policy proposals" (Center for Information, Technology, and Public Life, December 2020), https://citap.unc.edu/local-news-platforms-mis-disinformation/.

46 Guy Rolnick et al., *Protecting Journalism*, 54.

47 David McCabe, "Maryland Approves Country's First Tax on Big Tech's Ad Revenue", *The New York Times*, February 12, 2021, https://www.nytimes.com/2021/02/12/technology/maryland-digital-ads-tax.html.

48 Sarah Stonbely, Matthew S. Weber, and Christopher Satullo, "Innovation in Public Funding for Local Journalism: A Case Study of New Jersey's 2018 Civic Information Bill", *Digital Journalism* 8, no. 6 (2020): 740-757.

49 See Civic Information Consortium, "About the Consortium", accessed April 27, 2021. https://njcivicinfo.org/about/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3. Lack of Platform Accountability – Draft Laws to Shrink Section 230 Immunity

A large share of the US debate on online platform regulation revolves around immunity of platforms from liability and lack of effort on their part to intervene against the spread of harmful or illegal content within their own networks. Section 230 (c) in its current form prevents platforms as "providers of interactive computer services" from being treated as the publisher or speaker of information by another information content provider. Furthermore, the Good Samaritan clause in Section 230 (c) (2) excludes civil liability for removal or restriction of content in "good faith". Introduced in the mid-1990s to promote competition with the telecommunications network[50] and allow new and innovative internet services to establish themselves under protection from liability for third-party content,[51] the immunity provision has lately been cited as part of the problem in dealing with platforms. Critics from opposing ends of the political spectrum focus on different aspects, for example alleging left-leaning bias in content moderation[52] and "censorship" by platforms of political opinions,[53] or suggesting a systemic failure to sufficiently protect vulnerable groups and prevent crime.[54]

Over the last two years, a number of bills to reform platform immunity have been presented, but none have been passed so far. Just since January 2021, seven different draft bills have been introduced or re-introduced

50 Karen Kornbluh and Ellen P. Goodman, "Bringing Truth to the Internet", *Democracy Journal* no. 53 (2019), https://democracyjournal.org/magazine/53/bringing-truth-to-the-internet/.

51 Paul M. Barret, *Regulating Social Media: the Fight over Section 230 – and Beyond* (New York University Stern Center for Business and Human Rights, September 2020), 4, https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5f58df637cbf80185f372776/1599659876276/NYU+Section+230_FINAL+ONLINE+UPDATED_Sept+8.pdf.

52 See draft bill by Sen. Hawley, Ending Support for Internet Censorship Act, S. 1914, 116th Cong. (2019).

53 See proposal for the CASE-IT Act, introduced by Reps. Steube and Gregory excluding section 230 immunity for providers "stifling free expression", Curbing Abuse and Saving Expression In Technology Act, H.R. 285, 117th Cong. (2021); see also a bill recently passed in Florida, fining social media platforms for "deplatforming" (blocking) political candidates, S.B. 7072, 2021 Session (Fla. 2021).

54 See e.g. the Safeguarding Against Fraud, Exploitation, Threats, Extremism, and Consumer Harms Act, S. 299, 117th Cong. (2021).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

from the previous congressional session.[55] The proposals can be broadly categorized by the kind of content or involvement of the platform they wish to exclude from immunity in the future.

Chapter 3.a. Limiting the Scope for Specific Categories of Content

The most straightforward and least controversial approach to reforming Section 230 is exclusion of certain categories of content from immunity. Draft bills along these lines are most likely to reach consensus. They continue the idea of existing limitations[56] for federal crimes, intellectual property violations and sex-trafficking charges.[57]

According to the SAFE TECH Act[58] of Senator Mark Warner, Section 230 would no longer be viable as a defence against claims on grounds of civil rights violations, cyberstalking, and harassment.[59] However, the proposal does not introduce explicit liability; it only removes the immunity granted by Section 230 as a "categorical bar" against legal redress by victims.[60] A narrower carve-out is included in Senator Lindsey Graham's EARN IT Act[61] concerning child sexual abuse material.

---

55 See the legislative tracker by Kiran Jeevanjee et al., "All the Ways Congress Wants to Change Section 230", *Slate*, March 23, 2021, https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html.

56 Eric Goldman, "An Overview of the United States Section 230 Internet Immunity", in *Online Intermediary Liability*, ed. Giancarlo Frosio (Oxford University Press, 2020), 160 ff.

57 See 47 USC § 230 (e); see also Barret, *Regulating Social Media*, 5.

58 Safeguarding Against Fraud, Exploitation, Threats, Extremism, and Consumer Harms Act, S. 299, 117th Cong. (2021).

59 See Sec. 2 (2) Safeguarding Against Fraud, Exploitation, Threats, Extremism, and Consumer Harms Act, S. 299, 117th Cong. (2021).

60 See Mark Warner, "Warner, Hirono, Klobuchar Announce the SAFE TECH Act to Reform Section 230", Press Release, February 5, 2021, https://www.warner.senate.gov/public/index.cfm/2021/2/warner-hirono-klobuchar-announce-the-safe-tech-act-to-reform-section-230; The exception to this rule is the FOSTA bill, in force since 2018, which not only withdrew Section 230 protection for facilitation of prostitution, but also instated a new offence, see 18 USC § 2421A.

61 Eliminating Abusive and Rampant Neglect of Interactive Technologies Act of 2020, S. 3398, 116th Cong. (2020).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Other initiatives focus on platform interactions with and treatment of third-party content, rather than the content itself. A bill introduced by Representatives Malinowski and Eshoo in October 2020[62] seeks to limit the scope of Section 230 in cases where the platform's algorithm has influenced the display of content to individual users, for example by ranking, recommendation or amplification, and the affected information is directly relevant to the claim. A similar legal argument was presented by plaintiffs in the *Force v. Facebook* case.[63] In his partially dissenting opinion, Judge Katzmann concurred that the limitation of liability in Section 230(c) (1) did not extend to Facebook's friend- and content-suggestion algorithms as they constitute original and separate messages from the content itself.[64] The majority opinion, however, rejected this notion.[65] The immunity exception proposed by Malinowski and Eshoo is only applicable to civil action claims on grounds of civil rights violations or terrorism.[66] The bill also defines certain algorithmic actions as "obvious, understandable, and transparent" which do not trigger the immunity exception, such as sorting information chronologically, alphabetically, or by user rating.

The SAFE TECH Act, mentioned above, limits the scope of Section 230 from a different angle. The bill excludes immunity for content that users or providers have been paid to make available.[67] The provision is meant to apply to advertisements which are placed and disseminated on platforms against payment, but could also be interpreted as including paid cloud services or paid prioritization.

Both proposals draw a dividing line between content that is treated "neutrally" or "passively" and instances where services actively intervene in content dissemination. Only services in the former category would continue to be protected from liability, while Section 230 would no longer apply to the latter category.[68] This differentiation is similar to the EU's liability

62 Protecting Americans from Dangerous Algorithms Act, H.R. 8636, 116th Cong. (2020).
63 *Force v. Facebook*, Inc., 934 F.3d 53 (2d Cir. 2019).
64 *Force v. Facebook*, Inc., 934 F.3d 53 (2d Cir. 2019), 82.
65 *Force v. Facebook*, Inc., 934 F.3d 53 (2d Cir. 2019), 66.
66 42 USC § 1985, § 1986.; 18 USC § 2333.
67 Sec. 2 (1) (a) Safeguarding Against Fraud, Exploitation, Threats, Extremism, and Consumer Harms Act, S. 299, 117th Cong. (2021).
68 A similar approach was suggested by Rolnick et al., *Protecting Journalism*, 16.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

privilege for hosting services,[69] which also relies on determining whether the provider's relationship with third party content is "of a mere technical, automatic or passive nature".[70] According to European Court of Justice case-law, online platforms such as eBay start being "actively" involved once they help optimize and promote individual sale offers, for example by placing ads for the offer in search engines.[71] As a consequence, the hosting privilege does not apply to eBay in this case. However, just as under Section 230, excluding the liability privilege does not lead to automatic liability, which must be provided separately by national or European law.[72]

Chapter 3.c. Additional Obligations as Prerequisites for Immunity

Finally, different legislative and academic proposals seek to introduce new accompanying obligations for platforms either as prerequisites for Section 230 immunity or as separate duties. The idea of "earned" immunity has been discussed by Citron and Wittes, for example, on the condition of reasonable moderation practices,[73] and recommended in the Stigler report in the form of a "quid pro quo" for fulfilment of obligations mainly relating to transparency.[74] In the context of the recent congressional hearing, Mark Zuckerberg of Facebook expressed support for a similar system of conditional immunity, requiring compliance with best practice standards of content moderation and systems to identify and remove harmful content.[75] On the other hand, this approach has been criticized for conflating

---

69 Currently Art. 14 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce'), OJ L 178, 17.7.2000, p. 1–16; see also Art. 5 Digital Services Act proposal, COM(2020) 825 final.

70 Rec. 42 Directive on electronic commerce.

71 Case C-324/09, *L'Oréal SA v eBay International AG* [2011] ECR I-06011, marginal no. 116.

72 See Rec. 17 Digital Services Act proposal, COM(2020) 825 final.

73 Danielle Keats Citron and Benjamin Wittes, "The Problem isn't just Backpage: Revising Section 230 Immunity", *Georgetown Law Technology Review* (2018): 453.

74 Rolnick et al., *Protecting Journalism in the Age of Digital Platforms*, 195.

75 *Disinformation Nation: Social Media's Role in Promoting Extremism and Disinformation: joint hearing before the United States House of Representatives Committee on Energy and Commerce Subcommittees on Consumer Protection & Commerce and Communications & Technology*, March 25, 2021, Testimony of Mark Zuckerberg of Facebook, Inc., 7, https://docs.house.gov/meetings/IF/IF16/20210325/111407/HHRG-117-IF16-Wstate-ZuckerbergM-20210325-U1.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the question of liability with other policy goals, which should be regulated separately.[76]

A draft bill by Senators Schatz and Thune, the PACT Act,[77] contains both comprehensive transparency and moderation provisions, such as a duty to explain content moderation practices to users and establish a user complaint mechanism,[78] as well as a notice-and-takedown system tied to Section 230. According to the proposal, the liability privilege only applies to platforms who have either no knowledge of the content in question or have taken the necessary steps to review and remove or otherwise restrict the content after receiving notice.[79] This approach most closely resembles the current EU regime of the E-Commerce Directive and the Digital Services Act proposal, where the liability privilege and additional obligations are also regulated separately. Article 14 (1) of the E-Commerce Directive exempts hosting services from liability if they either have no knowledge of illegal activity or information or, upon obtaining such knowledge, restrict the content in question. The Digital Services Act proposal builds upon the principle of hosting privilege, but links it to a notice and action mechanism, mandatory for online platforms.[80] Qualified notices issued through this mechanism are "considered to give rise to actual knowledge or awareness", thereby obligating the platform to act upon the notice in order to benefit from the hosting privilege.[81] Other obligations of intermediary services in the Digital Services Act proposal[82] are not directly linked to liability but subject to enforcement and monetary penalties in case of non-compliance.[83]

76 Mark MacCarthy, "Back to the future for Section 230 reform", *Brookings Techtank*, March 17, 2021, https://www.brookings.edu/blog/techtank/2021/03/17/back-to-the-future-for-section-230-reform/

77 Platform Accountability and Consumer Transparency Act, S. 4066, 116th Cong. (2020).

78 Sec. 5 (a) and (b) Platform Accountability and Consumer Transparency Act, S. 4066, 116th Cong. (2020).

79 Sec. 6 (a) Platform Accountability and Consumer Transparency Act, S. 4066, 116th Cong. (2020).

80 Art. 14 Digital Services Act proposal, COM(2020) 825 final.

81 Art. 14 (3) Digital Services Act proposal, COM(2020) 825 final.

82 Art. 10 ff. Digital Services Act proposal, COM(2020) 825 final.

83 See Art. 42 Digital Services Act proposal, COM(2020) 825 final.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 4. Lack of Competition – Introducing Portability and Interoperability

Finally, an important characteristic of the current environment that facilitated the spread of information disorders is the high concentration in the platform market. General antitrust efforts in dealing with online platforms have increased in the USA[84] and elsewhere. There is considerably less hesitation in turning to antitrust law than to introducing media regulation.

Among the complex causes of platform dominance are so-called lock-in effects; these disincentivise users of one service from switching to alternate providers or using several services in parallel.[85] This, in turn, creates high entry barriers for competitors and renders users and the platform service as a whole more vulnerable to information disorders within the network.[86] In order to alleviate the barriers around online platforms that keep users in and competitors out, the introduction of interoperability[87] and portability[88] rules has been discussed.[89] In theory, data portability would empower users to take the information linked to their accounts from one platform to another platform,[90] the digital equivalent of moving apartments and bringing every piece of furniture along to the new apartment. Interoperability on the other hand would enable different platforms' systems to connect and communicate with one another through mutually established protocols.[91] Much as clients of different mobile providers are able to exchange calls and messages,[92] YouTube users might be able to send private messages to Instagram users and vice versa. In the context of information disorders, interoperability and data portability could potentially foster competition between different platforms' algorithms.

---

84 See Nadler and Cicilline, *Investigation of Competition in Digital Markets.*

85 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 384.

86 Judit Bayer et al., *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States* (Study for the European Parliament, 2019), 136, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf.

87 Wolfgang Kerber and Heike Schweitzer, "Interoperability in the Digital Economy", *JIPITEC* 8 no. 1 (2017): 39, https://www.jipitec.eu/issues/jipitec-8-1-2017/4531.

88 Ruth Janal, "Data Portability – A Tale of Two Concepts", *JIPITEC* 8 no. 1 (2017): 59, https://www.jipitec.eu/issues/jipitec-8-1-2017/4532.

89 Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 385 ff.

90 Janal, "Data Portability", 60; Nadler and Cicilline, *Investigation of Competition in Digital Markets*, 386.

91 Kerber and Schweitzer, "Interoperability in the Digital Economy", 40; Nadler and Cicilline, *Investigation of Competition in Digital Markets,* 385.

92 Rolnick et al., *Protecting Journalism in the Age of Digital Platforms*, 16.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

A draft bill from 2019 by Senators Warner and Hawley, the ACCESS Act,[93] proposed the introduction of a portability duty for platforms with more than 100,000,000 monthly active users in the USA. Platforms would be obligated to implement a system for the transfer of user data in a structured, commonly used and machine-readable format to other communication providers at the discretion of the user.[94] The bill also included an interoperability duty for the same platforms, requiring accessible interfaces to allow communications with users of competing providers.[95] Platform providers that operate several platforms or other products and services that are interoperable (such as Facebook and Instagram) are additionally required to provide a functionally equivalent version of their interface to competitors.[96] Finally, the interoperability requirement is also extended to custodial third party services that users may employ to manage their account settings, content, and online interactions.[97] Custodial services are bound by a duty of care and must be granted access to all functions available to the user on the same terms as the user. In theory, a third party service like this could be used across several platforms as a one-stop-shop for settings and communications, aggregating messages and other content for the user.

The ACCESS Act's interoperability requirements exceed the current EU framework.[98] The recent EU Commission proposal for a Digital Markets Act only includes interoperability requirements for gatekeepers' operating systems with third-party software and ancillary services, which do not apply to core platform services.[99] While the ACCESS ACT would obligate a platform like Facebook to enable its users to communicate with users

93 Augmenting Compatibility and Competition by Enabling Service Switching Act of 2019, S. 2658, 116th Cong. (2019).
94 Sec. 3 Augmenting Compatibility and Competition by Enabling Service Switching Act of 2019, S. 2658, 116th Cong. (2019).
95 Sec. 4 Augmenting Compatibility and Competition by Enabling Service Switching Act of 2019, S. 2658, 116th Cong. (2019).
96 Sec. 4 (3) Augmenting Compatibility and Competition by Enabling Service Switching Act of 2019, S. 2658, 116th Cong. (2019).
97 Sec. 5 Augmenting Compatibility and Competition by Enabling Service Switching Act of 2019, S. 2658, 116th Cong. (2019).
98 Rec. 68 General Data Protection Regulation only "encourages" development of interoperable formats instead of obliging data controllers, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016, 1–88.
99 Art. 6 sec. 1 lit. c, f Digital Markets Act proposal, COM/2020/842 final.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of other platforms, e.g. share content with them, the Digital Markets Act proposal only prevents operating systems or device manufacturers (such as Google Android) from restricting installation of third-party applications.[100]

In terms of portability, the wording of Sec. 3 in the ACCESS Act proposal is reminiscent of the Right to Data Portability in Article 20 of the EU General Data Protection Regulation (GDPR). Both concepts share one vital restraint, though: While they establish an obligation to transfer user data to the individual user or another provider, there is no equivalent obligation for other providers to enable reception of such data. As a competition tool, data portability requires a suitable destination for user data, i.e. a competing online platform with similar features for storing and displaying content that the user wishes to import. This will not be technically possible in many cases. Like an oversized couch that just will not fit into a new apartment, Facebook account data, including photos and other media, could hardly be imported to a platform such as Twitter. Nevertheless, interoperability and portability can be important building blocks in broader competition policy.[101]

Chapter 5. Conclusion

What does the future hold for regulation of information disorders on online platforms? Considering the fragmented policy proposals highlighted above, we cannot be sure. This is partially due to the diffuse nature of information disorders that do not lend themselves to traditional regulation. Rather, legislators can only try to reshape certain aspects of online communication in order to indirectly counteract information disorders.

The most prominent topic of recent policy debate in connection with platforms has been the reform of Section 230 immunity. To a certain extent, Section 230 has become a symbol of many things regarded as "wrong" with the current framework for online platforms. It is important to keep in mind, however, that Section 230 is not a blanket provision for content moderation, but a rule specifically addressing provider's liability

100 Rec. 52 Digital Markets Act proposal, COM/2020/842 final.

101 Paul de Hert et al, "The right to data portability in the GDPR: Towards user-centric interoperability of digital services", *Computer Law & Security Review* 34 no. 2 (2018): 194, https://www.sciencedirect.com/science/article/pii/S0267364917303333?via%3Dihub#fn0300; Rolnick et al., *Protecting Journalism in the Age of Digital Platforms*, 16.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

for illegal content. With its free speech protection famously one of the strongest in the world,[102] only very few categories of illegal content exist in the USA, in contrast to the European legal framework.[103] Especially pertaining to mostly legal but harmful disinformation, debates on general platform liability tend to generate more smoke than fire.[104] Among the reform approaches discussed above, only those including additional requirements for immunity as a "quid pro quo"[105] or implementing new regulatory obligations independent of liability[106] have the potential to go beyond this limited impact. The proposals share a stronger emphasis on systemic features, such as transparency, addressing platform design and not individual content, and are comparable to the approach in the European Digital Services Act proposal. This is preferable, as it avoids turning either the government or platforms into arbiters of acceptable speech.[107]

A positive approach against information disorders would be promotion of trustworthy news over sensationalist or dubious content. Above all, this requires a viable environment for local news providers. Proposed solutions tackling their current decline range from antitrust remedies to tax incentives, but shy away from providing direct government funding. In theory, promotion of local news as public value content on online platforms could also be mandated as a design feature in connection with immunity requirements.[108]

Lastly, antitrust efforts addressing the market dominance of (certain) online platforms have increased in the last years. In order to show a positive effect as a remedy for information disorders, competitors would first have to establish themselves in a very concentrated platform market. Interoperability and portability requirements as proposed by the ACCESS Act could be helpful in counteracting information disorders, although the

---

102 Kate Jones, *Online Disinformation and Political Discourse – Applying a Human Rights Framework*, (Chatham House, 2019), 19, https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf.

103 Barret, *Regulating Social Media*, 6.

104 MacCarthy, "Back to the future".

105 See Citron and Wittes, "The Problem isn't just Backpage", 471; Rolnick et al., *Protecting Journalism*, 195.

106 See Sec. 5 Platform Accountability and Consumer Transparency Act, S. 4066, 116th Cong. (2020).

107 Kornbluh and Goodman, "Bringing Truth to the Internet".

108 Bernd Holznagel and Sarah Hartmann, "Reforming competition and media law – the German approach" In *Regulating Big Tech: Policy Responses to Digital Dominance*, eds. Martin Moore and Damian Tambini (Oxford University Press, 2021).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

portability obligation already established in European law has not had a major impact in that regard.

Overall, most policy proposals do not specifically address information disorders, but rather are primarily geared towards other issues, like press subsidies, liability, and economic competition. Progress likely depends on many of the proposed measures interlinking to achieve a policy sum that is greater than its individual parts.

Bibliography

Abernathy, Penelope. *News Deserts and Ghost Newspapers – Will Local News Survive*?. University of North Carolina at Chapel Hill, Hussmann School of Journalism and Media, The Center for Innovation and Sustainability in Local media, 2020. https://www.usnewsdeserts.com/wp-content/uploads/2020/06/2020_News_Deserts_and_Ghost_Newspapers.pdf.

Allcott, Hunt and Gentzkow, Matthew. "Social media and fake news in the 2016 election." *Stanford University, Journal of Economic Perspectives* 31 no. 2 (2017): 211-236. https://web.stanford.edu/~gentzkow/research/fakenews.pdf.

Ardia, David, Ringel, Evan, Smith Ekstrand, Victoria and Fox, Ashley. "Addressing the decline of local news, rise of platforms, and spread of mis- and disinformation online – A summary of current research and policy proposals." *University of North Carolina, Center for Information, Technology, and Public Life*, December 2020. https://citap.unc.edu/local-news-platforms-mis-disinformation/.

Barret, Paul. *Regulating Social Media: the Fight over Section 230 – and Beyond*. New York University Stern Center for Business and Human Rights, September 2020. https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5f58df637cbf80185f372776/1599659876276/NYU+Section+230_FINAL+ONLINE+UPDATED_Sept+8.pdf.

Bayer, Judit, Bitiukova, Natalija, Bárd, Petra, Szakács, Judit, Alemanno, Alberto, and Uszkiewicz, Erik. *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States*. Study for the European Parliament, 2019. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf.

Cantwell, Maria. *Local Journalism – America's Most Trusted News Sources Threatened*. U.S. Senate Committee on Commerce, Science, and Transportation, October 2020. https://www.cantwell.senate.gov/imo/media/doc/Local%20Journalism%20Report%2010.26.20_430pm.pdf.

Citron, Danielle Keats and Wittes, Benjamin. "The Problem isn't just Backpage: Revising Section 230 Immunity." *Georgetown Law Technology Review* 2, no. 2 (2018): 453-473. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3218521.

Forman, Craig. "Covid Relief Bill Throws Lifeline to Transform Local news." *NiemanReports*, March 10, 2021. https://niemanreports.org/articles/covid-relief-bill-throws-lifeline-to-transform-local-news/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Goldman, Eric. “An Overview of the United States Section 230 Internet Immunity.” In *Online Intermediary Liability*, edited by Giancarlo Frosio, 155-171. Oxford University Press, 2020.

Goodman, Ellen. “Digital Information Fidelity and Friction.” *Knight First Amendment Institute at Columbia University*, February 26, 2020. https://knightcolumbia.org/content/digital-fidelity-and-friction.

de Hert, Paul, Papakonstantinou, Vagelis, Malgieri, Gianclaudio, Beslay, Laurent, and Sanchez, Ingacio. “The right to data portability in the GDPR: Towards user-centric interoperability of digital services.” *Computer Law & Security Review* 34 no. 2 (2018): 193-203. https://www.sciencedirect.com/science/article/pii/S0267364917303333?via%3Dihub#fn0300.

Holznagel, Bernd and Hartmann, Sarah. „Reforming competition and media law – the German approach.” In *Regulating Big Tech: Policy Responses to Digital Dominance,* edited by Martin Moore and Damian Tambini. Oxford University Press, 2021.

Janal, Ruth. “Data Portability – A Tale of Two Concepts.”, *Journal of Intellectual Property, Information Technology and E-Commerce Law* 8 no. 1 (2017): 59-69. https://www.jipitec.eu/issues/jipitec-8-1-2017/4532.

Jeevanjee, Kiran, Lim, Brian, Ly, Irene, Perault, Matt, Ruddock, Jenna, Schmeling, Tim, Vattikonda, Niharika, and Zhou, Joyce. “All the Ways Congress Wants to Change Section 230”, *Slate*, March 23, 2021. https://slate.com/technology/2021/03/section-230-reform-legislative-tracker.html.

Jones, Kate. *Online Disinformation and Political Discourse – Applying a Human Rights Framework*. Chatham House, 2019. https://www.chathamhouse.org/sites/default/files/2019-11-05-Online-Disinformation-Human-Rights.pdf.

Kerber, Wolfgang and Schweitzer, Heike. “Interoperability in the Digital Economy.” *Journal of Intellectual Property, Information Technology and E-Commerce Law* 8 no. 1 (2017): 39-58. https://www.jipitec.eu/issues/jipitec-8-1-2017/4531.

Kornbluh, Karen and Goodman, Ellen. „Bringing Truth to the Internet.“ *Democracy Journal* no. 53 (2019), https://democracyjournal.org/magazine/53/bringing-truth-to-the-internet/.

Luke, Timothy. “Democracy under threat after 2020 national elections in the USA: ‘stop the steal’ or ‘give more to the grifter-in-chief?’”, *Educational Philosophy and Theory* (2021). https://www.tandfonline.com/doi/pdf/10.1080/00131857.2021.1889327?needAccess=true.

MacCarthy, Mark. “Back to the future for Section 230 reform.” *Brookings Techtank*, March 17, 2021. https://www.brookings.edu/blog/techtank/2021/03/17/back-to-the-future-for-section-230-reform/.

McCabe, David. “Maryland Approves Country’s First Tax on Big Tech’s Ad Revenue.”, *The New York Times*, February 12, 2021. https://www.nytimes.com/2021/02/12/technology/maryland-digital-ads-tax.html.

Milanovic, Marko “Viral Misinformation and the Freedom of Expression: Part I.” *EJIL:Talk!, Blog of the European Journal of International Law*, April 13, 2020. https://www.ejiltalk.org/viral-misinformation-and-the-freedom-of-expression-part-i/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Nadler, Jerrold and Cicilline, David. *Investigation of Competition in Digital Markets – majority staff report and recommendations*. Subcommitee on antitrust, commercial and administrative law of the committee on the judiciary, 2020. https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf?utm_campaign=4493-519.

Nielsen, Rasmus and Grave, Lucas *'News you don't believe': Audience perspectives on fake news.* Oxford: Reuters Institute for the Study of Journalism, 2017. https://ora.ox.ac.uk/objects/uuid:6eff4d14-bc72-404d-b78a-4c2573459ab8/download_file?file_format=pdf&safe_filename=Nielsen%2B-%2BAudience%2Bperspectives%2Bon%2Bfake%2Bnews.pdf&type_of_work=Report.

Rolnick, Guy, Cagé, Julia, Gans, Joshua, Goodman, Ellen, Knight, Brian, Prat, Andrea, Schiffrin, Anya, and Raj, Prateek. *Protecting Journalism in the Age of Digital Platforms*. Stigler Center for the Study of the Economy and the State, University of Chicago Booth School of Business, July 1, 2019. http://www.columbia.edu/~ap3116/papers/MediaReportFinal.pdf.

Schiffrin, Anya, Clifford, Hannah, and Tumiatti, Kylie. *Saving Journalism: A Vision for the Post-Covid World.* Konrad Adenauer Stiftung, January 2021. https://www.kas.de/documents/283221/283270/KAS_Saving+Journalism.pdf/8ee31596-7166-30b4-551f-c442686f91ae?version=1.4&t=1611338643015.

Stein, Richard, Ometa, Oana, Shetty, Sarah, Katz, Adi, Popitiu, Mircea, and Brotherton, Robert. „Conspiracy theories in the era of COVID-19: A tale of two pandemics." *The International Journal of Clinical Practice* 75 no. 2 (2021): 1-5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7995222/pdf/IJCP-75-e13778.pdf.

Stonbely, Sarah, Weber, Matthew and Satullo, Christopher. „Innovation in Public Funding for Local Journalism: A Case Study of New Jersey's 2018 Civic Information Bill.", *Digital Journalism* 8, no. 6 (2020): 740-757.

Wardle, Claire and Derakhshan, Hossein. *Information Disorder: Toward and interdisciplinary framework for research and policy making.* Council of Europe report DGI(2017)09, 2017. https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Interoperability of Messenger Services. Possibilities for a Consumer-Friendly Approach

Jörg Becker, Bernd Holznagel, Kilian Müller[1]

Abstract: Messenger services connect consumers around the globe on a daily basis and help them share news, photos, videos and other information. To communicate, however, users need a shared messenger service. Information exchange between different service providers is rarely possible. Due to network effects, this leads to an increasing monopolisation of the messenger service market. Users are increasingly forced to use messenger services from a single manufacturer, which further extends its supremacy. Therefore, it is investigated to what extent the introduction of an interoperability obligation could counteract these effects in a consumer-friendly way. As a possible solution, the introduction of a federal XMPP-based system is presented, which could be used to implement an interoperability obligation in practice.

Keywords: Platform Regulation, Messenger Services, Federated Protocols, Competition, Interoperability and its legal possibilities under EU law

Chapter 1: Introduction

Messenger services are becoming increasingly popular. While in 2016, around 67% of all users aged 14 and above used messenger services[2], this figure increased to almost 90% in 2018. The figure is even higher among younger people aged between 14 and 29, where almost everyone (98%) now uses messenger services.[3] Facebook holds the largest market share. WhatsApp alone accounts for 96% of usage share. The second most-

1 The study was funded by the German Federal Ministry of Justice and Consumer Protection based on a resolution of the German Bundestag.

2 "Zwei von drei Internetnutzern verwenden Messenger", bitkom, accessed June 23, 2021, https://www.bitkom.org/Presse/Presseinformation/Zwei-von-drei-Internetnutzern-verwenden-Messenger.html.

3 „Zwei von drei Internetnutzern".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

used communication service, also from Facebook, is Facebook Messenger with 42%.[4] This market position enables Facebook to exercise a dominant position on the market, which is increasingly strengthened by existing network effects. These network effects increase the benefit of a particular service for all users involved as the number of users increases. In other words, the more users already use a service, the more attractive it becomes for new users, and for each new user the benefit of all existing users increases. In extreme cases, this can lead to so-called lock-in effects, through which consumers cannot move to another service provider without inconvenience.

One way to prevent or disrupt such a monopolistic position is to impose an interoperability obligation. If an interoperability obligation were to apply in the messenger service market, as introduced by the EECC (see chapter C), users would have to be able to exchange messages with users of other services without having to install the respective service. Thus, they can use their chosen service to contact users of all other services. Therefore, it will be investigated how an interoperability obligation can be technically executed and how it affects competition, innovativeness, data privacy, and usability for consumers.

In the following, the subjects of discussion (B I.), as well as the functionalities and the interoperability of messenger services, are presented (B II.). This is followed by an evaluation of the interoperability concept (B III.). In B IV, the design options for a successful introduction of interoperability are explained in more detail. Subsequently, possible interoperability obligations under the EKEK, the TKMoG-E, and the GWG-E are addressed (C I-III.). The paper ends with a conclusion (D).

4 Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, *Nutzung von OTT-Kommunikationsdiensten in Deutschland* (Bonn: Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2020), https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Berichte/2020/OTT.pdf?__blob=publicationFile.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 2. Technical/economical view

2.1. Subject

2.1.1. Messenger services

The German Federal Network Agency ("Bundesnetzagentur") classifies messenger services and other digital platforms in the category of "over-the-top" (OTT) services, which enable communication and other services via the Internet.[5] This means that, in contrast to traditional telephony and text-messaging services, the use of OTT services is not tied to the respective mobile or landline connection. They, therefore, do not require their own infrastructures but use existing infrastructures to build their services on. These services include for example sending messages, displaying a status, or sending images or video material.

The key task of messenger services is to connect users for the exchange of messages and other information. In contrast to social networks such as Facebook, YouTube, or Instagram, in which content can be broadcast by a sender and then be displayed to various consumers, in messenger services the addressee is specifically selected, which means that messages can only be received by this person. The message is sent via various open communication protocols such as XMPP, IRC, or Echo, or via proprietary protocols such as WhatsApp or Skype over the public Internet using a client. An addressee who has the same client or the same protocol can receive and read this message. Open protocols thus open up the theoretical possibility of connecting users from different clients with each other, whereas proprietary protocols usually exclude this possibility and only allow communications between users of their own client.

2.1.2. Interoperability

Communication between users of two different clients would constitute interoperable communication. In general terms, interoperability refers to the ability of different systems to communicate with each other and to be able to use the communicated information.[6] In a technical context,

---

5 Bundesnetzagentur, *Nutzung von OTT-Kommunikationsdiensten*.

6 "Stellungnahme der Digitalen Gesellschaft e. V. zur Konsultation des Bundesministeriums der Justiz und für Verbraucherschutz zu Interoperabilität und Daten-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

interoperability includes the ability of two or more software components to work together despite differences in language, interface, and execution environment. In the context of messenger services, interoperability refers to the possibility of exchanging messages and other types of communication not only between users of the same messenger A, but also between different messenger services A and B. According to Wegner, the two main mechanisms for creating interoperability are the creation of interfaces and standardization, with standardization being more scalable and interfaces being more flexible.[7] The use of a common communication protocol would also mean standardization of functionalities. The use of interfaces would make certain functionalities interoperable, independent of protocol. In this case, it could be individually selected which functionalities and data are passed on.

2.2. Introducing Interoperability

As stated above, according to Wegner, there are basically two ways to ensure interoperability between messenger services, standardization and the creation of interfaces.[8] The creation of federated systems uses an interface to mediate between different protocols or domains, using a common standard. A federation can be seen as a combination of both approaches.

2.2.1. Interfaces

One way to enable interoperability between messenger services is to create interfaces called application programming interfaces (APIs). APIs are data processing interfaces that define the interactions between multiple software components. APIs provide only the data that the other software component needs or requests. Using standardized formats, such as JSON or XML, certain components, such as the text content of a message, its re-

---

portabilität bei sozialen Netzwerken", Digitale Gesellschaft e. V., accessed June 23, 2021, https://digitalegesellschaft.de/2019/05/stellungnahme-der-digitalen-gesellschaft-e-v-zur-konsultation-des-bundesministeriums-der-justiz-und-fuer-verbraucherschutz-zu-interoperabilitaet-und-datenportabilitaet-bei-sozialen-netzwerken/.

7 Peter Wegner, "Interoperability", *ACM Computing Surveys* 28, no. 1 (March 1996): 285 ff., https://doi.org/10.1145/234313.234424.

8 Wegner, “Interoperability”, 285 ff.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

cipient, or telephone numbers, could be exchanged automatically between two or more different service providers.

2.2.2. Standardization

Standardization of communication protocols is another option for enabling interoperability between messenger services. Many functionalities of the messenger services depend on the respectively used protocols. It must therefore be ensured that a protocol is selected or created that supports all the functionalities that are to be interoperable. For this purpose, each messenger service must change its implementations to the new protocol or support multiple protocols.

2.2.3. Federation

In the case of a federation, the individual messenger services could retain their own protocols. However, a standardized protocol is used to which different parts of the respective messenger protocols can be "mapped". Thus, messages sent by users with the same messenger can still be transmitted using the messenger's own protocol. However, if a message is sent that is addressed to a user with a different messenger service, the interoperable parts of this message are passed to the federated protocol, transmitted to the target messenger, and finally forwarded to the recipient. For this purpose, a unified encryption protocol must be used to ensure end-to-end encryption.

2.3. Impact of interoperability

As interfaces are becoming increasingly impractical with an increasing number of participants and standardization hinders innovativeness the introduction of a federated system is examined with regard to its impact on competition, innovativeness, data privacy, and usability has to be examined.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2.3.1. Competition

Messenger services, being digital platforms, benefit from network effects. The more users such a digital platform has, the greater its benefit for all users involved (direct network effects).[9] This can influence new users in particular, as they increasingly opt for the largest provider in order to gain the greatest benefit from existing network effects. Furthermore, digital platforms with a large user base can use their existing network to expand into other areas by using the existing database for other purposes. This allows a company to continuously improve a product through the existing network (economies of scope), an opportunity not available to competitors with smaller networks. Another advantage for Facebook results from the use of indirect network effects. For both advertising and analysis purposes, Facebook can draw on a significantly larger user and database and thus generate further advantages over smaller competitors. In general, it should therefore be the case that the more users a messenger service has, the more useful this service is for all parties involved.[10] However, indirect network effects can put the users of a service at a disadvantage due to increased advertising or other use of their personal data. This is particularly problematic if the market share of a single service is sufficiently large, as this can lead to lock-in effects that make the user dependent on the respective service if there is no sufficient other alternative. As already mentioned, Facebook has market shares of over 90%. If "Facebook-external" users, i.e., users without a messenger service from Facebook, want to communicate with other users who use a Facebook messenger service such as WhatsApp, they are currently forced to use a messenger service from Facebook. Installing another shared messenger service on both sides (multihoming) would be an alternative, though it would involve additional effort. If consumers do not use a Facebook messenger service, they will not be able to reach most of their contacts via another messenger service unless the other person has installed another messenger service. In 65% of the cases, users have at least two different messenger services installed.[11] However, this does not always have to be the same additional messenger service, which would force consumers to install more than two different messenger services. The installation of multiple messenger services, though, consumes more

9 Michael L. Katz and Carl Shapiro, "Systems Competitionand Network Effects", *Journal of Economic Perspectives* 8, no. 2 (Spring 1994): 93 ff., https://doi.org/10.1257/jep.8.2.93.

10 Katz and Shapiro, “Systems competition”, 93 ff.

11 Bundesnetzagentur, *Nutzung von OTT-Kommunikationsdiensten*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

storage space, causes a wider distribution of personal data, and ultimately forces users to install a Facebook messenger service anyway in order to achieve full availability. At this point, it should be emphasized that multi-homing can of course be used to achieve almost full availability, but this is never possible without a Facebook Messenger service. Users are thus faced with the decision of passing on their personal data to Facebook or giving up full availability. To strengthen broader competition independent of Facebook, it is necessary to break up the prevailing network effects. Generally speaking, any form of interoperability obligation supports the intention to mitigate network effects. By enabling users to communicate with users of other services, they are no longer dependent on the provider with the largest user base but can select the messenger service according to other criteria. Functionality, graphical user interface, etc. are already criteria by which users select a messenger service, but they currently play a minor role compared to network effects.[12] An introduction of an interoperability obligation could therefore strengthen competition based on new functionalities or better interfaces. Furthermore, this would give new companies a better chance to enter the competition, as they are not measured by the size of their user base, but by their functionalities, data privacy, etc. However, an interoperability obligation must protect the vendors' Unique Selling Points (USPs) to ensure a justification for smaller vendors to exist. In terms of competition, no differences are expected regarding the three different design options.

2.3.2. Innovativeness

Innovativeness in the context of messenger services refers to the ability to continuously create new functionalities or to continuously improve existing functionalities. One incentive for companies to be innovative is the resulting competitive advantages. However, these would be lost if other competitors were able to adapt or copy an innovation immediately without major difficulties. New functionalities, for example, require considerable implementation and testing effort before they can be introduced in a stable form. It should also be noted that too much market power could cause new/small companies to shy away from investing in innovations, as these are unlikely to pay off.

12 Katz and Shapiro, "Systems competition", 93 ff.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

A federation circumvents the disadvantages of standardization in terms of innovativeness by still allowing companies to use their own protocols for intra-messenger communication. Only inter-messenger communication relies on a standardized protocol, to which functionalities can of course be added. However, messenger services can implement their own innovations in their own protocols and use these only for intra-messenger communications, thus maintaining their competitive advantage and continuing to generate incentives for future innovations. This avoids the problem that new functionalities based on a completely standardized protocol take a long time to be implemented. There would only be an additional effort in terms of implementation if further additional functionalities were made interoperable in the future, as these would then also have to be provided in inter-messenger communication.

2.3.3. Data privacy

Regarding data privacy, a distinction can be made between several aspects. On the one hand, end-to-end encryption can suffer from an interoperability obligation if no common encryption protocol is used, and the message thus must be decrypted and re-encrypted in several steps. However, this can be circumvented by using a common encryption protocol. Since inter-messenger messages automatically involve two companies in the communication, they inevitably generate more metadata than intra-messenger communications. The sending service needs information about the recipient of the message to deliver it. When the addressee replies, the service responsible for this also requires information about the original sender. The generated personal data is, however, significantly less (fewer providers are required) than when using multihoming, since the messenger service that is not installed and involved in the communication only obtains information about the username of the unknown user, but ideally does not obtain any further information due to end-to-end encryption. Thus, information about the address book, access to photos, etc. could be kept secret from the other messenger service. The transmission of the data by the sender and the storage of the data by the recipient qualify as processing within the meaning of Art. 4 No. 2 GDPR. Both must comply with the requirements of Art. 6(1) GDPR. Furthermore, the recipient is subject to the information obligations resulting from Art. 14 GDPR. In addition, the principles relating to processing of personal data according to Art. 5 GDPR must be considered when establishing interoperability. Of importance in this context are the principles of good faith, data minimization, data security, and

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

transparency. They must be considered when designing the interoperability system. Which remains true for federated systems.

WhatsApp already uses FunXMPP, an XMPP-based protocol that enables federated communication. To send a message to a user of another messenger service, the sender must know the exact address of the addressee. Similar to an email address, this is composed of a username and a domain (Username@Domain.de). WhatsApp, for example, replaces the username with the respective cell phone number. All common XMPP servers provide functionalities that support end-to-end encryption of messages between multiple clients, e.g., with OMEMO or other extensions. Using end-to-end encryption, the operators of the messenger services only know the addressee, message type, and time of message transmission, but the content of the messages remains hidden.

2.3.4. Usability

Each messenger service can implement functionalities that other messenger services do not support. If these functionalities are partially interoperable, consumers may miss familiar functionalities. For example, if a messenger service of an addressee does not support video telephony, or if video telephony, in general, should not be part of the interoperability regulation, consumers might be confused why this normally familiar functionality is not available to them. Furthermore, sharing data with yet another provider could discourage users from participating in inter-messenger communications. Different messenger services embody different types of emotional proximity of the communicators.[13] Arnold et al[14] note that users distinguish which messenger service or type of communication they used to communicate with certain people. However, this is only possible if the respective addressees have the selected messenger service or type of communication. An interoperability obligation would eliminate the choice of messenger service depending on the addressee and emotional proximity and could thus limit the user experience. It is questionable at this point, though, whether this is to be understood as a user experience or

13 René Arnold and Anna Schneider, "An App for Every Step: A psychological perspective on interoperability of Mobile Messenger Apps", *28th European Regional Conference of the International Telecommunications Society (ITS)* (July/August 2017).

14 René Arnold et al., "Interoperability of interpersonal communications services – A consumer perspective", *Telecommunications Policy* 44, no. 3 (April 2020), https://doi.org/10.1016/j.telpol.2020.101927.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a necessity. The argument is supported by a report of the Federal Network Agency.[15] Here, 53% of the users surveyed stated that they did not see any need to be able to contact users of other messenger services directly. However, 45% of the respondents would like to have this functionality. 67% of the respondents, though, would not like to be contacted directly by users of other messenger services. This would mainly be due to the parallel use of different free messenger services, which do not make interoperability necessary. This hypothesis should be questioned, though, since users might find it difficult to imagine interoperability of messenger services and thus no accurate statement can be made regarding actual future use.

The ability to communicate with users outside the same messenger service without having to install a new service, therefore, does not yet seem to be explicitly desired by consumers. This can also be attributed to a lack of experience in this area, though, and would have to be analysed in further scientific studies in case of an interoperability obligation.

2.4. Result

The introduction of an interoperability obligation would break up any existing network effects. This would particularly benefit small or new companies, as it would make it easier for them to enter the market. However, other competitive advantages or USPs should be protected to ensure continued innovativeness.

Standardization restricts the innovativeness and thus also the resulting competition too much. Creating interfaces for all other services requires an increased administrative effort for each newly added messenger service. What all these federations have in common is that the number of connections to be realized between the systems involved is 2*N (N is the number of systems involved), i.e., it grows linearly with the number of systems. For interfaces, the number of connections to be realized is N*(N-1), so it grows polynomially with the number of systems. This also speaks for federations instead of interfaces. Therefore, a federated system avoids these disadvantages and creates a platform on which companies can on the one hand maintain their own strengths (their own extended protocols for intra-messenger communication) and on the other hand are open to other providers to mitigate network effects.

---

15 Bundesnetzagentur, *Nutzung von OTT-Kommunikationsdiensten*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

As described above, an XMPP-based, federated system offers the possibility of establishing interoperability between different messenger services. XMPP is open source, so it does not belong to any company, and it can be extended by further services or functionalities. This means that additional functionalities that are not part of the general XMPP standard can be added and used within a messenger (see FunXMPP). Interoperable functionalities must be incorporated into the standard XMPP protocol in order to be generally accessible and thus interoperable. Via gateways, XMPP partially allows communication with non-XMPP-based messenger services, so-called legacy services.[16] These gateways could also be used to exchange messages between domains with more advanced XMPP-based protocols, such as WhatsApp, and domains using the standard protocol. However, the creation of these gateways requires a noticeable implementation effort, which must be considered in the context of the proportionality assessment of an order of the interoperability obligation by the Federal Network Agency.

Users of a messenger service would only need to know the XMPP address of their contact to be able to contact him across domains. Existing apps such as Quicksy[17] or Zom[18] can already contact services of other providers outside their own domains that use XMPP. In order to be able to contact WhatsApp users as well, another gateway solution is required.

Federations have already been used successfully in other areas and thus can help messenger services to become interoperable. Examples include the conversion of geometry data of a CAD system (e.g., AutoCAD) into that of another (e.g., CADdy) via e.g., STEP (Standard for the Exchange of Product Model Data) or the data exchange of business data (orders, shipping notifications, invoices, etc.) via UN/EDIFACT (United Nations Electronic Data Interchange for Administration, Commerce and Transport).

16 Peter Saint-Andre and Dave Smith, "XEP-0100: Gateway Interaction", accessed June 23, 2021, https://xmpp.org/extensions/xep-0100.html.

17 “Have some quick conversations“, Quicksy, accessed June 23, 2021, https://quicksy.im/.

18 “Be in the Zom“, Zom, accessed June 23, 2021, https://zom.im/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3. Interoperability obligation according to the EKEK

3.1. Applicability of the EKEK to messenger services

3.1.1. Extension of the scope of application

It has long been disputed whether messenger and other OTT communications services are subject to the traditional European legal framework for electronic communications. In Germany, the discussion was triggered by a ruling of the Cologne Administrative Court (VG Köln) [19], which classified the webmail service Gmail as a telecomm5unications service within the meaning of § 3 No. 24 TKG (German Telecommunications Act of 22 June 2004).[20] The question of whether this legal assessment is compatible with the European legal framework was referred to the ECJ. The ECJ[21] rejected the functional understanding of the term "signal transmission" within the meaning of Art. 2(c) of the Framework Directive of 7 March 2002[22] advocated by the Administrative Court of Cologne, and thus the application of the EU legal framework to Web mail services. From a strictly technical perspective, according to the ECJ, signal transmission is carried out exclusively by Internet access and communications network providers. It was not sufficient for this characteristic to be affirmed "that the provider of the Internet service takes active steps in the sending and receiving of messages, whether by assigning to the e-mail addresses the IP addresses of the corresponding terminal equipment or by breaking down the messages into data packets and feeding them into the open Internet or receiving them from the open Internet so that they can be forwarded to their recipients." In its European Electronic Communications Code of 11 December 2018 (EECC)[23], the EU has reversed the trend and decided to base the definition of electronic communications services on a "more functional approach" to regulation. Such an understanding of the term could also cover services other than traditional services that enable communication. The background to this paradigm shift is the changes in user behaviour that have

19 *Verwaltungsgericht Köln*, Judgment of 11 November 2015, Az. 21 K450/15.
20 Telekommunikationsgesetz, Bundesgesetzblatt I 2004, 1190.
21 European Court of Justice, Judgment of 13 June 2019, *Case C-193/18*, ECLI:EU:C:2019:498.
22 Directive (EU) 2002/21, Official Journal of European Commission, L 108/33 of 24 April 2002.
23 Directive (EU) 2018/1972, Official Journal of European Commission, L 321/36 of 17 December 2018.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

been observed in recent years. The European Commission has not failed to notice that voice telephony, text messaging and e-mail transmission services are increasingly being replaced by online services with equivalent functionality, such as Internet telephony, messaging services and Web-based e-mail services. The central feature of the new definitional approach is the abandonment of the characteristic of signal transmission. In the future, electronic communications services are to include three types of services, some of which may overlap. According to Art. 2 No. 4 EECC an electronic communication service is:

> "a service normally provided for remuneration via electronic communications networks, which encompasses, with the exception of services providing, or exercising editorial control over, content transmitted using electronic communications networks and services, the following types of services: (a) 'internet access service' as defined in point (2) of the second paragraph of Article 2 of Regulation (EU) 2015/2120; (b) interpersonal communications service; and (c) services consisting wholly or mainly in the conveyance of signals such as transmission services used for the provision of machine-to-machine services and for broadcasting".

3.1.2. Messenger services as interpersonal communication services

Messenger services can obviously only fall into category b). According to Art. 2 No. 5 EECC, an "interpersonal communication service" is

> "a service normally provided for remuneration that enables direct interpersonal and interactive exchange of information via electronic communications networks between a finite number of persons, whereby the persons initiating or participating in the communication determine its recipient(s) and does not include services which enable interpersonal and interactive communication merely as a minor ancillary feature that is intrinsically linked to another service".

Recital 17 generally places messaging services in the category of interpersonal communications services. However, the most important messenger services available on the market must be examined to determine whether they meet the criteria of Article 2 No. 5 EECC.

With messenger services, users are firstly given the opportunity to reply. This means that communication is interactive. This feature distinguishes messenger services from linear services such as broadcasting, which ad-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

dresses its content to users as a one-to-many service. Secondly, no other person is involved in the exchange of information between these people. Communication is therefore direct and interpersonal. Third, the users themselves determine all the people involved in the communication process. For example, they must enter the telephone number as an identifier in order to reach the desired addressee.

Fourth, a finite number of people also participate in communication using messenger services. In the case of Telegram, there is the possibility to send messages to an unlimited number of users via the so-called "channels". However, this broadcasting function is only a partial function of a messenger service, whose main function remains the transmission of messages to a finite group of recipients. This is because the „sender“ of the communication content determines all participants. WhatsApp, for example, has a maximum group size of 256, which has been increased from 100. Fifth, the communication process takes place in messenger services via electronic communication networks. It is not decisive here that these are not networks of the messenger service providers, but of the Internet access service providers.[24]

Sixthly, it could still be questionable whether the characteristic of remuneration is present in the case of messenger services. This is because no direct monetary payments are made for these services. Rather, in the case of WhatsApp, Facebook Messenger and Skype, personal data is disclosed, or data is made available in return for the use of the services. According to recital 16, however, this should be sufficient to be able to regularly assume that a payment has been made. The Telegram and iMessage services also meet the remuneration criterion. Telegram is financed by donations. iMessage is part of the system software, which is paid for in the purchase price. According to the provider, no personal data is processed or sold for either service.

Finally, as a seventh characteristic, Art. 2 No. 5 EECC requires that the service to be assessed is a communication service in its main function. As an example of a service that only enables a subordinate secondary function, Recital 17 mentions a communication channel in online games. The purpose of this characteristic is to prevent over-extension of the scope of the directive and regulation. In particular, content providers are not to be covered, as can be seen from Art. 2 No. 4 EECC. Whether this characteristic is present can only be decided in individual cases. Facebook Messenger, for example, was included in the user interface of the Facebook portal for a

24 Directive (EU) 2018/1972, Recital 15.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

long time. Today, it is an independent service. The application can be downloaded via an app store. This should be sufficient to assign the service more than just a subordinate secondary function. With regard to the WhatsApp or Skype services, for example, there is no doubt from the outset that the requirements of Art. 2 No. 5 EECC are met.

Table: Characteristics of messenger services

Messenger service	Interactive communication	Direct communication	Direct addressing	Finite number of participants	Electronic communication networks	Remuneration	Mainly communication service
WhatsApp	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Facebook-Messenger	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Skype	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Threema	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Telegram	Yes	Yes	Yes	Yes	Yes	No	Yes
iMessage	Yes	Yes	Yes	Yes	Yes	System software	Yes

3.1.3. Types of interpersonal communication services

According to the EECC, there are two subcategories of interpersonal communications services: number-based and number-independent services. Since their distinction is of paramount importance for the application of numerous provisions of the EECC, the messenger services to be found on the market shall be assigned to these two categories in advance.

The decisive factor for the distinction is whether the respective service is either a

> "service which connects with publicly assigned numbering resources, namely, a number or numbers in national or international numbering plans, or which enables communication with a number or numbers in national or international numbering plans".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

If the answer to this question is in the affirmative, the service is a number-based interpersonal communications service in accordance with Art. 2 No. 6 EECC If the answer is no, the service is a number-independent interpersonal communications service in accordance with Art. 2 No. 7 EECC.

For classification purposes, it should be noted that the mere use of a number as an identifier cannot be equated with the use of a number to establish a connection with publicly assigned numbers. Only when the number is used to connect to a publicly assigned number is it a number-based interpersonal communications service. It is part of the regulatory course set by the EECC to subject number-based services to stricter requirement because they use publicly allocated numbering resources and establish end-to-end connectivity to end users via the (number) mechanism.

With regard to the messenger services examined here, only Skype opens up the possibility of reaching another end user via the public number space for a (small) fee. In this case, Skype is to be classified as a number-based interpersonal communications service. The other services examined here are to be qualified as number-independent interpersonal communications services. This also applies to Skype if the service is used in such a way that calls are only made between Skype users.

3.2. Interoperability of services according to EECC

3.2.1. Authorization to promote and ensure interoperability

Requirements for promoting and ensuring the interoperability of services can be found in Art. 61(1) EECC. The measures shall serve to achieve the objectives set out in Art. 3 EECC. The national regulatory authorities or the other competent authorities in the case of Art. 3(2) subpara. 1(b) and (c) are responsible for ordering them.

Interpersonal communications services are explicitly addressed in Article 61(2) subpara. 1(b) and (c) EECC. Notwithstanding any access obligations for companies with significant market power (cf. Art. 68 EECC), the authorities may take the measures specified in subparagraph 1(b) of the provision for number-based interpersonal communications services and the measures specified in subparagraph 1(c) for number-independent interpersonal communications services. The addition of "in particular" makes it clear that the measures listed are not exhaustive. It is already clear from the wording that the obligation is not part of asymmetrical regulation.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

3.2.2. Interoperability of number-based communication services

Pursuant to Art. 61(2) subpara. 1(b) EECC, the regulatory authorities may impose obligations on companies subject to a general licensee and controlling access to end users to make their services interoperable. These must be justified cases. In addition, the obligation can be ordered only to the extent necessary.

Which companies are subject to a general license can be seen from Art. 12 EECC (cf. Art. 15(1) EECC). The term "general permit" is misleading. The natural usage of the term suggests that it means the general permission of a certain activity. However, according to the legal definition in Art. 2 No. 22 EECC, it refers to "the legal framework" by which rights for the provision of electronic communications networks or services are guaranteed and in which sector-specific obligations are laid down. However, the introduction of a general license by Art. 3(2) of the Licensing Directive of 24 April 2002 eliminated the obligation, dating back to monopoly times, for companies to obtain an explicit permit or license from the regulatory authority before carrying out their activities and exercising their rights. They were to be bound only by the provisions of the regulatory framework. The European legislator hoped that this would strengthen the internal market. However, the member states were given the power to introduce a (declaratory) reporting obligation for companies subject to a general license. In this way, an overview of the players active in the market could be maintained. The EECC maintains this conception (cf. Art. 12(3) EECC, recitals 42 f.).

With regard to number-based interpersonal communications services, Art. 12(2) EECC clarifies that there must also be no authorization or license required for the providers of these services prior to commencing their activities. This is because the services "may only be made subject to general authorization." Special obligations may only be imposed with regard to the specifications mentioned in Art. 13(2) EECC and the rights of use mentioned in Arts 46 and 94 EECC.

However, there is no comparable regulation for number-independent interpersonal communications services and thus messenger services. This raises the question of whether their providers are subject to the regulations for general authorizations at all, such as a notification requirement. Recital 44 answers this question by stating that it is "not appropriate" to apply these regulations. This makes it particularly clear that the European legislator subjects number-independent interpersonal communications services to a lower level of regulation than is the case for number-based interpersonal communications services. They are to be subject to obligations only

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

if this is justified by a public interest. The reason given for this is that number-independent services do not benefit from the "use of public numbering resources" and do not participate in the "publicly secured interoperable ecosystem". This justification is not convincing, as it no longer does justice to the current economic and social significance of this category of services in comparison to number-based services. Nevertheless, an analogous application of Art. 13(2) EECC to number-independent services is out of the question. This is because, as recital 44 shows, the legislator has seen the regulatory problem, so there is no regulatory gap.

It can thus be stated that providers of number-independent interpersonal communications services are not companies subject to a general license. This means that ensuring the interoperability of messenger services in accordance with Art. 61(2) subpara. 1(b) EECC is generally ruled out from the outset. An exception applies only to messenger services such as Skype, insofar as the service uses publicly assigned numbers.

3.2.3. Interoperability of number-independent communication services

3.2.3.a). Regulatory approach

This does not mean, however, that providers of number-independent messenger services cannot in any case be required to make their services interoperable. However, the hurdle for this is significantly higher than is the case for number-based services. According to Art. 61(2) subpara. 1(c) of the EECC, the prerequisites for this are that, in a justified case, end-to-end connectivity between end users is threatened due to a lack of interoperability between interpersonal communications services and that the addressee of the obligation has a significant coverage and user base. These prerequisites and the possible legal consequences of ensuring interoperability are further specified in terms of content in Art. 61 EECC.

In procedural terms, a two-step approach is envisaged. First, the Commission determines which threats to connectivity in the internal market exist. On this basis, it also clarifies whether and with what instruments action can be taken to counter these threats. In a next step, the regulatory authority is responsible for deciding whether to take action in view of the national circumstances. In doing so, they must also be able to take action on their own initiative in order to ensure that the policy objectives listed in Art. 3 EECC are observed (cf. Art. 61(6) EECC). T This procedural sequence alone shows that the regulatory authorities have to overcome high hurdles if they want to impose an interoperability obligation. In this re-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

gard, however, it is an exaggeration to speak of a "regulation without teeth", because the assessment of whether the relevant factual prerequisites are met can change dynamically depending on market conditions.[25]

3.2.3.b). Threats to connectivity between end users

The European legislator's restraint with regard to number-independent interpersonal communications services is also shown by the fact that an interoperability obligation under Art. 61(2) subpara. 1(c) EECC, unlike in the cases of lit. a and lit. b, can only be considered if an "appreciable" threat to a regulatory objective of Art. 3 EECC can be identified. A higher danger threshold is required. End-to-end connectivity between end users must already be "threatened" by a lack of interoperability between interpersonal communications services.

End-to-end connectivity between end users is ensured when there is the possibility of communication between the end users. In the English-language version of Art. 61(2) subpara. 1(c) EECC, this classic task of telecommunications is vividly described when it speaks of "end-to-end connectivity between end-users". This terminology is similar to the "end-to-end interconnection of services" formula used in Article 5 (1) (2) (a) of the Access Directive of March 2002[26], which is replaced in the EECC by the phrase "end-to-end connectivity". However, end-to-end connectivity between end users requires that the systems and technologies used are interoperable. That is, they must be capable of working together and exchanging information with each other or making it available to the user as efficiently as possible. Achieving interoperability is therefore also one of the classic objectives of telecommunications law. The European legal framework therefore has a number of instruments, such as the specification of an interface for the end-to-end connection or the standardization of technical standards, to ensure that this objective is achieved.[27]

However, it is questionable when a threat to connectivity can be assumed. The literature is cautious in this regard. Certainly, it cannot simply be pointed out here that conventional voice telephony in PSTN mode

---

25 Stefan Bulowski, *Regulierung von Internetkommunikationsdiensten. Zur Anwendbarkeit des Telekommunikationsrechts auf Voice over IP, Instant Messaging und E-Mail-Dienste* (Baden-Baden: Nomos, 2019).

26 Directive (EU) 2002/19., Official Journal of European Commission, L 108/7 of 24 April 2002.

27 Directive (EU) 2018/1972, Recital 148.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

provides the necessary connectivity between end users. After all, this is not an interpersonal communications service. However, as can be seen from the wording, the provision is concerned precisely with ensuring the interoperability of these services ("lack of interoperability between interpersonal communications services"). The reference to the possibility of multihoming, i.e., the frequently observed parallel use of several number-independent interpersonal communication services such as WhatsApp and Facebook Messenger, does not lead anywhere either.[28] This is true even if a limit on the reasonableness of the available multihoming service is read into the law in the event of serious data protection concerns. This is because it remains the case that each of these services is proprietary in its own right and does not have an end-to-end connection with another of these services. Recital 149 therefore also comments exclusively on interpersonal communications services: With regard to this category of services, end-to-end connectivity is "currently" present because end-users use number-based interpersonal communications services. However, it could not be ruled out that "future technical developments or increased use of number-independent interpersonal communications services" would lead to a significant threat to connectivity between end users. This could result in significant market entry barriers and obstacles to further innovation.

The latest market analysis by the Federal Network Agency indicates that the frequency of use of interpersonal communications services in Germany has already shifted sharply in the direction of number-independent services. WhatsApp is the most-used service by a wide margin at 85.4%. Facebook Messenger follows this with 4% and Instagram with 3.3%. Only then comes Skype with 1.3%. This is an interpersonal communication service that can (also) be used on a number-dependent basis. These data indicate that the market development referred to in recital 149 has already occurred in Germany. However, this alone is not sufficient to justify an interoperability obligation.

Rather, if interoperability problems arise, the procedure provided for in Art. 61(2) subpara. 2(ii). EECC must be followed. In this case, the Commission is first required to have BEREC assess the situation at the Union and Member State level. On the basis of this report, the Commission must then decide whether regulatory intervention by the regulatory authority is necessary. However, according to Art. 61 (2) subpara. 2(ii). EECC, this can

28 Jürgen Kühling, "What to do with OTT? - Die Regulierung von Gmail, WhatsApp & Co. de lege ferenda", in *Regulierung – Wettbewerb – Innovation*, ed. Torsten Körber and Jürgen Kühling (Baden-Baden: Nomos, 2017), 181.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

only be considered if end-to-end connectivity between end users throughout the Union or in at least three Member States is threatened to a significant extent. If such intervention is contemplated, the Commission should, as a next step, adopt implementing measures specifying the nature and scope of any regulatory measures. For this purpose, an examination procedure pursuant to Art. 118(4) of the EECC is to be carried out.

3.2.3.c). Providers with significant coverage and user base

The addressees of an interoperability obligation can only be providers of number-independent interpersonal communications services that have "significant coverage and user base". According to Recital 151, notable should mean that the geographical coverage and the number of end users ensure a "critical mass" with regard to the objective of end-to-end connectivity to be achieved. Accordingly, interoperability obligations should not apply as a rule if providers with a limited number of end users or limited geographic coverage can make "only a marginal contribution" to achieving this objective. The market data of the Federal Network Agency suggest that the requirements are met for Germany, at least for the WhatsApp service belonging to Facebook.[29]

However, the regulatory authorities are not to determine whether a provider has significant market power within the meaning of Art. 63 et seq. EECC. This is because the interoperability obligation is designed as a symmetrical regulatory measure, as are the other possible orders to be imposed under Art. 61 EECC. This is indicated by recital 157, which states that obligations to ensure connectivity and interoperability could be imposed "irrespective of the designation as an undertaking with significant market power". This is confirmed by the systematic position of Art. 61 in the EECC. The provision is located in the chapter on "Access and Interconnection" (Arts 61 f. EECC), but not in the chapter on "Access Obligations for Companies with Significant Market Power" (Arts 63 ff. EECC.).

3.2.3.d). Scope of the obligation

With regard to the legal consequences of an order issued by the regulatory authority, Art. 61(2), (5) sentence 1 EECC emphasizes the principle of pro-

29 Bundesnetzagentur, *Nutzung von OTT-Kommunikationsdiensten*, 16.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

portionality, which also applies elsewhere in European law. Interoperability obligations may only be imposed to the extent "necessary" to ensure end-to-end connectivity between end users (subparagraph 1(c)) or may not exceed the "extent necessary" for this purpose (subpara. 2(i)). The regulatory authority may also only intervene in "justified cases". The objective of proportionality is also explicitly mentioned in (Art. 61(5) p. 1 EECC).

There is also a requirement that regulatory measures must be "objective, transparent, and non-discriminatory". These are also general requirements of access regulation under telecommunications law. The application of these criteria is governed by the procedures set out in Articles 23, 32 and 33 of the EECC (Art. 61(5) sentence 1 EECC). As part of the notification procedure governed by these provisions, the Member States shall ensure that the European Commission and the national regulatory authorities are informed of the intended obligation and are given the opportunity to comment on it.[30]

The evaluation obligation of the regulatory authorities in Art. 61(5) sentence 2 EECC is also to be understood as an expression of the principle of proportionality. According to this, they must review the results of the obligation and condition within five years of the measure's enactment and whether its amendment or repeal would be appropriate in light of changing circumstances. The results of this review must be announced (Art. 61(5) sentence 3 EECC).

Art. 61(2)(i) EECC allows the regulatory authority to attach conditions to the interoperability obligation. The provider concerned may be required, in order to ensure the interoperability of interpersonal communications services, to publish "relevant information" itself or to authorize its use, modification and further dissemination by public authorities or other providers. In this way, guidance can be provided so that, as stated in Art. 61(1) sentence 2 EECC, small and medium-sized enterprises and operators with a limited geographical reach can benefit from the obligations imposed.

In addition, the provider concerned may be required to use and implement in practice standards and specifications listed in the directory referred to in Art. 39(1) EECC or other relevant European or international standards. According to Recital 148, Member States shall encourage the use of the published standards or specification for the provision of services, technical interfaces or network function as strictly necessary to ensure the interoperability of services.

30 Directive (EU) 2018/1972, Recital 157.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 4. Conclusion

The interoperability of messenger services would allow users of different service providers to exchange messages, photos, videos, and many other data formats across domains. The concept of a federated system, specifically the use of the XMPP protocol, was presented as one form of a technical design option. XMPP was chosen because it is already used by WhatsApp (in a modified form), which would simplify interoperability. However, other protocols, e.g., Matrix Protocol (which offers a bridge to XMPP), are also suitable for an interoperable design. This concept was analysed in terms of its impact on competition, innovativeness, data privacy, and usability. In summary, the following can be stated for these four points:

Competition is likely to benefit significantly from an interoperability obligation, as network effects are reduced.[31] This can be advantageous for smaller existing messenger services as well as facilitate the market entry for new developments. The unique selling point is thus no longer the size of the user base, but the extensiveness of the functionalities.

Innovativeness should not be restricted by an interoperability obligation based on a federated approach. With this approach, companies remain free to use their own (XMPP-based) protocols. Thus, no functionalities are lost. However, each company must create a gateway that can also interact with other domains based on the standardized XMPP protocol to enable data exchange.

Since at least two messenger services (more than two in the case of group messages) are involved for cross-domain data exchange, personal metadata is usually generated twice. From a data privacy perspective, however, this does not pose a problem, as the General Data Protection Regulation has established a sufficient level of protection regarding the generated personal data.

Users do not seem to see any clear advantage in an interoperability obligation so far. Thus, no further benefits are expected to arise from a usability perspective, although sending messages to multiple messengers does mean an increase in functionality.

From the perspective of consumer convenience, the introduction of a federated system proves to be the most reasonable solution, as the competition is strengthened, and innovation and usability are not restricted. However, meta-information will always be shared with multiple providers, but this should not be a problem from a data privacy perspective.

---

31 Bundesnetzagentur, *Nutzung von OTT-Kommunikationsdiensten*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

In legal terms, the European and German legislators have opted for the possibility of an interoperability obligation. The Federal Network Agency is responsible for issuing such an order. It can issue this order if connectivity between end users is threatened due to a lack of interoperability between interpersonal telecommunications services. Providers of number-independent interpersonal telecommunication services that have a significant coverage and user base can be considered as addressees. For Germany, a study by the Federal Network Agency on OTT communication services in Germany suggests – as mentioned - that these conditions are met for the messenger services belonging to Facebook. However, the interoperability obligation can only be imposed if the European Commission has taken the necessary enforcement measures beforehand and a planned order by the Federal Network Agency is in line with this. In addition, the principle of proportionality must be applied in each individual case, so that the issues discussed (competition, innovativeness, data privacy, and usability) must once again be weighed against each other.

Bibliography

Arnold, René, and Anna Schneider. "An App for Every Step: A psychological perspective on interoperability of Mobile Messenger Apps." *28th European Regional Conference of the International Telecommunications Society (ITS)* (July/August 2017).

Arnold, R., A. Schneider and J. Lennartz. "Interoperability of interpersonal communications services – A consumer perspective." *Telecommunications Policy* 44, no. 3 (April 2020). https://doi.org/10.1016/j.telpol.2020.101927.

Bitkom. "Zwei von drei Internetnutzern verwenden Messenger." Accessed June 23, 2021. https://www.bitkom.org/Presse/Presseinformation/Zwei-von-drei-Internetnutzern-verwenden-Messenger.html.

Bulowski, Stefan. *Regulierung von Internetkommunikationsdiensten*, *Zur Anwendbarkeit des Telekommunikationsrechts auf Voice over IP, Instant Messaging und E-Mail-Dienste.* Baden-Baden: Nomos, 2019.

Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen. *Nutzung von OTT-Kommunikationsdiensten in Deutschland*. Bonn: Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2020. https://www.bundesnetzagentur.de/SharedDocs/Mediathek/Berichte/2020/OTT.pdf?__blob=publicationFile.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Digitale Gesellschaft e.V. "Stellungnahme der Digitalen Gesellschaft e. V. zur Konsultation des Bundesministeriums der Justiz und für Verbraucherschutz zu Interoperabilität und Datenportabilität bei sozialen Netzwerken." Accessed June 23, 2021. https://digitalegesellschaft.de/2019/05/stellungnahme-der-digitalen-gesellschaft-e-v-zur-konsultation-des-bundesministeriums-der-justiz-und-fuer-verbraucherschutz-zu-interoperabilitaet-und-datenportabilitaet-bei-sozialen-netzwerken/.

Katz, Michael L., Carl Shapiro. "Systems Competition and Network Effects." *Journal of Economic Perspectives* 8, no. 2 (Spring 1994): 93-115. https://doi.org/10.1257/jep.8.2.93.

Kühling, Jürgen. "What to do with OTT? – Die Regulierung von Gmail, WhatsApp & Co. de lege ferenda." In *Regulierung – Wettbewerb – Innovation*, edited by Torsten Körber and Jürgen Kühling, 165-185. Baden-Baden: Nomos, 2017.

Quicksy. „Have some quick conversations." Accessed June 23, 2021. https://quicksy.im/.

Saint-Andre, Peter and Dave Smith, "XEP-0100: Gateway Interaction." Accessed June 23, 2021. https://xmpp.org/extensions/xep-0100.html.

Wegner, Peter. "Interoperability." *ACM Computing Surveys* 28, no. 1 (March 1996). 285-287. https://doi.org/10.1145/234313.234424.

Zom. „Be in the Zom." Accessed June 23, 2021, https://zom.im/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Six Problems with Facebook's Oversight Board. Not enough contract law, too much human rights.

Mårten Schultz

Abstract: After intense criticism against Facebook's content moderation process, CEO Mark Zuckerberg stated in 2018 his intention to set up a "Supreme Court" for the company. In January 2021 the idea became reality when Facebook's Oversight Board started reviewing complaints against Facebook's decisions. While there are reasons to be hopeful that the Oversight Board will turn out to be a positive step forward in the discussion on online speech governance, there are also reasons to be worried. This article addresses six problems with Facebook's Oversight Board in its current form.

Keywords: Oversight Board, Facebook, content moderation, self-regulation, community standards.

Chapter 1. Introduction

1.1. Background

In 2018, Facebook's CEO Mark Zuckerberg first presented the idea. An independent institution would be given the task of reviewing appeals against Facebook's content moderation decisions. "You can imagine some sort of structure, almost like a Supreme Court, that is made up of independent folks who don't work for Facebook, who ultimately make the final judgment call on what should be acceptable speech in a community that reflects the social norms and values of people all around the world." It took some time, more time than Facebook initially thought would be

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

needed.[1] Facebook's Oversight Board (OB/the Board) opened for business in January 2021, after a couple of years of preparation.[2]

Facebook is, arguably, the most important catalyst for freedom of expression in human history. When Facebook set up an independent institution and gave it the power to overrule its decisions and build its own "case law" it also established the most influential arbitrator of expression in human history. That alone is cause for concern. There are other reasons to be concerned as well. This article puts forward six problems with the OB, as it has developed in its still early stage.

Before getting on to these at times critical arguments, I want to make clear that my perception of the process behind the OB is that it was formed with the best intentions and that the first line of people that have been put in charge of the project have the best of credentials. There are reasons to be hopeful that the OB will turn out to be a starting point in the development of a new kind of institutions that can tackle the balancing act between different interests and rights in social media.[3] This makes it even more important to early on address issues where the project seems to be taking a bad turn.

1.2. The Oversight Board: A very brief description

Facebook is one of the world's largest companies. It controls not only the Facebook social media platform but also Instagram and Whatsapp (and other companies as well).

The OB is an independent legal person that was set up by Facebook.[4] The function of the OB is to enable Facebook and Instagram users to "appeal" decisions made by the company regarding content on the platform, such as decisions to remove posts that Facebook moderators have found

1 Kate Klonik, "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression", *The Yale Law Journal* 129 (2020): 2450.

2 Transparency: in Berlin, June 2019, I participated in one of the six brainstorming meetings Facebook organised around the world in the process of setting up the OB, and thereafter expressed interest in being a member of the OB. My main grievance, however, is that I failed to convince Facebook to place the headquarters of the Board in Stockholm.

3 *See* for a sympathetic take on the value of the project Evelyn Douek, "Facebook's "Oversight Board": Move Fast with Stable Infrastructure and Humility", *North Carolina Journal of Law and Technology* 21 (2019): 7.

4 Klonik, "Facebook Oversight Board", 2481-2487 (Discussing different kinds of independence criteria with regard to the OB).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to be in violation of the Community Standards.[5] In April 2021, the OB also started to take on cases where users appealed decisions *not* to remove content.[6]

The Oversight Board Charter ("the Charter") is the foundational steering document for the OB.[7] The Charter makes clear that a case can be submitted to the OB either by a user or by Facebook itself (which is how the decision to remove president Donald Trump from the platform came before the Board). It is up the Board to decide which cases it should take on, but the Charter states that it should prioritize cases "that have the greatest potential to guide future decisions and policies."[8]

A trust has been set up to oversee the financing of the Board and to safeguard the independence of the Board. (The OB itself is a limited liability company based in Delaware.) The trust also oversees administration of the Board.

According to the Charter, the OB must include at least 11 members and, when fully staffed, is "likely to be forty members."[9] The members work part time for the OB, and are paid for their work. Facebook has allocated 130 million dollars to the trust to fund the board.[10]

In the Charter Facebook commits "to the board's independent oversight on content decisions and the implementation of those decisions."[11] The Board not only has the power to overrule Facebook decisions regarding content on the platform. It can also make advisory statements on Facebook/Instagram policy.[12] Facebook can choose whether to follow these recommendations or not.

---

5 Hereinafter Facebook should be understood as a short term for Facebook and Instagram.

6 "The Oversight Board is accepting user appeals to remove content from Facebook and Instagram", Oversight Board, accessed June 2, 2021, https://oversightboard.com/news/267806285017646-the-oversight-board-is-accepting-user-appeals-to-remove-content-from-facebook-and-instagram/.

7 See "Trustees", Oversight Board, accessed June 2, 2021, https://oversightboard.com/governance/.

8 Art. 2, sect. 1.

9 Art. 1, sect. 1.

10 Kate Klonik, "Inside the Making of Facebook's Supreme Court", *The New Yorker*, February 12, 2021.

11 Art. 5, sect. 3.

12 The process is pictured in "Rulebook for Case Review and Policy Guidance", Oversight Board, accessed June 2 2021, https://oversightboard.com/sr/rulebook-for-case-review-and-policy-guidance. This opportunity was used already in one of the first decisions, 2020-003-FB-UA (2021-01-28).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Charter provides the structure and basic rules, but it is supplemented by other documents. More detailed procedural guidelines are found in the Oversight Board Bylaws.[13] In addition, there is a Rule Book for Case Review and Policy Guidance.[14]

An outline of the arguments of this article

This article describes six problems with the OB as it has developed. These problems are partly intertwined. Under the heading "The Narrative" I criticize the use of a public law narrative, especially the language of human rights, in the discussion of content moderation. "The Bias" argues that the OB has a bias in favour of freedom of speech arguments, which may have negative effects on Facebook's legitimate interest to control content on its platform. In "The rules" I question the OB's choice of the sets of norms that are used in its decision-making. "The process" discusses whether Facebook and the OB has missed an opportunity to give all Facebook users access to an appeals process, to instead focus on producing guiding decisions. In "The decisions" I wonder whether a policy to highlight differences in opinions between board members in the Board's decisions, instead of aiming at consensus, would better promote the purpose of providing guidance. Lastly, "The power shift" asks whether transfer of power of content moderation decisions to a small group of experts is dangerous.

Chapter 2. The narrative

When Mark Zuckerberg first floated the idea of establishing an external institution that would have the capacity to independently review decisions by Facebook, he referred to it as Facebook's "Supreme Court".[15] The me-

13 "Oversight Board Bylaws", Oversight Board, accessed June 2, 2021, https://oversightboard.com/sr/governance/bylaws.

14 "Rulebook."

15 Ezra Klein, "Mark Zuckerberg on Facebook's hardest year, and what comes next", *Vox*, April 4, 2018, https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

dia quickly caught on.[16] Everybody knows that the OB is not a court at all. Still, many use the description as a metaphor even today.[17]

In this context it is not necessary (or possible) to explain what a court is, but a simple description of what characterizes a court in a modern *Rechtstaat* illustrates why the label is misleading also as a metaphor. A court is, at least, an institution within a national state that exercises public authority. The OB is nothing of the sort. Its scope is narrow (content moderation decisions by Facebook), its authority is narrow (it can decide either that Facebook needs to put back content it has removed or that its decision stands) and it lacks the possibility to exercise any public power.

The court metaphor is part of a larger narrative.[18] Companies such as Facebook – but especially Facebook – have for some time been described as nation-like entities, in Facebook's case under labels such as Facebookistan.[19] The company's representatives are partly to blame for this. More than 10 years ago Mark Zuckerberg described Facebook as something more like a government than a traditional firm.[20] As a result of this narrative, private law issues – such as questions about the contractual relationship between companies and their customers – are discussed in the language of public law.

A particular and important example is how the terminology of fundamental human rights and freedoms is employed: the question of content moderation has taken the form of a human rights problem. As will be discussed below, the OB and Facebook are parts of the explanation for this language. This kind of language is often misleading and perhaps harmful.

Whether fundamental rights and freedoms have a role to play in private law relationships is one of the most debated questions in private law

16 *Cf.* Casey Newton, "Why Facebook needs a Supreme Court for content moderation", *The Verge*, August 21, 2018, https://www.theverge.com/2018/8/21/17762354/facebook-supreme-court-content-moderation.

17 See, e.g., Kate Klonik, "Inside the Making of Facebook's Supreme Court", *The New Yorker*, February 12, 2021 and Klonik, "Facebook Oversight Board," 2476 ("The analogy to courts is valuable, but also imperfect."). The parable is used in Sweden as well, Anni Carlsson, "Tyst vår?," *Svensk Juristtidning* (2021): 170.

18 Evelyn Douek calls the OB "one of the most ambitious *constitution-making* projects of the modern era", Douek, "Facebook's "Oversight Board", 1 (Emphasis added.).

19 Anumap Chander, "Facebookistan", *North Carolina Law Review* 90 (2012): 1807.

20 See David Kirkpatrick, *The Facebook Effect* (New York: Simon & Schuster, 2010), 254.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

in recent decades, especially in tort law.[21] The most contested issue in this context is whether private entities (companies and persons) could be held responsible under human rights rules, an issue discussed under the heading of "horizontal human rights" or "direkte Drittwirkung".[22] It has also been a hot topic in international law.[23]

However, to my knowledge, there are no examples in any jurisdiction of direct application of a general human rights catalogue as a basis for duties of private companies. There are examples of constitutions that apply human rights law to (humans and) companies, but only in a limited sense.[24]

Furthermore, there is a risk of an intellectual fallacy here. A company's duty to distribute another person's piece of a information, is also a limitation of that company's (or its owners') right to decide what information it wants distribute.[25] Nuance and detail are thus necessary if one wants to frame responsibilities of a company in human rights language.

The public law narrative in general, including the sweeping usage of the language of fundamental human rights and freedoms, is dangerous in two different ways. Firstly, it is dangerous because it suggests that Facebook has special duties that other companies do not have; that for some reason it

---

21 A fresh example from Sweden is Karolina Stenlund, *Rättighetsargument i skadeståndsrätten* (Uppsala: Iustus, 2021). See also Mårten Schultz, "Rights Through Torts," *European Review of Private Law* 17, no 3 (2009): 305 ff.

22 To take Sweden as an example, the Supreme Court shut the door on a direct application of human rights rules as a direct basis for holding a private company liable in tort in Högsta Domstolen, NJA 2007, 747. However, in 2015 the Supreme Court stated human rights rules may in some circumstances affect the assessment of a private party's obligation to compensate for pure economic loss (an indirect horizontal effect of human rights), Högsta Domstolen, NJA 2015, 899. See also Håkan Andersson, *Ansvarsproblem i skadeståndsrätten* (Uppsala: Iustus, 2013), 618 ff., Jan Kleineman, "Konstitutionell skadeståndsrätt", *Juridisk Tidskrift* (2018-19): 23 ff., and Mårten Schultz, "Nya argumentationslinjer i förmögenhetsrätten: Rättighetsargument", *Svensk Juristtidning* (2011): 996 ff. (All discussing horizontal applications of the European Convention of Human Rights and Fundamental Freedoms.)

23 See, e.g., Andrew Clapham, *Human Rights Obligations of Non-State Actors* (Oxford: Oxford University Press, 2006) and John H. Knox, "Horizontal Human Rights Law", *American Journal of International Law* 102 (2008): 1.

24 *Cf*. art. 8 of the Bill of Rights in the South African constitution.

25 There has been a debate on whether the social media giants should follow under some kind of must carry obligations. See for an early discussion on must carry obligations and digital publications European Audiovisual Observatory, *To Have or Not to Have Must-Carry Rules* (Strasbourg: European Audiovisual Observatory, 2005), https://rm.coe.int/168078349b.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

should be treated fundamentally differently than, say, Tesla, IKEA or Pindo's Pizzeria in Ösmo outside of Stockholm. A common argument for this standpoint is Facebook's size and dominance. A company that dominates a market may have obligations under anti-trust or consumer legislation, for instance. However, if there is no legislation that states something else, then it is the contract that sets up the rules. This obvious starting point is too often missing or underestimated in the debate on tech companies' content moderation.

From the perspective of the company, there is a risk that this narrative may have negative effects on the right to property. An owner of property, for instance the owner of a company, has a fundamental right to use her property any way she likes. The law may set limitations but such limitations are only acceptable under some conditions, for instance "in so far as is necessary for the general interest" (to use the formulation in art. 17 of the European Union Charter on Fundamental Rights).

Secondly, it is dangerous to treat a private company as a state because it suggests that it has rights which it does not have. Statehood comes with privileges. One privilege is sovereignty. One facet of sovereignty is the right to control the law within a territory. But Facebook does not have the power to control the rules that govern its platform. States, and sometimes international bodies such as the European Union, control the law, not companies. It is sometimes difficult to ascertain which country's rules apply and which country's courts have jurisdiction. In the case of the big tech companies there is also, from a practical point of view, a complication in the fact that platforms have the possibility to unilaterally formulate dispute resolution clauses in the contract with users. Nevertheless, the law – in the true sense of the word – is written by legislators and in some countries the courts, not companies. Even if they are wealthy and have global reach.

Chapter 3. The bias

"Freedom of expression is a fundamental human right. Facebook seeks to give people a voice so we can connect, share ideas and experiences, and understand each other.

Free expression is paramount, but there are times when speech can be at odds with authenticity, safety, privacy, and dignity. Some expression can

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

endanger other people's ability to express themselves freely. Therefore, it must be balanced against these considerations."[26]

The quote is taken from the preamble to the Charter. Freedom of expression is indeed a fundamental human right. But so is, for instance, the right to respect for private and family life, the right to property and many other interests. If one takes a look at the European Union's Charter on Fundamental Rights there are several rights that will often conflict with freedom of expression, such as the right to protection of personal data.[27]

The idea that freedom of speech is in some way more fundamental than other freedoms and rights is associated with the constitutional tradition in the United States.[28] European countries, on the other hand, do not generally consider that freedom of speech *a priori* weighs heavier than other rights and freedoms.[29] Sometimes freedom of speech outweighs privacy. Sometimes it is the other way around.

Comparative law observations aside, it is clear that the OB is based on a bias in favour of facilitating speech. This follows from the quoted mission statement in the Charter. Moreover, the same sentiment is iterated in the other steering documents that govern the Board. The introduction to the Bylaws starts off with the following sentence: "The purpose of the Oversight Board is to protect freedom of expression by making principled,

---

26 "Trustees".

27 As Maroussia Lévesque points out, "The Board's narrow focus on freedom of speech excludes other pertinent human rights", "Applying the UN Guiding Principles on Business and Human Rights to Online Content Moderation", Maroussia Lévesque, accessed June 2, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3789311.

28 "Applying the UN Guiding Principles on Business and Human Rights to Online Content Moderation." See, for a strong case in favour of setting freedom of speech protection at the centre against an analysis of international law, Evelyn Mary Aswad, "To Protect Freedom of Expression: Why Not Steal Victory from the Jaws of Defeat", *Washington & Lee Law Review* 77 (2020): 609.

29 There is a large body of literature comparing US and "European" freedom of speech traditions. See, e.g., Sionaidh Douglas-Scott, "The Hatefulness of Protected Speech: A Comparison of the American and European Approaches", William & Mary Bill of Rights Journal 7 (1999): 305 (focusing on hate speech). This characterization is oversimplified. For instance, it does not hold in a comparison between constitutional protection of speech in Sweden and the USA. Arguably, Sweden has the strongest protection of free speech in the media in the world, if one considers both substantive as well as procedural rules. See Mårten Schultz, *Det här får man inte säga i det här landet!* (Stockholm: Stiftelsen Juridisk Fakultetslitteratur, 2021), 11.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

independent decisions about important pieces of content and by issuing policy advisory opinions on Facebook's content policies."[30]

The OB thus has protection of freedom of expression as its primary goal.[31] This is an unfortunate formulation. The Board here uses the term in the way Facebook's critics have often used it, when the company is accused of "censorship". Removal of content by Facebook restricts the possibility to reach other people but it is not a restriction of freedom of speech. It may be a breach of contract, if Facebook has failed to follow the terms of the agreement, but it is not censorship.

It is also unfortunate because this bias entails that Facebook's legitimate interest in excluding different types of content from its platform is undermined. It is perfectly legitimate to want to exclude nudity, profanity, hate speech, false information and pictures of snakes, even if this means that the platform excludes information that may be legally published in every country on the planet. When the Board taps into the language of freedom of speech and thereafter, in its first batch of decisions, overrides most of Facebook's content moderation decisions (of which none were clearly in conflict with the terms of service) it sent a signal, "When in doubt: restore".[32]

Most of all, however, it is unfortunate because it is questionable to assign freedom of speech – or any fundamental (negative) human right or freedom – a *general* priority.[33] The issue whether there is a hierarchy of

---

30 "Oversight Board Bylaws." The Rulebook expresses it somewhat differently in its introduction: "The Oversight Board was created to make principled, independent, and binding decisions on what content Facebook and Instagram should allow or remove, based on respect for freedom of expression and human rights", "Rulebook".

31 *Cf*. Klonik, "Facebook Oversight Board", 2475.

32 The decisions are published on "Board Decisions", Oversight Board, accessed June 2 2021, https://oversightboard.com/decision/. A good example of this is decision 2021-005-FB-UA (2021-05-20), which dealt with a Turkish meme that questioned the Armenian genocide.

33 This assertion rests on a distinction between negative and positive human rights and freedoms, which rests upon Isaiah Berlin's famous dichotomy of negative and positive concepts of liberty. (Isaiah Berlin, *Four Essays on Liberty* (Oxford: Oxford University Press, 1969). This distinction has been the subject of lively political, moral and conceptual debate, but in this context a short description will have to suffice. Negative human rights and freedoms oblige someone (typically the Government) to not act so that another person's freedoms are restricted. To take freedom of speech as an example, this right protects any person from being actively silenced by the government, or from being punished for speaking. A negative right does not, however, oblige the government to act to make sure that

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

human rights has been debated.[34] However, in decision-making such as the one that the OB is involved in, which necessarily involves weighing interests against each other, a presumption in favour of one of these interests may have a negative effect. If my speech may risk causing another person's death it makes no sense to view my right to expression as a *prima facie* prioritized right over the other person's right to life. In cases involving a conflict of rights or freedoms, or interests of this kind, a decision maker must or at least should aim at neutrally weighing the interests against each other taking into account the circumstances of the individual case.

Chapter 4. The rules

The relationship between Facebook and its users is contractual. When conflicts arise between two parties to a contract the first question is: "What does the contract say?" When a decision maker, for instance a judge, settles a contractual dispute the starting point of the analysis is always the set of rules that forms the contract. Only in special circumstances will the decision maker need to set aside that term of the contract, for instance if it does not meet the requirements of consumer protection laws or if it is discriminatory. There are thus situations in which "external" rules enjoy priority over the "internal" rules in the contract. Still the main rule is that the contract applies and exceptions are only made if there is a clear legal rule that says otherwise.

The OB has taken another path. Already in the first decisions it became clear that the Board uses three sets of norms in its handling of cases:

---

everyone can be heard. In the category of negative rights we thus find the provisions of the European Convention of Human Rights. A positive right, on the other hand, obliges someone, often the Government, to act to help someone get or achieve something. To take an example from the freedom of speech sphere in Sweden, the Swedish constitutional Freedom of the Press Act includes arguably the world's most far-reaching obligation to disclose public documents. More often, perhaps, positive rights are thought of as social rights, such as the right to education and medical treatment. Many have been critical of the distinction between positive and negative rights (see, e.g., Henry Shue, *Basic Rights* (New Jersey: Princeton University Press, 1996, Second Edition)). In this context – which focuses on the obligations of a private company and not a government – I will presuppose that the distinction is helpful and indeed necessary, rather than arguing for it.

34 *Cf.* Tom Farer, "The Hierarchy of Human Rights", *American University International Law Review* 8 (1992): 115.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Facebook's Community Standards, Facebook's values, and international human rights law.

Facebook's Community Standards are part of the terms of service in the contract between Facebook and its users. The Community Standards include rules against violence and incitement, bullying and harassment, and hate speech, to give a few examples.

The introduction to the Community Standards states that Facebook limits expression "in service of one or more of the following values": "Authenticity", "Safety", "Privacy", and "Dignity". These values make up a set of general principles that the more specific Community Standards rest upon and make up a second set of norms that the OB apply in its decision making.

The third norm source used by the OB comes from international human rights law.[35] The OB uses the formulation "Relevant Human Rights Standards considered by the Board". More specifically, the Board refers to "The UN Guiding Principles on Business and Human Rights (UNGPs)" which were endorsed by the UN Human Rights Council in 2011. These principles establish "a voluntary framework for the human rights responsibilities of private businesses".[36]

There are, at least, two problems with this selection of normative sources. The first problem is that it does not take sufficient account of the priority of the contract. When someone sets up an account with Facebook a contract is formed. The contract includes different terms that the parties agree upon. These terms include the community standards but also Facebook's values, but not any reference to the UNGPs. When a dispute between Facebook and a user is resolved under principles of human rights law it means not only that Facebook's actions are tested against a normative framework it has not accepted but also that the decision maker overrides the rules that both parties had agreed upon. The inclusion of human rights principles in the OB's set of rules thus amounts, in a way, to disregard of the will of both Facebook and its users as expressed through the contract.

A second problem with the norm sets the OB has chosen is unpredictability. It is often not too difficult to assess if a post adheres to the Community Standards or not. We know, for instance from Facebook's

35 See, for arguments for using international human rights law in the OB, Aswad, "Freedom of Expression", 609.

36 "Applying the UN Guiding Principles on Business and Human Rights to Online Content Moderation."

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

experiences of handling content with nudity, that there will always be difficult cases. In most cases, however, it is not too difficult to foresee how the Community Standards would be interpreted in a particular case. In contrast, it is much more difficult to predict the result of an interpretation based on Facebook's general values or human rights principles.

Chapter 5. The process

> "I think in any kind of good-functioning democratic system, there needs to be a way to appeal."[37] This statement comes from Mark Zuckerberg, in one of the earlier interviews in which he talked about the need for independent judicial review. Zuckerberg later wrote, in an open letter in connection with publication of the Charter: "If someone disagrees with a decision we've made, they can appeal to us first, and soon they will be able to further appeal to this independent board."[38]

One of the purposes of the OB was to provide Facebook users with a channel to voice their dissatisfaction with the company's decisions, for instance a decision to take a post down. If a moderator at Facebook unfairly decides to remove a picture that someone has published in a Facebook group, the OB is able to overrule and correct the decision. The OB is thus, in a way, supposed to provide access to justice.

When this is written, in April 2021, more than 220 000 complaints have been appealed to the Board.[39] Only a few cases have been decided. It is clear that most of the millions of people that will appeal to the OB will never be heard by the Board.[40] This is primarily a result of the sheer number of complaints and how the organization is currently set up.

How many decisions the OB will produce is also affected by how the decision-making process is construed. The first decisions indicate, even if they do not show, that the OB has chosen quality over quantity. Each decision rests upon thorough analysis. The Board will not only take into

---

37 Klein, "Mark Zuckerberg on Facebook's hardest year".

38 "Establishing Structure and Governance for an Independent Oversight Board", Facebook, accessed June 2, 2021, https://about.fb.com/news/2019/09/oversight-board-structure/.

39 "Announcing the Board's next cases and changes to our Bylaws," Oversight Board, accessed June 2, 2021, https://oversightboard.com/news/288225579415246-announcing-the-board-s-next-cases-and-changes-to-our-bylaws/.

40 *Cf.* Evelyn Douek, "Facebook's "Oversight Board", 5 f.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

account the material put forward by the appellant and Facebook but will also, if it thinks it is necessary, conduct its own research. This costs not only money but time, which likely affects the number of decisions it will be able to produce.

It remains to be seen how many cases the OB will take on. Out of the billions of decisions Facebook make every year, only a few – maybe a couple of dozen – will be heard.[41] These cases will likely be high profile disputes, regarding influential people (Donald Trump) or with connections to world politics (genocide or military conflicts). In a special document, Overarching Criteria for Case Selection, the Board has stated the following: "The Oversight Board will select cases for review that raise important issues pertaining to respect for freedom of expression and other human rights and/or the implementation of Facebook's Community Standards and Values. These cases will be of critical importance to public discourse, directly or indirectly affect a substantial number of individuals, and/or raise questions about Facebook's policies. These cases will reflect the user base of Facebook and ensure regional and linguistic diversity."[42] The practicalities of the selection process are regulated in the Bylaws.[43]

In other words, the OB will not provide every user with a fair and equal chance to get the Board to review their case. The decision to focus on issuing guiding decisions and policy recommendations instead of a general possibility to appeal may seem obvious in light of how many Facebook users there are and how many content moderation decisions Facebook and Instagram make every single day. It is still a lost opportunity to provide all users with an internal access-to-justice mechanism. The scale of such a system would, of course, be enormous. But, as a comparison, the European Court of Human Rights in Strasbourg covers 47 nations and a population of more than 800 million people and still manages to work as a "full" court in the real sense of the word.[44]

---

41 See Shira Ovide, "Facebook Invokes its Supreme Court", *The New York Times*, January 22, 2021.

42 "Overarching Criteria for Case Selection", Oversight Board, accessed June 2, 2021, https://oversightboard.com/sr/overarching-criteria-for-case-selection.

43 "Bylaws", Art. 1, sect. 3.

44 "The European Convention of Human Rights – how does it work?", Council of Europe, accessed June 2, 2021, https://www.coe.int/en/web/impact-convention-human-rights/how-it-works.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 6. The decisions

It can be concluded already now that the OB will produce first-class decisions. The process seems rigorous and the Board has based its assessments on thorough research. But one thing seems to be missing: transparent minority opinions.

The Bylaws do allow for dissenting opinions. In 3.1.7, "Draft Decision and Recommendation", the following is stated: "After concluding deliberations, a board panel will draft a written decision, which will include: a determination on the content; the rationale for reaching that decision; and, if desired, a policy advisory statement. The decision will also include any concurring or dissenting viewpoints, if the panel cannot reach consensus."

The last sentence indicates that the Board strives towards unanimous decisions. This is underlined in a "procedural note" that accompanies many of the OB's decisions:

> "The Oversight Board's decisions are prepared by panels of five Members and must be agreed by a majority of the Board. Board decisions do not necessarily represent the personal views of all Members."

In the Oversight Board decision on whether Facebook was right to restrict then president Donald Trump from posting on the platform – the Board found that Facebook's decision was not in itself wrong but that the sanction, indefinite suspension, was not supported by the company's rules – it was mentioned that a minority had a different opinion on some issues, albeit not on the main issue of whether it was within Facebook's right to suspend the president.[45] However, the minority view is not clearly elaborated and is just briefly noted in the majority decision.

Whether dissenting opinions are a good thing or not has long been widely discussed in legal circles, but it is a fact that a dissent can provide important contributions to a discussion of how to weigh different interests against each other. Particularly good examples of this can be found in the area of freedom of speech. Oliver Wendell Holmes' dissent in *Abrams v. the United States* sparked a debate that changed and broadened freedom of speech discourse in the USA.[46] In the further development of the OB

45 "Decision 2021-001-FB-FBR", Oversight Board, accessed June 2, 2021, https://oversightboard.com/decision/FB-691QAMHJ/.

46 See e.g., Thomas Healy, *The Great Dissent, How Oliver Wendell Holmes changes his mind - and changed the history of free speech in America* (New York: Metropolitan Books, 2014), discussing Holmes dissent in *Abrams* v. *United States* from 1919.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

– and I say this in spite of my background as a lawyer in perhaps the most consensus-driven country in the world – surely it would be fruitful to emphasize the differences rather than the compromise.[47]

Chapter 7. The power shift

It is worth mentioning, since it is sometimes forgotten, that when Facebook decides to remove a user's content because of alleged violations against the rules there is always a possibility for the user to go to court if she believes that the decision is in violation of the contract. There has always been a way to "appeal" Facebook decisions – the national courts.

In practice, however, it is often difficult and risky to bring a company such as Facebook to court. Moreover, it is not always clear what it would mean to win a case regarding wrongful moderation of content.[48] Even if one believes the company has made the wrong decision it will not be worth the trouble or cost to take Facebook to court. Not even Donald Trump has thought it worth the effort.

Many countries have independent and private appeals functions that deal with complaints against media companies. Facebook is not only a tech company, but has also taken over some functions traditionally associated with media companies (for instance through Facebook News).[49] The OB has been established to fill a function similar to that of private institutions that have been developed in many countries to address complaints against traditional media.

Early sceptics of the OB project saw Facebook's actions as a strategy to deflect criticism against the company for its decisions on content moderation issues.[50] The suspicion was that Facebook would keep doing what it was doing – getting rid of users and content that the people in Facebook's headquarter in Menlo Park don't like – while using the OB for whitewash-

---

47 See for a general discussion of the merits of public reasoning and the OB, Douek, "Facebook's "Oversight Board"," 66-76.

48 *Cf*. Matthias C. Kettemann et al., "Back up: can users sue platforms to reinstate deleted content?", *Internet Policy Review* 2 (2020): 9.

49 Facebook News is still only available in the USA, "Get Started with Facebook News", Facebook, accessed June 2, 2021, https://www.facebook.com/news/getstarted/.

50 See for a discussion on the OB as a way to outsource controversy Douek, "Facebook's Oversight Board", 23-26 f. Kate Klonik says that this is perhaps "the most common criticism against the Board", Klonik, "Facebook Oversight Board", 2488.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ing purposes. The company would keep the power and the OB would take the responsibility. This line of criticism can still be heard.[51] There is nothing in the first round of decisions that indicates that the OB sees itself as having the role of helping Facebook with public relations.[52] However, as the project has developed, a very different risk has emerged. The members of the OB are becoming the most powerful people in deciding the limits of speech in human history. This concentration of power is in itself worrying.

A reminder of how the process behind content moderation at Facebook used to work.[53] A person that wanted to use the company's product signed a contract and agreed to various terms such as the Community Standards.[54] The Community Standards were continuously changed. Before changing the rules, Facebook would seek input from people and organizations around the world.[55] At the end of the day, it was Facebook that decided what kind of rules it wanted and users' decision whether to stay on the platform or to leave.

The introduction of the OB has changed the power structure. Now the power is concentrated in a small group of experts.[56] A few dozen people get the last word on how to interpret the rules that govern the possibility to use the largest platform for communication and interaction that ever existed. They have also been given the power to affect the rules

---

51 See, e.g. the statements by Marietje Schaake, international policy director at Stanford University's Cyber Policy Center and a member of an alternative organization, called the "the Real Facebook Oversight Board", in Billy Perrigo, "Facebook's New Oversight Board Is Deciding Donald Trump's Fate. Will It Also Define the Future of the Company?", *Time*, January 29, 2021, https://time.com/5934393/facebook-oversight-board-big-tech-future/.

52 Rather, there are signs that it sees itself as a watchdog: Oversight Board (@OversightBoard), "Where Facebook limits users' expression without good reason, we will call them out. Over time, we hope this will ground Facebook's decisions in human rights and benefit users everywhere.", Twitter, May 26, 2021, 2:08 p.m., https://twitter.com/OversightBoard/status/1397524951909941252

53 See for a background Klonik, "Facebook Oversight Board", 2427-2448.

54 These standards were previously not communicated to public/users. See Nicholas P. Suzor, *Lawless. The Secret Rules that govern our Digital Lives* (Cambridge: Cambridge University Press, 2019).

55 In fact, Facebook still listens to stakeholders in the development of community standards. *See* "Stakeholder Engagement", Facebook, accessed June 2, 2021, https://www.facebook.com/communitystandards/stakeholder_engagement.

56 There are other ways to interpret this development. One interpretation is that this is a shift from "Mark [Zuckerberg] decides" to "a transparent process", see Chinmayi Arun, "Facebook's Faces", *Harvard Law Review Forum* 135 (2021).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

that govern them and decide how their own work should be organized.[57] The members of the OB are not only "judges": they are also partly in charge of their own legislation. This is a unique concentration of power over access to freedom of expression to billions of people. At no time in human history have so few people exercised this much control over so many other people's possibility to be heard.

Chapter 8. Concluding Remarks

Facebook's Oversight Board is the most ambitious attempt at construing a private access-to-justice function for content moderation issues in social media. The project in itself is laudable, but there are also problems or potential problems that need further discussion. This article raises six such problems of different kinds.

The most important objection could be boiled down to: "not enough contract law, too much human rights law". To iterate: Facebook's relationship with its users is based on contract. A user that signs the contract has accepted its rules. If the user breaks the rules, the company has a right to use the remedies that follow from the contract, if no clear rules speak to the contrary. This banal observation is sometimes lost in a discussion where Facebook is compared to states, the OB is compared to a Supreme Court and the interest of users in accessing Facebook is labelled as a freedom of speech-issue. "My house, my rules" is still a good starting point.

Bibliography

Andersson, Håkan. *Ansvarsproblem i skadeståndsrätten*. Uppsala: Iustus, 2013.

Arun, Chinmayi. "Facebook's Faces." *Harvard Law Review Forum Volume* 135 (2021): *to be published*.

Aswad, Evelyn Mary. To Protect Freedom of Expression: Why Not Steal Victory from the Jaws of Defeat." *Washington & Lee Law Review* 77 (2020): 609-659.

Berlin, Isaiah. *Four Essays on Liberty*. Oxford: Oxford University Press, 1969.

Carlsson, Anni. "Tyst vår?." *Svensk Juristtidning* (2021): 169-178.

Chander, Anumap. "Facebookistan." *North Carolina Law Review* 90 (2012): 1807-1842.

57 The Board can only decide its own rules if 2/3 of the Board agrees upon it and if the amendment does not conflict with the Charter, see "Bylaws", art. 5, sect. 1.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Clapham, Andrew. *Human Rights Obligations of Non-State Actors*. Oxford: Oxford University Press, 2006.

Council of Europe. "The European Convention of Human Rights – how does it work?" Accessed June 2, 2021, https://www.coe.int/en/web/impact-convention-human-rights/how-it-works.

Douglas-Scott, Sionaidh. "The Hatefulness of Protected Speech: A Comparison of the American and European Approaches." *William & Mary Bill of Rights Journal* 7 (1999): 305-346.

Douek, Evelyn. "Facebook's "Oversight Board": Move Fast with Stable Infrastructure and Humility." *North Carolina Journal of Law and Technology* 21 (2019): 1-78.

European Audiovisual Observatory. *To Have or Not to Have Must-Carry Rules*. Strasboug: European Audiovisual Observatory, 2005. https://rm.coe.int/168078349b.

Farer, Tom. "The Hierarchy of Human Rights." *American University International Law Review* 8 (1992): 115.

Healy, Thomas. *The Great Dissent, How Oliver Wendell Holmes changes his mind - and changed the history of free speech in America*. New York: Metropolitan Books, 2014.

Kettemann, Matthias C. Tiedeke, Anna Sophia. "Back up: can users sue platforms to reinstate deleted content?" *Internet Policy Review* 2 (2020): 9.

Kirkpatrick, David. *The Facebook Effect*. New York: Simon & Schuster, 2010.

Klein, Ezra. "Mark Zuckerberg on Facebook's hardest year, and what comes next." *Vox*, April 4, 2018, https://www.vox.com/2018/4/2/17185052/mark-zuckerberg-facebook-interview-fake-news-bots-cambridge.

Kleineman, Jan. "Konstitutionell skadeståndsrätt." *Juridisk Tidskrift* (2018-19): 23-40.

Klonik, Kate. "Inside the Making of Facebook's Supreme Court." *The New Yorker*, Feb. 12, 2021.

Klonik, Kate. "The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression." *The Yale Law Journal* 129 (2020): 2418.

Knox, John H. "Horizontal Human Rights Law." *American Journal of International Law* 102 (2008): 1-47.

Lévesque, Maroussia. "Applying the UN Guiding Principles on Business and Human Rights to Online Content Moderation." Accessed June 2, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3789311.

Newton, Casey. "Why Facebook needs a Supreme Court for content moderation." *The Verge*, August 21, 2018, https://www.theverge.com/2018/8/21/17762354/facebook-supreme-court-content-moderation.

Ovide, Shira. "Facebook Invokes its Supreme Court." *The New York Times*, January 22, 2021.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Perrigo, Billy. "Facebook's New Oversight Board Is Deciding Donald Trump's Fate. Will It Also Define the Future of the Company?." *Time*, January 29, 2021, https://time.com/5934393/facebook-oversight-board-big-tech-future/.

Schultz, Mårten. *Det här får man inte säga i det här landet!*. Stockholm: Stiftelsen Juridisk Fakultetslitteratur, 2021.

Schultz, Mårten. "Nya argumentationslinjer i förmögenhetsrätten: Rättighetsargument." *Svensk Juristtidning* (2011): 989-1018.

Schultz, Mårten. "Rights through Torts." European Review of Private Law 17 (2009): 305-333.

Shue, Henry. *Basic Rights*. New Jersey: Princeton University Press, 1996, Second Ediiton.

Stenlund, Karolina. *Rättighetsargument i skadeståndsrätten*. Uppsala: Iustus, 2021.

Suzor, Nicholas P. *Lawless. The Secret Rules that govern our Digital Lives*. Cambridge: Cambridge University Press, 2019.

Facebook and Oversight Board Sources

Facebook. "Establishing Structure and Governance for an Independent Oversight Board." Accessed June 2, 2021, https://about.fb.com/news/2019/09/oversight-board-structure/.

Facebook News. "Get Started with Facebook News." Accessed June 2, 2021, https://www.facebook.com/news/getstarted/.

Facebook. "Stakeholder Engagement." Accessed June 2, 2021, https://www.facebook.com/communitystandards/stakeholder_engagement.

Oversight Board. "Announcing the Board's next cases and changes to our Bylaws." Accessed June 2, 2021, https://oversightboard.com/news/288225579415246-announcing-the-board-s-next-cases-and-changes-to-our-bylaws/.

Oversight Board. "Bylaws." Accessed June 2, 2021, https://oversightboard.com/sr/governance/bylaws.

Oversight Board. "Decisions." Accessed June 2, 2021, https://oversightboard.com/decision/.

Oversight Board. "Decision 2021-001-FB-FBR." Accessed June 2, 2021, https://oversightboard.com/decision/FB-691QAMHJ/.

Oversight Board. "Overarching Criteria for Case Selection." Accessed June 2, 2021, https://oversightboard.com/sr/overarching-criteria-for-case-selection.

Oversight Board. "The Oversight Board is accepting user appeals to remove content from Facebook and Instagram." Accessed June 2, 2021, https://oversightboard.com/news/267806285017646-the-oversight-board-is-accepting-user-appeals-to-remove-content-from-facebook-and-instagram/.

Oversight Board. "Rulebook for Case Review and Policy Guidance." Accessed June 2, 2021, https://oversightboard.com/sr/rulebook-for-case-review-and-policy-guidance.

Oversight Board. "Trustees." Accessed June 2, 2021, https://oversightboard.com/governance/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Screenshots: A Glance beyond the Transatlantic

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

„Open with Caution".
How Taiwan Approaches Platform Governance in the Global Market and Geopolitics

Kuo-Wei Wu, Shun-Ling Chen, Poren Chiang[1]

Abstract: Originated in the US, platform governance has relied on self-governance. To make GAFAM and other tech companies accountable to values in democratic societies, the EU proposes a more interventionist model. The rise of Chinese platforms has led to new concerns about state censorship and surveillance. Yet, often considered as exotic exceptions, neither paradigm effectively addresses the accountability problems of Chinese platforms. As a major ICT manufacturer and with a peculiar position in global geopolitics, Taiwan finds both models inadequate. This paper explains Taiwan's specific concerns and offers examples of how it seeks to strike a balance between effective platform governance, free speech, industrial growth and national security.

Keywords: platform governance, GAFAM, big tech, Chinese platforms, disinformation campaigns, free speech, infiltration, geopolitics, Taiwan-China relationship, national security

Introduction

In the past years, the European Union has led important discussions on platform governance and introduced new regulations. While the United States has largely relied on platform self-governance and given providers

---

1 Kuo Wei Wu, M.S. in Computer Science, Columbia University; Chair of TWIGF; Former Board Member of ICANN; Former Board Member of Chunghwa Telecom. Email: kuoweiwu@gmail.com.
Dr. Shun-Ling Chen, S.J.D, Harvard Law School; Associate Research Professor and Co-Director of the Information Law Center, Institutum Iurisprudentiae, Academia Sinica (Taipei, Taiwan). Email: shunlingchen@sinica.edu.tw.
Poren Chiang, LL.M., UCLA School of Law; Research Assistant at Institutum Iurisprudentiae, Academia Sinica (Taipei, Taiwan). Email: hi@poren.tw.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

much leeway to shape their own terms of content removal and privacy policies, the EU approach tends to intervene more. For example, Germany's NetzDG mandates platforms to set up effective systems to manage complaints regarding hate speech and unlawful content; France introduced new legislation against disseminating disinformation during elections; the EU overtakes the US in terms of setting a higher standard for user privacy (i.e., the GDPR) and seeks to export it as a new paradigm. With Google, Apple, Facebook, Amazon and Microsoft (GAFAM)—all American companies—leading the global market, EU regulators have much concern about the European competitiveness.

Taiwan has mostly followed the US model and has a rather hands-off approach in platform governance. It does share some of the above concerns and has begun to look to the EU as an alternative regulatory model. However, with its own specific socio-political context, its industrial structure and population size, Taiwan may hesitate to accept the European model and will find its own way to position itself in the global market and geopolitics. As the US–China decoupling continues to unfold, TSMC alone has allowed the world to take note of Taiwan's strategic importance in ICT manufacturing. Taiwan's geographical location also gives it a critical role in the submarine cable network, which has been further boosted as Hong Kong's political instability grows. With a close yet thorny relationship with China, Taiwan has been a target for disinformation campaigns. Taiwanese government and civil society share the same goal of fighting disinformation with a robust and free internet. Like other countries, internet platform governance issues intersect various fields: national security, democracy, business opportunities, etc. With Taiwan's unique role in geopolitics, how it approaches platform governance may be of interest to regulators and scholars in other countries.

Chapter 1. Taiwan, geopolitics, internet, and platforms

The current thorny relationship between Taiwan and China began in 1949 when Kuomintang (KMT, the Chinese Nationalist Party) retreated to Taiwan. Towards the end of the cold war, Taiwan lifted the martial law that went into effect in 1947 and gradually opened up cross-strait traffic. Taiwanese investment in China gradually increased and broadened to

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

include ICT related industries.[2] Taiwan's capital investment in China gradually increased from 1991 (US$174 million) and peaked between 2010 and 2012 (US$12.8–14.6 billion) under the KMT government. After President Tsai Ing-wen (Democratic Progressive Party, DPP) was elected in 2016, the cross-strait tension heightened. Tsai's administration took proactive measures to divest from China, and the number gradually dropped to 4.1 billion US dollars in 2019.[3]

With the long martial law history under the KMT government and the constant threat from China, the Taiwan–China relationship and the Taiwan identity have been the most paramount issues in the democratization process. The amount of traffic between Taiwan and China has significant impacts on Taiwanese domestic politics. According to China's 2010 official census, more than 1.5 million to 2 million Taiwanese were working, studying or living in China.[4] During the 2012 presidential election, more than 200,000 Taiwanese expats in China flew back to vote.[5] (Taiwan does not have absentee ballots.) The number of Taiwanese citizens in China has also declined in the past years. Yet, in 2019, more than half of the Taiwanese working overseas were in China (including Hong Kong and Macau).[6] Chinese visitors (including Hong Kong) to Taiwan grew from 2.5 million in 2011 to the high point at 5.5 million in 2015, and down to 4.3 million (including 1.6 million from Hong Kong) in 2019.[7] Until 2016, almost half a million people have immigrated to Taiwan from China (including Hong

2 Lin Chu-chia 林祖嘉, "台商在兩岸經貿發展的過去與未來" [The past and future of Taiwanese merchant in cross-strait trade development], *National Policy Foundation*, March 25, 2011, https://www.npf.org.tw/2/8948.

3 Taiwan Ministry of Economic Affairs, Investment Commission, "Investment to Mainland China," Statistics Chart, https://www.moeaic.gov.tw/business_category.view?seq=3&lang=en (accessed May 7, 2021).

4 *Apple Daily (HK)*, "近 200 萬台灣人居大陸" [Near 2 million Taiwanese lives in mainland], November 25, 2014,. https://collection.news/appledaily/articles/4UIFDGBV6NLHERYKX5JDFDFLVQ (archived).

5 Peter Shadbolt, "Taiwan's expats seen as key in presidential poll", *CNN*, January 14, 2012, https://edition.cnn.com/2012/01/13/world/asia/taiwan-election/.

6 Taiwan Directorate-General of Budget, Accounting, and Statistics", 108 年國人赴海外工作人數統計結果" [2019 statistical result of nationals working overseas], news release, December 17, 2020, https://www.dgbas.gov.tw/public/Attachment/01217147167RLW6M7Z.pdf. There were more than 739,000 Taiwanese working overseas in 2019. Specifically, 395,000 nationals were working in China (including Hong Kong and Macau), which was about 53.4%.

7 Tourism Statistics Database of the Taiwan Tourism Bureau, "Changes in the number of visitor arrivals from Japan, South Korea, Malaysia, Mainland China and Hong Kong from 2011~2020," https://stat.taiwan.net.tw/ (accessed May 7, 2021).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Kong and Macau).[8] The Taiwanese identity has grown over time. People identifying themselves as Taiwanese grew from 17.6% In 1992 to 67% in 2020, and people identifying themselves as Chinese dropped from 25.5% to merely 2.4%.[9] This political and demographic context is critical for understanding Taiwan's fight against disinformation campaigns, especially in recent elections.

Despite the instability of being on the seismic belt, Taiwan's location makes it an important node in the submarine cable network. Most of the undersea cables connecting the US to Asia make landfall in Japan, then past Taiwan, across the South China Sea to ASEAN countries. Taiwan's south and east coasts are crowded with submarine cables. An earthquake in southern Taiwan in 2006 caused interruption for several cables in the area, which severely disrupted telecommunication in Southeast Asia. Internet access slowed down as much as 98% for Taiwan, Malaysia, Singapore, Thailand and Hong Kong.[10] With the recent US-China decoupling and the deteriorating political situation in Hong Kong, submarine cable has become a heated issue. In 2020, Washington partially objected to the building of an undersea internet cable that connects the United States and Asia through Hong Kong and instead recommended that it goes through Taiwan and the Philippines to prevent any direct control by China.[11]

Similar to many European countries, Chunghwa Telecom used to operate as the only national telecommunication carrier until the revision of

---

8 Taiwan National Immigration Agency", 大陸地區人民、港澳居民、無戶籍國民來臺居留、定居人數統計表 11001" [January 2021 statistics chart for Mainland China, Hong Kong, Macau resident, and stateless person setting up residence or registering permanent residence], https://www.immigration.gov.tw/5382/5385/7344/7350/8883/?alias=settledown&edate=202101.

9 Lin Kelun 林克倫, "政大民調：台灣人認同感 67% 創歷年新高" [NCCU poll: 67% identify as Taiwanese, a historic high], *CNA*, July 3, 2020, https://www.cna.com.tw/news/firstnews/202007030346.aspx.

10 Winston Qiu, "Submarine cables cut after Taiwan earthquake in Dec 2006", *Submarine Cable Network*, March 19, 2011, https://www.submarinenetworks.com/news/cables-cut-after-taiwan-earthquake-2006.

11 U.S. Department of Justice, "Team Telecom recommends that the FCC deny Pacific Light Cable Network System's Hong Kong undersea cable connection to the United States", news release, June 17, 2020, https://www.justice.gov/opa/pr/team-telecom-recommends-fcc-deny-pacific-light-cable-network-system-s-hong-kong-undersea; Jennifer Elias, "Google gets federal OK to operate subsea cable from Taiwan to US as it nears maximum capacity in Asia", *CNBC*, April 8, 2020, https://www.cnbc.com/2020/04/08/google-gets-federal-ok-to-operate-subsea-cable-from-taiwan-to-us.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the 1996 (Taiwan) Telecommunication Act.[12] After the liberalization in telecommunications, Chunghwa has transitioned into a private company, although the government remains its largest shareholder[13] and HiNet (a Chunghwa subsidiary) is still Taiwan's biggest internet service provider.[14] The 1996 Act requires the chairperson of a Tier 1 company to be a Taiwan citizen, and foreign institutions or individuals are barred from owning more than 49% of the company's share.[15] The Act also requires communication equipment for Tier 1 and Tier 2 to be government-certified.[16] (The above requirements remain unchanged in the Telecommunication Management Act, which has replaced the Telecommunication Act since July 2010.)[17] In addition, while not explicitly stated in the law, the Taiwan government generally does not allow the deployment of China-manufactured network equipment at the infrastructure level.[18] When 4G was first introduced, operators' attempt to adopt Huawei products was rejected by the Taiwan National Communications Commission (NCC).[19] The three major 5G operators, Chunghwa Telecom, Taiwan Mobile, and Far EasTone, use either Nokia or Ericsson, which are both European companies.[20]

---

12 Chen Wen-sung 陳文生 and Wang San-chi 王三吉, "台灣網際網路發展歷程研究之初探" [A preliminary study on Taiwan's internet development history], in *Proceedings of Taiwan Academic Network Conference (TANet) 2005*, http://nccur.lib.nccu.edu.tw/handle/140.119/113242.

13 See generally Taiwan Directorate-General of Telecommunications, "我國電信自由化效益分析研究報告" [Analysis report on the benefits of domestic telecom liberalization], 2003, https://www.ncc.gov.tw/chinese/news_detail.aspx?site_content_sn=475&sn_f=955.

14 "About HiNet", HiNet, last modified March 18, 2021, https://www.hinet.net/globe/en/about.html.

15 Telecommunications Act art. 12 (1996) (Taiwan).

16 *See id.* art. 13, 18, 39, 40, 46, and 52.

17 See Telecommunications Management Act (Taiwan), https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=K0060111. Citizenship requirement of the chairperson for a public telecom operator is in art. 36. Specifications for core communication equipment are in art. 37, 38, 40, 53, and 81 respectively.

18 See, e.g., *Asia Times*, "Taiwan may ban all Chinese equipment, apps", March 12, 2019, https://asiatimes.com/2019/03/taiwan-may-ban-all-chinese-equipment-apps/.

19 Keoni Everington, "After report on Huawei's 'Trojan Horse,' Taiwan retains ban on China-made gear", *Taiwan News*, December 10, 2018, https://www.taiwannews.com.tw/en/news/3593407.

20 Nokia, "Nokia wins exclusive Taiwan Mobile 5G deal", news release, June 29, 2020, https://www.nokia.com/about-us/news/releases/2020/06/29/nokia-wins-exclusive-taiwan-mobile-5g-deal/; Ericsson, "Far EasTone Taiwan expands exclusive Ericsson 5G partnership", March 2, 2021, https://www.ericsson.com/en/press-releases/2021/3/far-eastone-taiwan-expands-exclusive-ericsson-5g-partnership.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Unlike in the field of telecommunications, Taiwanese regulations about platforms are rather scant. Three of the top five websites in Taiwan are owned by foreign companies (Google, YouTube, and Yahoo HK).[21] The top two food delivery platforms are Foodpanda (79.6%) and UberEats (60.8%),[22] both foreign companies. Shopee and Ruten (Chinese and Japanese companies respectively) are among the top online shopping websites. Shopee also takes the lead in mobile shopping.[23] With a reach rate of 98.5%, Facebook is the top social media platform, followed by Instagram (38.8%). LINE is the most popular mobile communication app (99.2%), followed by Messenger (26.8%) and WeChat (21.4%). As previously mentioned, a considerable portion of Taiwanese population has maintained close ties with China. They rely heavily on Chinese apps, such as WeChat,[24] Weibo (microblogging), Baidu Map, Taobao (online shopping), or Didi (the Chinese version of Uber). Although the Taiwan government has been able to ban Chinese telecom equipment in the public infrastructure and prohibit using Chinese apps in public offices,[25] such hardline approaches are rarely taken for Chinese apps and platforms.[26]

---

21 Alexa, "Top Sites in Taiwan", https://www.alexa.com/topsites/countries/TW (accessed May 7, 2021).

22 Xiao Junhui 蕭君暉, "foodpanda 市占率達八成 穩坐美食外送龍頭" [foodpanda holds its throne on food delivery industry with 80% market share], *Economic Daily News*, August 5, 2020, https://money.udn.com/money/story/5612/4756742.

23 U.S. International Trade Administration, "Taiwan – ECommerce," last modified November 8, 2019, https://www.export.gov/apex/article2?id=Taiwan-ecommerce.

24 Yujie Chen, Zhifei Mao, and Jack Linchuan Qiu, *Super-Sticky WeChat and Chinese Society* (Emerald Publishing, 2018). WeChat began as a messaging app like WhatsApp, but has gradually become a mega gateway platform that connects many other third party providers and serves different parts of users' daily activities.

25 Ku Chan 顧荃, "公務資通訊禁中國產品 政院：國安無灰色地帶" [Chinese products banned from official ICT duty; Executive Yuan: no gray area on national security], *CNA*, January 24, 2019, https://www.cna.com.tw/news/aipl/201901240160.aspx.

26 *Cf.* BBC News, "Zoom banned by Taiwan's government over China security fears", April 7, 2020, https://www.bbc.com/news/technology-52200507. One special case may be Zoom. Although Zoom is an American company, its Chinese connection raised national security and censorship concerns. When the Covid-19 pandemic began to unfold in Spring 2020, Zoom quickly became a popular online meeting platform. Since April 2020, the Taiwan government has prohibited public offices, universities and schools from using Zoom. This might have been a less controversial case, as Zoom was not yet widely adopted among Taiwanese users.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 2. Why do Taiwan's approaches (must) differ from the EU?

Section 1. GAFAM is only part of the problem

For regulators in EU, a key policy goal is to release EU countries from the dominance of GAFAM. There is no doubt that these global tech giants are also major players in Taiwan.[27] However, as a significant number of the population constantly travels across the Taiwan Strait, there is heavy reliance on major Chinese platforms as well. Of the top messaging apps, the reach rate of Facebook Messenger (26.8%) is only slightly higher than WeChat (21.4%). LINE, the most popular messaging app, is ultimately a Japanese company. While LINE may be more willing to adopt the US model of self-regulation (e.g., issuing transparency reports) and comply with European laws that are more intervening,[28] it is less likely to see WeChat joining the course. In fact, neither of these regulatory frameworks has shown efficacy when it comes to regulating Chinese platforms—even though the GDPR has set a higher standard for user privacy, it does not address the potentially regular access of private platforms' user data by the government, as what can happen in China.[29] This kind of data access by the Chinese government is of particular concern for Taiwan, as it may lead to cyber security and national security issues, as well as arrest and detention of Taiwanese citizens by the Chinese authorities. For example, the 2020 Hong Kong National Security Law criminalizes secession and sedition. As the law applies to people who do not reside in Hong Kong, China could theoretically charge Taiwanese citizens who supported Hong Kong protesters on social media platforms for violating this law. When a government requests user data, platforms often have to comply with local laws. Even if Taiwanese citizens might bet on GAFAM to decline unreasonable data requests by the Chinese government, the same cannot be said for WeChat and Weibo. In Taiwan's threat model, Chinese platforms and services pose much bigger problems than GAFAM. Nevertheless, with

---

27 Taiwan Network Information Center (TWNIC), "2018 年台灣網路報告" [2018 Taiwan Internet Report], December 2018, https://www.twnic.tw/doc/twrp/201812e.pdf.

28 See, e.g., "Transparency Report", LINE Corporation, https://linecorp.com/en/security/transparency/.

29 Wang Zhizheng, "Systematic Government Access to Private-Sector Data in China", in *Bulk Collection: Systematic Government Access to Private-Sector Data*, ed. Fred H. Cate and James X. Dempsey, 241–58 (New York: Oxford University Press, 2017), https://doi.org/10.1093/oso/9780190685515.003.0011.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

part of its population locked into Chinese platforms, banning Chinese companies for failure of compliance is usually not an option for the Taiwan government.

The oligarchy of GAFAM presents only one set of problems in Taiwan's internet governance. The various efforts to address privacy and ethics in ICT development (e.g., whether to restrict the application of facial recognition, how to avoid algorithmic discrimination) have appeared to be addressing the "western" platforms, leaving out the Chinese platform ecosystem.[30] The fact that these approaches are not effective in regulating Chinese platforms may be particularly problematic for Taiwan, but it is certainly not a Taiwan-only issue. As the traffic between the EU and China continues to grow, and as Chinese platforms seek to expand in the global market, EU countries may also face the same regulatory obstacle. The ban of WeChat and TikTok in the US app stores in 2020 had already met with criticisms for causing hardship for American citizens and residents with connections in China.[31] The rationale for the ban may not have received the credit it should have, partly because the ban was issued by the Trump administration. EU countries may have to tackle the privacy and national security concerns accompanying these platforms in the future.

Section 2. GAFAM as potential partners

Taiwan shares the concerns about GAFAM with EU. But on the other hand, Taiwan also sees business opportunities with big tech. Taiwan is well known for its strength in ICT hardware, and big tech companies heavily depend on Taiwanese manufacturers. In nano-electronics, TSMC dominated the market in both quality and quantity.[32] Most advanced

30 José van Dijck, Thomas Poell, and Martijn de Waal, *The Platform Society: Public Values in a Connective World* (New York: Oxford University Press, 2018). The online geopolitics is roughly divided into two platform ecosystems, the Western and the Chinese, each is completed with separate infrastructure and sectoral platforms, and operates with different political and ideological views.

31 Ana Swanson, David McCabe and Jack Nicas, "Trump administration to ban TikTok and WeChat from U.S. app stores", *New York Times*, September 18, 2020, https://www.nytimes.com/2020/09/18/business/trump-tik-tok-wechat-ban.html.

32 Kathrin Hille, "TSMC: how a Taiwanese chipmaker became a linchpin of the global economy", *Financial Times*, March 24, 2021, https://www.ft.com/content/05206915-fd73-4a3a-92a5-6760ce965bd9. TSMC has 90% of the global market share in the 5–10nm category, 70% in 12–32nm, and 45% in 45–90nm. The car industry mostly uses chips in the 28–65nm category.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

chips are heavily used in 5G, smartphones, high-performance computing, cloud computing and machine learning. Apple, AMD, and Qualcomm are among TSMC's top customers.[33] Since March 2021, TSMC has begun to manufacture CPU chips for Intel.[34] Aside from TSMC, MediaTek is also a major player and has become the biggest smartphone chipset vendor in 2020."[35] Five Taiwanese major IT companies manufacture almost 90% of the notebooks in the world.[36] Taiwanese companies supply over 80% servers to Google, Facebook, Amazon, and Microsoft for their public cloud data centers worldwide.[37] Taiwan is also a data hub for GAFAM companies. Two of three Google's data centers in Asia are already located in Taiwan, and the company announced a plan to build a third one.[38] Google acquired HTC mobile design talents for US$1.1 billion and made Taiwan its main hardware R&D hub outside the US.[39]

The EU has sought to contain the big tech with various approaches, e.g., setting and exporting new legal frameworks, developing EU's own platforms, and having EU's own cloud and data centers. For example, the

33 Whitney Huang, "AMD is becoming TSMC's second largest customer", *TechOrange*, March 23, 2021, https://buzzorange.com/techorange/en/2021/03/23/amd-tsmcs-customer.

34 Paul Alcorn, "Intel to outsource some key CPU production for 2023 chips to TSMC", *Tom's Hardware*, March 24, 2021, https://www.tomshardware.com/news/intel-to-outsource-some-key-cpu-production-for-2023-chips.

35 Ankit Malhotra, "MediaTek becomes biggest smartphone chipset vendor for first time in Q3 2020", *Counterpoint*, December 24, 2020, https://www.counterpointresearch.com/mediatek-biggest-smartphone-chipset-vendor-q3-2020/.

36 Wang Yulun 王郁倫, "鴻海、和碩領電子 6 哥 2020 年營收創新高、4 家入兆元俱樂部" [Foxconn, Pegatron lead the electronics Big Six to record high earnings in 2020, 4 made it to trillion], *Business Next*, January 11, 2021, https://www.bnext.com.tw/article/60891/2020-6-ems-companies-revenue-shipment-comparision.

37 Wang Yihong 王宜弘, "伺服器產業牛氣沖天" [The server industry is as bullish as the sky], *United Daily News*, January 27, 2021, https://udn.com/news/story/6851/5205451.

38 Yu Nakamura, "Google embraces Taiwan as Asia hub with third data center", *Nikkei Asia*, September 4, 2020, https://asia.nikkei.com/Business/Technology/Google-embraces-Taiwan-as-Asia-hub-with-third-data-center.

39 Chris Welch, "Google is buying part of HTC's smartphone team for $1.1 billion", *Verge*, September 20, 2017, https://www.theverge.com/2017/9/20/16340108/google-htc-smartphone-team-acquisition-announced; Cheng Ting-fang and Lauly Li, "Google to make Taiwan its main hardware R&D hub outside US," *Nikkei Asia*, January 27, 2021, https://asia.nikkei.com/Business/Technology/Google-to-make-Taiwan-its-main-hardware-R-D-hub-outside-US.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Gaia-X project is to provide a federated data infrastructure for Europe.[40] Taiwan does have a government cloud, but other than that,[41] Taiwan approaches platform governance differently from the EU. While having local platforms as alternatives to the big tech is ideal, as a late starter and with a rather small domestic market on the global scale, such an ideal is not very realistic. With the small domestic market and without an attempt to become a competing alternative, Taiwan does not pose itself as a potential exporter of normative frameworks. Unlike the EU, Taiwan may see GAFAM more as potential business partners than foe, and is less likely to challenge GAFAM like the EU does.

Chapter 3. The uneven regulatory landscape in Taiwan

Taiwan does not yet have a well-charted legal framework for platform governance.

There were isolated attempts to regulate internet companies and transactions in the early days. Recent administrative and legislative efforts seek to update the regulatory framework on a larger scale but the progress remains sectorial. Aside from addressing the issues brought by recent technological developments, the Taiwan–China relation remains one of the most important concerns.

Section 1. Early clashes

Yahoo acquired Taiwan's major internet portal website in 2000,[42] its most popular blog platform in 2006,[43] and one of the top eCommerce com-

40 "GAIA-X: A Federated Data Infrastructure for Europe", accessed May 7, 2021, https://www.data-infrastructure.eu/GAIAX/.

41 E.g., Chunghwa Telecom, "行政院及所屬委員會雲端資料中心傲視亞洲 首座榮獲國際雙認證之政府雲端資料中心" [Executive Yuan and affiliated commissions take pride in their cloud datacenter among Asia: first government cloud datacenter with two international certifications], news release, October 24, 2014, https://www.cht.com.tw/zh-tw/home/cht/messages/2014/msg-141024-152141.

42 Hong Shuzhen 洪淑珍, "雅虎買下台灣人的眼珠—奇摩" [Yahoo bought Kimo, the eyes of Taiwanese people], *Global Views*, December 1, 2000, https://www.gvm.com.tw/article/6566.

43 Dan Nystedt, "Yahoo given go-ahead to buy Taiwanese blog site", *Network World*, March 29, 2007, https://www.networkworld.com/article/2297219/yahoo-given-go-ahead-to-buy-taiwanese-blog-site.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

panies in 2008.[44] As worrying as these mergers might seem, The Taiwan Fair Trade Commission nonetheless approved all of them (the last one was approved on condition),[45] securing Yahoo's dominance in the Taiwanese market in the upcoming years. It achieved near-monopoly in the domestic online auction and web portal market, coining iconic social platform services like Yahoo Answers.[46]

Around the time Yahoo took over Taiwanese' digital life, digital content vendors and app stores ran into obstacles. The Consumer Protection Act poses a mandatory 7-day rescind period for all goods purchased through door-to-door or distance selling.[47] In July 2011, The Taipei City government found Google's Android Market[48] and Apple's iTunes Store[49] noncompliant to this statute, fining the former for only offering a 15-minute refund window.[50] While Apple swiftly revised its terms,[51] Google delisted all paid apps in Taiwan and filed suit to appeal the fine. The court ruled in favor of Google on jurisdictional grounds, although it agreed with the city that the refund window was insufficient.[52] The incident led to criticism from app developers and the IT industry, denouncing governmental bodies for their obliviousness and "risking the opportunity of industrial

---

44 Zhao Yuzhu 趙郁竹, "Yahoo!奇摩將併購興奇科技 加碼電子商務" [Yahoo Kimo is acquiring MONDAY Tech, staking on e-commerce], *iThome*, April 8, 2008, https://www.ithome.com.tw/node/48346.

45 Liao Qianyin 廖千瑩 and Wang Peihua 王珮華, "雅虎奇摩併興奇科技 公平會有條件同意" [Yahoo Kimo's merger with MONDAY Tech conditionally approved by Fair Trade Commission], *Liberty Times*, June 19, 2008, https://ec.ltn.com.tw/article/paper/220767.

46 Wang Xiaowen 王曉玟, "Yahoo!奇摩巨人 主導全民生活" [Internet giant Yahoo Kimo dominates the peoples' lives], CommonWealth, April 21, 2012, https://www.cw.com.tw/article/5032228.

47 Consumer Protection Act art. 19 (Taiwan), https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=J0170001.

48 "Android Market Business and Program Policies", Android Market, Google, 2011, https://web.archive.org/web/20110902130151/http://www.google.com/mobile/android/market-policies.html (archived on September 2, 2011).

49 "Mac App Store, App Store and iBookstore Terms and Conditions", Terms and Conditions, Apple, last updated June 21, 2010, https://web.archive.org/web/20110521103525/http://www.apple.com/legal/itunes/us/terms.html#APPS (archived on May 21, 2011).

50 Jason Tan, "Google, Taipei City still at apps odds", *Taipei Times*, July 16, 2011, http://www.taipeitimes.com/News/front/archives/2011/07/16/2003508342.

51 Id.

52 Under the Act, local and municipal governments do not have jurisdiction on this particular issue.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

upgrading and the nation's economic benefits in general."[53] This strong public outcry perhaps has contributed to a long period of regulatory inertia, in which agencies turned a blind eye to companies that are too big to regulate.

Section 2. Updating the legal framework for ICT innovations

This passive attitude shifted in 2014, when Uber began to operate in Taiwan. Taxi drivers and cab companies, a heavy-regulated industry and a traditionally significant voter base, raged to demonstration and blocked the street in protest.[54] Instead of giving Uber a free pass, the Ministry of Transportation and Communication (MOTC) kept ordering Uber to register as a taxi service,[55] fining the firm and its drivers for illegal operation per ride. The legislature further revised the Highway Act in 2016, increasing the maximum penalty for Uber to NT$25 million (US$780,000) and threatening to revoke the driving license of those who drove Uber without a taxi operator permit. Despite the sanctions, Uber kept rolling the wheels until it accumulated US$10 million in fines.[56] It even orchestrated a huge media campaign to pressure Taiwan to "progress together."[57]

---

53 Zheng Shaofan 鄭少凡, "北市府與 Google 的 Android Market 消保大戰" [The consumer protection war between Google's Android Market and Taipei City government], *WatChinese*, February 5, 2013, https://www.watchinese.com/article/2013/4936.

54 Josh Horwitz, "Uber hits first backlash from taxis in Asia as Taipei cabbies block streets in protest", *Tech in Asia*, July 8, 2014, https://www.techinasia.com/taipei-taxi-industry-drivers-protest-uber-backlash-in-asia.

55 Aries Poon, "Uber fights to stay on the road in Taiwan", *Wall Street Journal*, Dec 22, 2014, https://www.wsj.com/articles/uber-fights-to-stay-on-the-road-in-taiwan-1419243209.

56 Reuters, "Uber will suspend service in Taiwan after being slapped with over $10 million in fines", *Fortune*, February 2, 2017, https://fortune.com/2017/02/02/uber-suspend-service-taiwan-fines/. Uber encouraged users through email and on their app to voice their dismay toward the government.

57 *Up Media*, "好諷刺！不繳稅卻砸重金買廣告 Uber「想和台灣一起進步」" [How ironic! Squandering on ads while not paying taxes, Uber 'seeks to progress with Taiwan together'], November 28, 2016, https://www.upmedia.mg/news_info.php?SerialNo=8134; Sharing Economy Industry Association", 你的力量，帶領台灣前進" [Your power leads Taiwan to move forward], https://web.archive.org/web/20170509032014/http://www.movingtaiwan.com/petition (archived on May 9, 2017). The website was featured several times in Uber's newspaper campaign, urging the public to join the petition against "obsolete transport regulations and

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Standing firm on its assertion to "regulate, insure, and tax" Uber, MOTC nevertheless admitted the potential of a sleek and streamlined taxi experience as represented by Uber. Starting from 2015,[58] the agency had worked rigorously with the public to relax the Regulations Governing Motor Carriers, establishing a new category of "diversified taxi services."[59] Uber was actively involved in the drafting process.[60] Ultimately, Uber complied with the new rule and began working with only this new category of drivers.[61] Uber was a case where new foreign actors accelerated the overhaul of the legal framework for ICT innovation.

Taiwan enjoyed its own Personal Data Protection Act since 1995 (limited to computer-processed information) and 2000 (for personal information in general),[62] but the lack of civic awareness and enforcement had sidelined the law until the EU introduced the GDPR. Since 2018, the National Development Council (NDC) and Taiwan's industry at large have put heavy efforts to achieve compliance.[63] Further protections and a new data protection authority is expected to be introduced in an upcoming amendment bill.[64]

---

conservative government attitude" and "bring Taiwan back among the tiers of Asian Tigers."

58 Audrey Tang, "Uber responds to vTaiwan's coherent blended volition", *Pol.is Blog*, May 23, 2016, https://blog.pol.is/uber-responds-to-vtaiwans-coherent-blended-volition-3e9b75102b9b.

59 *See* Regulations Governing Motor Carriers art. 91 (Taiwan), https://law.moj.gov.tw/LawClass/LawSingle.aspx?pcode=K0040003&flno=91.

60 "UberX private car taxi service", vTaiwan, https://vtaiwan.tw/topic/uberx/; Richard D. Bartlett, "How Taiwan solved the Uber problem", Medium, June 12, 2016, https://richdecibels.medium.com/how-taiwan-solved-the-uber-problem-29fd2358a284.

61 J.R. Wu, "Uber resumes ride-hailing service in Taiwan after talks with authorities", *Reuters*, April 13, 2017, https://www.reuters.com/article/us-uber-tech-taiwan-idUSKBN17F0KB.

62 Personal Data Protection Act (Taiwan), https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=I0050021.

63 Chen Meiyin 陳梅英, "國發會力拼 2 年內取得歐盟 GDPR 適足性認定" [NDC strives to obtain EU GDPR adequacy decision in 2 years], Liberty Times, July 4, 2018, https://ec.ltn.com.tw/article/breakingnews/2478465. Note that the Ministry of Justice handed over its jurisdiction to the NDC in 2019. (*See* "Legislative History", Personal Data Protection Act (2015) (Taiwan), Taiwan Laws & Regulations Database, https://law.moj.gov.tw/ENG/LawClass/LawHistory.aspx?pcode=I0050021.)

64 Taiwan National Development Council (NDC), "國發會推動個資法修法，力拼 GDPR 適足性認定"[NDC pushes for PDPA amendment, striving to obtain EU

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

As for telecommunication regulations, the NCC proposed two notable bills: (1) the Digital Communications Act (will be further discussed in chap. 4), and (2) the Internet Audiovisual Service Management Act, which will put Netflix and over-the-top (OTT) media services under scrutiny.[65] The bill was introduced to address illegally operating Chinese OTT operators, e.g., iQiyi.com and Tencent Video. The former has accumulated about 6 million subscribers in Taiwan.

Section 3. Combating disinformation

Taiwan has been a target for disinformation campaigns from China. As one tool to influence Taiwanese politics, these campaigns tend to escalate during the election seasons and focus on controversial and dividing topics. The impacts felt during the 2018 local elections and referendums sent shock waves to the Tsai administration and Taiwan civil society.[66] The then upcoming 2020 presidential and congressional elections called for immediate and proactive actions. Like in the US and Europe, social media has become a main channel for disinformation in Taiwan. Meanwhile, eager to get out of the swamp of criticism, major social media platforms and messaging service providers (e.g., Google, Facebook, LINE, Yahoo) has been eager to display a commitment to defend democracy since 2016. New initiatives included better reporting and removal mechanisms, more transparency for political ads, closer collaboration with independent organizations and civil tech communities on fact-check, and so on.[67] Towards

---

GDPR adequacy decision], news release, December 29, 2019, https://www.ndc.gov.tw/nc_27_33660.

65 Shelley Shan, "Commission bill aims to halt services to illegal Chinese over-the-top providers", Taipei Times, July 16, 2020, https://www.taipeitimes.com/News/front/archives/2020/07/16/2003740010.

66 Judit Bayer, Bernd Holznagel, Katarzyna Lubianiec et al, "Disinformation and propaganda: impact on the functioning of the rule of law and democratic processes in the EU and its Member States — 2021 Update", requested by European Parliament INGE Committee (Brussels: European Union, 2021), PE 653.633.

67 Wong Qianru 翁芊儒, "網路平臺聯手打擊不實消息，臉書、Google、Line 皆在臺啟動事實查核，更聯手在地平臺共擬自律準則" [Internet platforms coordinate to fight false information: Facebook, Google, LINE all kicked off fact-checking in Taiwan and even partnered with local platforms to draft self-regulation standards], *iThome*, June 21, 2019, https://www.ithome.com.tw/news/131416; Alice Su, "Can fact-checkers save Taiwan from a flood of Chinese fake news?", *Los Angeles Times*, December 16, 2019, https://www.latimes.com/world-nation/story/2

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the end of the 2020 election, Facebook even set up a "war room," which worked around the clock to allow for expeditious responses.[68] These efforts also help to demonstrate the companies' willingness and capability to self-regulate, diverting the government from taking heavy-handed approaches.

Disinformation operations pre-exist online platforms. Want Want—a food conglomerate with vested interests in China—acquired the China Times and CTiTV. Both received instructions from the Chinese government on news stories related to cross-strait relations.[69] Civil society has protested against the "red media" since 2012, when the Want Want Group sought to acquire a cable TV operator. The NCC did not approve the acquisition in the end. Policy debates surrounding this incident have continued and renewed after 2018, resulting in a 2019 NCC-proposed bill on "Media Monopolization Prevention and Diversity Preservation" in 2019.[70] Congress did not pass the bill until its re-election in 2020, though. In 2014, CTiTV's license renewal was issued with conditions as it had repeatedly violated regulations. Although CTiTV had a good track record between 2014 and 2017, its violations began to pile up again after 2018, with multiple incidents involving inadequate fact-checks. The NCC refused to renew CTiTV's license in November 2020.[71] CTiTV has begun to broadcast via YouTube and OTT, moving itself into the less regulated field of platform governance.

019-12-16/taiwan-the-new-frontier-of-disinformation-battles-chinese-fake-news-as-elections-approach.

68 Jeffery Wu and Joseph Yeh, "Facebook to establish 'war room' in Taipei ahead of elections", *Focus Taiwan*, December 30, 2019, https://focustaiwan.tw/sci-tech/201912300015.

69 Kathrine Hille, "Taiwan primaries highlight fears over China's political influence", Financial Times, July 17, 2019, https://www.ft.com/content/036b609a-a768-11e9-984c-fac8325aaa04.

70 Lin Shangzuo 林上祚, "反媒體壟斷法捲土重來！NCC 新版草案審查完成 媒金分離不溯及既往" [Anti-media monopolization act comes back! NCC passed the new draft bill; separation of media and financial institutions does not apply retroactively], Storm Media, January 16, 2019, https://www.storm.mg/article/833489.

71 Su Siyun 蘇思云, "NCC 委員一致決議否決中天新聞台換照：違規嚴重 內控失靈" [NCC members unanimously rejected CTiTV's license renewal: serious violations, failed internal control], CNA, November 18, 2020, https://www.cna.com.tw/news/firstnews/202011185006.aspx; NCC, "國家通訊傳播委員會決議予以駁回「中天新聞台」衛廣事業執照換發申請" [NCC votes to reject CTiTV's broadcast service license renewal application], news release, November 18, 2020, https://www.ncc.gov.tw/chinese/news_detail.aspx?site_content_sn=8&sn_f=45332.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

China sees Taiwan as a renegade province. On the other hand, until Taiwan adopts a new constitution or amends the article that defines territory in the constitution of "the Republic of China" (Taiwan's official name), mainland China is technically still a part of its territory. China, however, would consider either a new constitution or an amendment as inciting for breaking the "status quo." This knotty political reality causes much agony in Taiwan's foreign affairs. In addition to the difficulties in diplomacy and international participation, Taiwan also places "China" and "Chinese" affairs in a distinct category. As mentioned earlier, the Taiwanese government is reluctant to ban Chinese apps, even though there are considerable national security concerns. Nevertheless, Chinese platforms may be barred from providing services in Taiwan. For example, Didi entered the Taiwan market in January 2018 but discontinued services at the end of the year.[72] Unlike Uber, Didi and other Chinese companies are the subject of the "Cross-Strait Relations Act,"[73] which sets stricter requirements and procedures for Chinese investment in Taiwan. Taobao, an Alibaba subsidiary cloaked as a British company, began operating in Taiwan in October 2019 and was shut down by the end of 2020 for violating the same act.[74]

Since 2017, there was a digital communications bill in Congress aiming to safeguard the communication environment and facilitate digital transformation.[75] The initial bill allowed platforms much room to self-regulate. In its first term (2016–2020), the Tsai administration[76] did deliberate on whether to revisit that bill to give platforms more responsibilities, including making platforms liable for hosting questionable contents or for not responding timely. By the end of 2018, the government had concluded

72 Mia, "退出台灣？罰款 4.3 億後滴滴出行暫止服務" [Leaving Taiwan? Didi halted service after 430 million fine], *Inside*, December 20, 2018, https://www.inside.com.tw/article/15060-didi-stopped-services-in-taiwan-for-now.

73 The full name of the statute is the Act Governing Relations between the People of the Taiwan Area and the Mainland Area.

74 Liu Jiqin 劉季清, "震撼！淘寶台灣今關閉平台 年底退出台灣" [Astonishing! Taobao Taiwan closes down its platform today, leaving Taiwan at the end of the year], *Business Today*, October 15, 2020, https://www.businesstoday.com.tw/article/category/80392/post/202010150019/.

75 Taiwan National Communications Commission (NCC), *2019 NCC Performance Report* (Taipei, 2020), 24, https://www.ncc.gov.tw/english/files/20091/382_5243_200918_1.pdf.

76 Tsai was reelected in 2020, serving her second term.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to follow the Manila Principles and support a self-regulatory model.[77] Instead of revising the digital communications bill, the DPP government proposed to review and fortify existing laws with clauses that penalize the intentional dissemination of rumors which may cause harm or public panic. One of the main concerns is that making platforms more responsible may inadvertently lead to private censorship.[78] Nevertheless, Congress did not pass the digital communications bill before the 2020 Congressional re-election, and as a rule the legislative process must start anew. Congress did pass the Anti-infiltration Act in December 2019 to combat Chinese influences on the domestic political processes. The Act targets agents of foreign hostile forces and their activities in lobbying and campaigning.[79] After the 2020 election, the NCC revisited the digital communications bill under a new commissioner. Although not yet revealed, early discussions suggest that the new bill mandates platforms to remove certain content at the request of the government for reasons such as national security, communications security, or criminal offenses.[80] The opposition parties[81] and The Asia Internet Coalition led by global internet and technology firms expressed serious concerns, regarding the bill as a potential threat to free expressions.[82] The Taiwan Association for Human Rights also calls for

77 "SayIt database of Taiwan Public Digital Innovation Space (PDIS)", 2018-12-13 行政院第 3630 次會議後記者會" [2018-12-13 Executive Yuan No. 3630 post-meeting press conference], https://sayit.pdis.nat.gov.tw/2018-12-13-%E8%A1%8C%E6%94%BF%E9%99%A2%E7%AC%AC-3630-%E6%AC%A1%E6%9C%83%E8%AD%B0%E5%BE%8C%E8%A8%98%E8%80%85%E6%9C%83.

78 "SayIt database of Taiwan Public Digital Innovation Space (PDIS)".

79 See Anti-infiltration Act (Taiwan), https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=A0030317.

80 Zhang Yifeh 張逸飛, "NCC 擬重提「數位通訊傳播法草案」 賴香伶：別用傳統思維治理網路世界" [NCC considers to bring 'Digital Communications Act' bill back to table; Lai Hsiang-Ling: stop governing the internet with old-fashioned mind], *Newtalk*, December 29, 2020, https://newtalk.tw/news/view/2020-12-29/515825.

81 Lin Yu-hsuen and Joseph Yeh, "Digital communications draft bill not internet censorship: NCC", *Focus Taiwan*, December 14, 2020, https://focustaiwan.tw/society/202012140014.

82 Jeff Paine, "AIC on digital communications act", editorial, *Taipei Times*, December 18, 2018, https://www.taipeitimes.com/News/editorials/archives/2018/12/18/2003706312.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

more public hearings to revise or stop the Act.[83] We still await the actual bill.

Chapter 5. Conclusion

The EU is taking a proactive role in setting a new regulatory paradigm for platforms to address privacy and ethical issues involved in platforms' business models, as well as the oligarchical structure of the market. While Taiwan shares many of EU's concerns, it may not find EU's platform governance strategies the best fit. Aside from the differences in the industrial make up, the cross-strait relations are often the trumping factor in policy discussions in Taiwan. Seeking to strike a balance between effective regulation, free speech, industrial growth and national security, platform governance is a complicated and contentious issue. Taiwan is less likely to directly challenge GAFAM as the EU does. Taiwan appreciates the EU for setting higher regulatory standards. However, to adequately address Taiwan's national security concerns, frameworks that are not effective in regulating major Chinese platforms can only be a partial solution. Such threats are not specific to Taiwan, but they are easily overlooked in other countries as Chinese platforms or companies are not as dominant as GAFAM. Nevertheless, It would be naive to leave out the problems posed by Chinese platforms in the platform governance debates.[84] Taiwan does not have a well-crafted solution for platform governance either, but the government and the civil society tackle it from other angles to ensure national security and sustain a healthy democracy.

Bibliography

Bayer, Judit, Bernd Holznagel, Katarzyna Lubianiec, Adela Pintea, Josephine B. Schmitt, Judit Szakács, and Erik Uszkiewicz. *Disinformation and propaganda: impact on the functioning of the rule of law and democratic processes in the EU and its Member States — 2021 Update*. Requested by European Parliament INGE Committee. Brussels: European Union, 2021.

---

83 Ho Ming-Syuan 何明諠, "恐危害言論自由的數位通訊傳播法" [Digital Communications Act that could harm the freedom of speech], *Taiwan Association for Human Rights*, May 22, 2017, https://www.tahr.org.tw/news/1999.

84 *Economist*, "The most dangerous place on Earth", May 1, 2021, https://www.economist.com/leaders/2021/05/01/the-most-dangerous-place-on-earth.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Cate, Fred H., and James X. Dempsey, ed. *Bulk Collection: Systematic Government Access to Private-Sector Data*. New York: Oxford University Press, 2017.

Chen, Yujie, Zhifei Mao, and Jack Linchuan Qiu. *Super-Sticky Wechat and Chinese Society*. Emerald Publishing, 2018.

Van Dijck, José, Thomas Poell, and Martijn de Waal. *The Platform Society: Public Values in a Connective World*. New York: Oxford University Press, 2018.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Digital Platform Regulation in Japan – does the soft approach work?

Izumi Aizu

Abstract: With the increasing use of digital platforms, new challenges are growing also in Japan. The areas most visible in this regard are the socio-political, economic, as well as privacy and personal data protection.
In the socio-political area, hate speech targeting Korean residents in Japan is most concerning. The negative emotions root in the historical relationship of Japan and Korea. Counteractions by citizens appealed to the international community such as the United Nations that led to a new law, Hate Speech Elimination Act (HSEA) in 2016.
The Act lacks the enforcement tools, but its soft approach has been supplemented by local ordinances and court decisions that effectively reduced hate speech in physical spaces. While hate speech seemed to have migrated to the Internet, the combination of industry self-regulation, a new local ordinance with criminal penalty and an emblematic court decision may be capable to tackle that as well.
The economic concerns around the rise of global Big Tech and increased use of big data and AI drove the enactment of new laws: the Act on Improving Transparency and Fairness of Digital Platforms (AITFDP) in 2021 and the revision of the Act on the Protection of Personal Information (APPI) to be enacted in 2022. Again, the legal approaches are soft, lacking the enforcement tools. However, enhanced capabilities of the Ministries and a new industry self-regulation system are expected to bring a better balance between consumer protection and the digital innovation. This "co-regulation" approach may suit to the Japanese societal structure. It also reflects the multi-stakeholder approach, largely exercised among the Internet Governance concerns.

Keywords: Digital Platform Regulation, Hate Speech, Privacy protection, Personal Data protection, Freedom of Speech, Human rights, Big Tech, Big Data, Co-regulation, Soft and hard approach

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Introduction: Three areas and two approaches to platform regulation

Under the digital platform regulation concerns in Japan, there are three major policy areas:

1) Social and political issues including hate speech, harmful and illegal content, and fake news;
2) Economic concerns including protection of domestic small and medium businesses (SMEs) against Big Tech;
3) Consumer protection including protection of privacy and personal data.

When it comes to regulatory frameworks, two different approaches are observed:

a) Hard approach – use of existing legal framework or establishing a new legislation with strong enforcement;
b) Soft approach – relying on voluntary activities of citizens, local autonomy, and industry self-regulation.

This paper will examine these three policy areas and discuss the effectiveness of both hard and soft approaches and their combinations.

A recent book "*Hate Speech in Japan: The Possibility of a Non-Regulatory Approach*" provides a comprehensive and in-depth analysis of the regulatory approaches of hate speech targeting Korean residents in Japan.[1] The Author of this paper highly acknowledges the rich knowledge and insights contained in this large volume and would like to examine the value of Japan's non-regulative soft approach that this book puts forward.

Chapter 1. Hate speech regulation in Japan

This Chapter discusses the regulation on offline and online hate speech in Japan. The hate speech against ethnic Korean residents in Japan has been outstandingly persistent due to their complex historical relationship.[2] The critical issue has been how to effectively eliminate the hate speech targeting the Korean residents.

---

1 Shinji Higaki and Yuji Nasu, eds., *Hate Speech in Japan* (Cambridge: Cambridge University Press, 2021).
2 For the historical background, read references in the Annex.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Aggressive acts by xenophobic Japanese people toward Korean residents have been present until today. Korean youth in Japan often experience assaults online or in real space. The most visible cases are the direct threats given to Korean students in the street going to their Korean Schools in Japan who wear traditional Korean folk-style outfits called *Chima jeogori*. They receive such dirty words as "Go home!" or "We will kill you" during commuting on trains. But that is just the tip of the iceberg: Korean residents frequently encounter other hostile acts by some Japanese citizens.

1.1. Hate speech in 2000s preceding the new legislation

The recent strong hate speech activities targeting Korean residents originated in around 2006. They first took the form of public rallies and street demonstration staged by a xenophobic activist group called "*Zaitokukai*". They claim that Korean residents in Japan are granted special privileges, misinterpreting the meaning of special permanent residency and alleging special welfare and preferential tax treatment, and insist that granting such special privileges to Korean residents in Japan amounts to reverse discrimination against Japanese people.[3]

Their hate speech had such an impact that it became a serious social concern. These activities partly reflected the growing tendency by Japanese public to become more patriotic, or xenophobic, in view of the territorial disputes with Korea over small islands in Sea of Japan. The disputes of historical issues around the so-called "comfort women" during the World War II and the forced labour workers of Koreans under the Japanese Imperial system added fuel to the fire. These disputes are still ongoing and brought over to the court in Korea, and to the public eyes from time to time in Japan, often more visible over the Internet and in social media. Many Korean residents feel threatened, some became furious, all of them got some form of psychological scars.

Zaitokukai's aggressive hate speech in public spaces ignited strong counteractions by the citizens' group of both Korean residents as well as Japanese. It was also brought to the court, that *Zaitokukai* organized threatening hate demonstrations three times between December 2008 and March 2010 in front of the Kyoto Korean Elementary School and distribut-

3 Shinji Higaki, *The Hate Speech Elimination Act*, eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021), Chapter, 11, 368, https://doi.org/10.1017/9781108669559.012.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ed the videos of these demonstrations on YouTube and their website. The school filed a lawsuit against *Zaitokukai* and the Kyoto District Court ordered to pay about 12 Million Yen for the damages and provided injunction to prohibit further demonstrations in the school neighbourhood in 2013. [4]

In the end, the Supreme Court dismissed *Zaitokukai*'s further appeal in 2014. This is the first case where the Japanese court recognized the illegality of a hate speech based on race or nationality. The Supreme Court concluded: Not only were the acts hate speech in general, but also constituted racial discrimination resulting in serious material damage, based on the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD). [5]

Zaitokukai's demonstration threatening the Kyoto Elementary Korean School was also brought to the criminal court that had preceded the civil case and members of the group were found to be guilty of the interruption of business and insulting action affirmed by the Supreme Court in 2011. [6]

Zaitokukai did not give up, however. In 2015, they tried to attack the *Sakuramoto* district of Kawasaki City where many Korean residents have been living peacefully with Japanese citizens and forming an extensive network of community activities such as church, nursery school, social welfare facilities and activities to support elderly citizens and persons with disabilities. It is this positive relationship of Korean and Japanese residents that *Zaitokukai* tried to destroy by staging violent street demonstrations. When the demonstration was attempted, hundreds of local citizens of both Korean and Japanese nationals gathered and blocked the demonstration from entering into the heart of the district.[7] *Zaitokukai* tried again later but in no vail.

---

4 Ryangok Ku, *The Current Movement of Hate Speech*, eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021), Chapter 5, 203, https://doi.org/10.1017/9781108669559.006.

5 Katsuo Yakura, *The Legislative Process Leading to the Hate Speech Elimination Act*, eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021), Chapter 10, 349, https://doi.org/10.1017/9781108669559.011.

6 Kazushi Ogura, *Hate Speech on the Internet*, eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021), Chapter, 18, 617 and 631, https://doi.org/10.1017/9781108669559.019.

7 Kanagawa Shimbun, "ヘイトデモ、我が街に通さず　川崎・桜本 " Kanagawa Newspaper, last modified 2015, https://www.kanaloco.jp/news/social/entry-67417.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

1.2. International voices pushed Japan to the New HSEA

Given these aggressive anti-Korean campaigns and hate speeches by *Zaitokukai* and some growing support for them, citizens groups proactively started to lobby both domestic law and policy makers as well as international organisations such as the United Nations Human Rights Council and active NGOs engaged in these policy areas. After they gave them plenty of chances to make their point, the UN Human Rights Committee (HRC) and UN Committee on the Elimination of Racial Discrimination (CERD) concluded with strong recommendations that the Japanese government must take steps to curb hate speeches.[8]

Domestic voices alone were not enough, but these international voices functioned as an extra pressure to the lawmakers, most of them conservative politicians who had been reluctant to act. Thus, the Act on the Promotion of Efforts to Eliminate Unfair Discriminatory Speech and Behaviour against Persons Originating from Outside Japan (known as the Hate Speech Elimination Act, or HSEA), was finally passed on May 24, 2016, as the first law against hate speech in Japan, much sooner than most had expected.

1.3. The Effect of HSEA challenged

After two failed attempts, *Zaitokukai* announced the third attack on *Sakuramoto* to be held on June 5, 2016, just two days after the new Act (HSEA) was enforced. It was a clear strategic move to deny the practical effectiveness of HSEA.[9]

The local citizens filed a petition that requested a court injunction to prohibit the demonstration. The local court issued an injunction with direct reference to HSEA as well as that of ICERD ratified by Japan and

8 Ayako Hatano, *Hate Speech and International Law*, eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021), Chapter 3, 105, https://doi.org/10.1017/9781108669559.004.

9 Toshihide Yamamura, *A Chronology of Events and Legislation Related to Hate Speech in Japan*, eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021), Appendix A, 723, https://doi.org/10.1017/9781108669559.022.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Article 14 of Japan's Constitution that prohibits discrimination based on race or other attributes.[10]

There has been no attack by the *Zaitokukai*'s on *Sakuramoto* since then. A similar case was found in the *Shin-Okubo* district in Tokyo where *Zaitokukai* staged several demonstrations, but in the end, they were far outnumbered by the citizens and effectively shut out.[11]

Now, as Hatano asks: "Does the HSEA effectively respond to the recommendations from the UN human rights treaty bodies, as is claimed? Specifically, does it in fact 'internalize' international human rights norms at the domestic level?" [12]

HSEA imposes no penalty provisions at all. After defining hate speech as "unfair discriminatory speech and behaviour against persons originating from outside Japan" (Article 2), it prescribes the moral duty of the general public (Article 3) and assigns both central and local governments duties in tracking and eliminating hate speech (Article 4). Articles 5, 6, and 7 provide for measures such as consultation, education, and other awareness campaigns to achieve the goals.[13] Thus, it provides neither any concrete steps, nor sanctions to enforce the law. Therefore, some remain very doubtful and call for additional provision of penalties; whereas others argue it has a unique value worth to maintain. Ogura argues "it will be subject to interpretation in the civil law courts, and that it may exert certain influence on local government permission or rejection of meetings using public facilities, such as demonstrations and rallies."[14] Shinji Higaki, the co-editor of the book "*Hate Speech in Japan*" argues that although the Act lacks the penalty, it "may offer a modest model that strikes an appropriate balance between the freedom of expression and anti-racism." [15] Higaki points out that there are no hate speech laws in the United States either, as the US put absolute value on freedom of expression.[16]

Higaki continues to examine the unique value of the Japanese approach with HSEA.

---

10 Toru Mori, *An Injunction Banning a Xenophobic Group from Demonstrating, Kawasaki Case*, eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021) Chapter 14, 493, https://doi.org/10.1017/9781108669559.015.

11 Personal Interview with Mr. Chun Kang Heon, Secretary, Culture Center Arirang in Shin-Okubo district of Shinjuku on May 11, 2021. Mr. Chun is a second generation Zainichi Korean.

12 Ayako Hatano, "Hate Speech and International Law", 115.

13 Shinji Higaki, *The Hat Speech Elimination Act*, 371.

14 Kazushi Ogura, *Hate Speech on the Internet*, 624.

15 Shinji Higaki, *The Hat Speech Elimination Act*, 366.

16 Shinji Higaki, *The Hat Speech Elimination Act*, 365.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

> "Under current circumstances, hate speech regulation must be implemented deliberately in Japan. The HSEA may be a second-best way of preventing hate speech, at the very least, *but it may be the most suitable model of hate speech law in the world.*" *[emphasis added by the Author]*
> "There are several points that we might highlight as its strengths of the HSEA. First, it respects the 'marketplace of ideas', which is based on the fundamental principles of modern law, such as freedom, autonomy, and self-realization. Second, numerous works on hate speech have argued that the criminal regulation of public discourse will cause undesirable backlash, produce martyrs, or drive dangerous speech underground, but the Japanese non-regulatory model is immune to these problems." [17]

The Author considers that Higaki's high evaluation of HSEA and Japanese model is too optimistic, especially describing it as "the most suitable model in the world", since different societies have different structures and historical and cultural contexts, and therefore no model could work "best" singlehandedly. Yet, as Higaki points out, the respect for "market of ideas" and avoiding undesirable backlash are worth to acknowledge as the merits of HESA.

1.4. Hate Speech on the Internet

After HSEA was established, the *Zaitokukai*'s aggressive hate speech activities seemed to have subsided greatly. HSEA is seemingly working to suppress hate speech in the physical space. However, expressions of hate speech have not entirely vanished. There are still many manifestations of hate speech on the Internet as of today. The centre of gravity has shifted from the offline to the online world. Clearly, the Internet and social network services (SNS) on digital platforms are widely used to spread and amplify hate speech. Kazushi Ogura points out that there is significant co-relation between "offline" and "online" hate speech activities as follows.

> The cases that follow are not cases of discriminatory expressions simply being posted on the Internet but are examples of public demonstrations or rallies that have taken place in offline contexts and have

17 Shinji Higaki, *The Hat Speech Elimination Act*, 379.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

subsequently been filmed, photographed, and uploaded to video-sharing sites.[18]

Hate speech on the Internet has been present in Japan since the early days of Internet use started in the 1990s. As the use of the Internet has grown, the amount of hate speech has increased exponentially. The Internet's potential to hide the identity, the low barrier to send offensive messages to the Internet or SNS, and the easy amplification by copying and spreading these messages are relevant factors.

In addition to targeting Korean residents, there are also aggressive assaults against other ethnic minorities such as Chinese residents as well as expat workers mostly from the developing countries in Asia, Middle East, and South America. But the author believes it is fair to say that the hate speech towards Korean residents has been the most vocal and problematic in Japanese society.

The Human Rights Bureau of the Ministry of Justice published a survey report on the foreign residents in Japan in 2017. The respondents were of many nationalities including Chinese (32.5%), South Korean (22.1%), the Philippines (6.7%), Brazilian (5.2%), Vietnamese (4.8%) and others. North Koreans was only 1.4%. They did not include Japanese nationals who have foreign origins such as Korean Japanese.[19]

In this report, 41.6% of the 3,400 respondents answered that they have seen discriminatory messages on the Internet against foreign residents in Japan. 33.3% of them answered that they have seen hate speech actions such as street demonstrations or rallies against them over the Internet, while 42.9% answered that they have seen them on newspapers or TVs. 65% of them answered that they felt uncomfortable, while 19% said these should not be allowed, 22% felt threatened and only 7% answered they did not feel much.[20]

Clearly, hate speech on the Internet still has negative impacts to foreign residents in Japan. While the general level of emotions against foreigners has not increased that much, in part due to various efforts including the

18 Kazushi Ogura, *Hate Speech on the Internet*, in: *Hate Speech in Japan* (Cambridge, Cambridge University Press, 2021), Chapter 18, 614.

19 Center for the Promotion of Human Rights Education and Encouragement, *外国人住民調査報告書－訂正版*. (FY Center for the Promotion of Human Rights Education and Encouragement, 2017), 8, http://www.moj.go.jp/content/001226182.pdf.

20 Center for the Promotion of Human Rights Education and Encouragement, *外国人住民調査報告書－訂正版*, 45.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

provision of HSEA and other institutional measures, hate speech on the Internet remains as a very serious problem.

The HSEA does not specifically mention the Internet or any electronic means. However, in the supplementary resolutions of the parliament the clause 'implement countermeasures to deal with individuals or groups promoting unjustifiable discriminatory expression against persons from overseas or from outside Japan and to eliminate acts that promote unfair discriminatory behaviour on the Internet' is mentioned as an issue for special consideration. This indicates the problematic nature of the Internet with regard to harmful content.[21]

Nevertheless, today, in 2021, concerns on social and political dimensions of the platform regulation, especially on hate speech and free speech issues are not that high. This does not mean that there are no issues at all, but the general awareness among the Japanese public about the hate speech has become quite low compared with five to ten years ago.

1.5. Industry self-regulation on Internet content

While the explicit scope of HSEA remains outside of the online space, industry self-regulation on the illegal and harmful content in general in voluntary manner has been implemented over the past 20 years.

In 2001, to regulate the illegal and harmful content on the Internet, the Act on the Limitation of Liability for Damages of Specified Telecommunications Service Providers and the Right to Demand Disclosure of Identification Information of the Sender, or the "Provider Liability Limitation Act" in short, was established. This Act helps deleting illegal and harmful material posted over the Internet, yet it only provides the procedural guidelines, and not for not legally binding duties, showing a soft approach again.

When an internet service provider is either requested to delete content that is illegal or harmful by any subject or to disclose the name and contact information of such senders by following the Act, their liabilities will be immured. When providers are asked by "trusted parties" such as a lawyer or the police defined in their voluntary code, they will share the IP address of the sender, but not the identity information. The providers will submit the sender's identity information only when they are asked by a legitimate court order. The reason behind this cautious process is

21 Kazushi Ogura, *Hate Speech,* 613.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

that the Japanese Constitution Article 21 explicitly protects the "secrecy of communication." [22]

There is a "model contract article" jointly published by four Internet related industry associations which sets a standard model of contract article with their customers. Many providers are using this model contract to prohibit defamation, discrimination, and other offensive acts, and to delete certain messages unilaterally without the sender's consent. After the HSEA was enacted in 2016, this model was revised in 2017 adding languages that define hate speech. Irrespective of using this model contract or not, most major commercial providers publish their own contract that explicitly prohibits the posting of material that promotes hate speech and actions. Yahoo! Japan and Twitter Japan are examples of such providers. [23]

1.6. Local ordinances implemented

With institutions such as the Human Rights Bureau of the Ministry of Justice, the administrative branch of the government is engaged in providing remedy for damages caused by human rights violations including hate speech on the Internet.[24] Several local governments are also working on other actions to prevent and delete hate speech, most visibly in the form of issuing a local ordinance against hate speech. The City of Osaka and Kawasaki are leading in this regard as they have a large community of Korean residents.

In 2016, the Osaka City Ordinance to Deal with Hate Speech was established in Osaka. This ordinance deals with instances in which Osaka citizens or organizations suffer damage because of the diffusion of hate speech, including ones via Internet, in or around Osaka City, and citizens – or the mayor of the city of Osaka – may request that steps be taken to curb hate speech.[25]

In July 2020, Kawasaki City established an Ordinance on Establishing a City with No Discrimination and Respecting Human Rights. What is

---

22 Prime Minister of Japan and His Cabinet, *The Constitution of Japan,* 1947, https://japan.kantei.go.jp/constitution_and_government_of_japan/constitution_e.html

23 Human Rights Protection Committee of Daini Tokyo Bar Association, "*Internet and Hate Speech* (in Japanese)" (Gendai Jimbunsha, 2019), 14.

24 Human Rights Protection Committee of Daini Tokyo Bar Association, "*Internet and Hate Speech* (in Japanese)", 15.

25 Kazushi Ogura, *Hate Speech in Japan*, in: "*Hate Speech on the Internet*", Chapter 18, 625.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

unique about this ordinance is that this is the first case where the criminal punishment including financial penalty is included in the official regulation. Hate speech over the Internet had been excluded from criminal punishment since the authority considered the balance between freedom of expression and hate speech. In both cases, the civic groups' active engagement and lobbying played vital roles.[26]

Under this ordinance, a citizen could request the City to become a proxy of him/her so that the City makes the formal request to the Internet Service Provider of taking down the offensive material from the Internet space. Ms. Che Kainjya who is a third-generation Korean living in Kawasaki City filed a lawsuit against the city to request deletion of offensive tweets in 2020. However, it took five months to investigate through a third-party review board who recognized only two tweets out of 332 as offensive.[27] There is no information available as to the basis of this judgement, but the author speculates that the review board weighed the freedom of expression for many of the tweets which had some vagueness in their texts.

Thus, even though citizens' active engagements are pushing the local governments, the case in Kawasaki illustrates the difficulty to materialize an effective solution over hate speech on the Internet in practice.

As mentioned above, hate speech has been included in the industry self-regulation framework. There are several cases reported where Korean residents who used the disclosure procedure of the self-regulation model won compensation payment in the court for having their dignities damaged or defamed.[28]

The latest case was reported on May 13, 2021. The Tokyo High Court ordered a man to pay 1.3 million yen in damages for posting discriminatory comment about Korean residents on his blog. "The posted comments were extremely vicious," presiding Judge Yukio Shirai said, adding that racial discrimination is illegal per se.

The damage's amount is unusually high for comments made via a single post, and it is expected to have a deterrent effect on hate speech, the plaintiff's lawyer said. The plaintiff obtained the identity of the man who

---

26 Naoto Higuchi, *ibid*., in Chapter 16 "Japan's Postcolonial Hate Speech" 546.

27 Joji Mochida, "ヘイトスピーチは止まったか：川崎市が全国初の罰則付き条例" *Nippon.*com, November 12, 2020, https://www.nippon.com/ja/in-depth/d00648/.

28 Human Rights Protection Committee of Daini Tokyo Bar Association, "*Internet and Hate Speech* (in Japanese)" (Gendai Jimbunsha, 2019), 11.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

posted these insulting comments by asking the Internet service provider involved to disclose it.[29]

Since HSEA's text only protects the right of specific individuals, excluding collective term such as race or nationality, some general or abstract expressions such as "Koreans go home" or "kill them" had not been regarded as the subject of this Act.[30] However, this latest court ruling suggests that such comments are largely illegal. The judge took the spirit of the Act, not the letter, and recognized that they hurt the plaintiff's personal rights and constitute racial discrimination.

This latest ruling is expected to bring further potential to reduce hate speech on the Internet. But it may still require active engagement of citizens who dare to file suits in the court.

1.7. Political and Social areas

In addition to hate speech, offensive speech, fake news, mis-information campaigns, cyber-bullying, and communication fraud, all are persistent problems in Japan's digital media at large.

In Japan, the use of an online medium and SNS for political purposes is neither so widely exercised nor so influential as that of the United States or Korea. There is an Election Law that strictly limits the use of email services during the public election period. Only the officially recognized candidates and registered political parties can send emails calling for voting to their candidates. Unsolicited bulk emails calling for voting for a specific candidate or party is prohibited; candidates and parties who plan to send such campaign emails are mandated to obtain the consent of the addressees in advance in opt-in or opt-out manner.

Moreover, the general public is not that much interested in or affected by the use of these electronic media for political campaigns.[31] Therefore, the room for fake news or misinformation aimed to attack the opposing candidates is relatively small, which is why such methods are much less practiced than in some other countries.

---

29 "Tokyo court orders Oita man to pay ¥1.3 million in damages over 'vicious' racist comments against boy", *the japan times*, May 13, 2021, https://www.japantimes.co.jp/news/2021/05/13/national/crime-legal/tokyo-court-ruling-racist-comments/.

30 "*Internet and Hate Speech*", 13.

31 "Japan's first 'Internet election", *the japan times*, July 10, 2013, https://www.japantimes.co.jp/opinion/2013/07/10/editorials/japans-first-internet-election/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Of course, there still exist diverse kinds of offensive messaging and other online activities that could defame, offend, or provide fake news and mis-information in public. We have not yet observed well-organized online negative campaigns so far; they are mostly spontaneous and solitary ad hoc reactions and casual criminal acts for fun until today.

There is some hate speech and offensive speech against sexual, ethnic, and social minorities of various dimensions, but again they are less organized and more personal in general except in the case of hate speech against Koreans and also against Burakumins. Burakumins are ethnic Japanese people who were historically discriminated and still are targets of online hate speech. It is also a very serious and long-standing human rights violation issue in Japan.[32]

On the individual level, offensive bullying among juveniles, for example, or vicious speeches related to domestic violence using the Internet are often observed, and they have led to suicide or homicide cases at worst. Sexual seductions, illegal drug sales, and other anti-social uses of online media also exist and sometimes promoted by organized criminal groups. Phone or communication fraud ,especially targeting the senior citizens, by these criminal groups are rather serious and widespread.

Most large SNS platform operators are requested to monitor criminal use of their services, with varying degrees of regulatory mechanisms. Child pornography and direct seduction for committing suicide are strictly prohibited and could legally be filtered out online, while other forms of offensive or illegal messages are regulated on a more voluntary basis including "Notice and Takedown" process or legal measures in the court.

A new wave of fake news and misinformation was observed in 2020 with the outbreak of the COVID-19 pandemic in Japan. People had difficulties in finding accurate information, and a loft of false information that came from outside Japan was translated into Japanese and led to confusion. The government took some action and asked Internet platform providers such as Google and LINE to take measures to send notices of caution automatically once the term "Corona virus" was found in any use of online instances.

In any case, the issue of how to strike a balance between conflicting values such as freedom of expression vs. hate speech remain important, and we will examine the effectiveness of hard and soft approaches after

32 The Headquarters of Buraku Liberation league, "*What is Buraku Discrimination?*" Last modified: Dec 25, 2005 http://www.bll.gr.jp/en/index.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

discussing other areas of digital platform regulation approaches in the next Chapter.

Chapter 2: Privacy and Personal Data Protection and Economic Concerns

2.1. Economic concerns

The second area of digital platform regulation is the one of economic concerns. Referring to this, the obvious concerns are aimed at the excessive power and behaviours of the global Tech Giants such as Google, Amazon, Facebook, and Apple. Policy makers in Japan have been taking these concerns seriously for the past years and now they are starting to put some institutional measures to regulate the excessive behaviours on the digital platforms in domestic markets. Even though it is difficult to place a regulatory framework directly upon this challenge, the new platform regulation enacted in February 2021 can be interpreted as such a manifestation.

The global rankings of the market cap of large corporations are often referred to as the indicator of the economic strength (and weakness). In 1989, there were six Japanese companies among the global top ten as shown in the table. After more than three decades, there are no Japanese companies in the top ten in 2021, while all top five are American Big Tech companies with strong digital platform services, one from China, Alibaba, is also offering platform services, and one from Taiwan, TSMC, is supporting these digital platform infrastructures with its huge supply of semiconductors.

Table 1. Most Valuable Global Companies in 1989[33]

Rank	Company	Country	Full Market Cap (in USD M)
1	Industrial Bank of Japan	Japan	104,291.49
2	Sumitomo Bank	Japan	73,304.65
3	Fuji Bank	Japan	69,403.38
4	Dai-Ichi Kangyo Bank	Japan	64,036.45
5	Exxon Corp	United States	63,838.00

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

6	General Electric USA	United States	58,187.00
7	Tokyo Electric Power	Japan	56,499.62
8	IBM Corp	United States	55,656.99
9	Toyota Motor Corp.	Japan	53,251.22
10	American Tel & Tel	United States	48,951.00

Table 2. Most Valuable Global Companies in 2021[34]

Rank	Company	Country	Full Market Cap (in USD Bn)
1	Apple	United States	2,226.60
2	Microsoft	United States	1,901.40
3	Amazon	United States	1,660.00
4	Alphabet (Google)	United States	1,591.30
5	Facebook	United States	904.7
6	Berkshire Hathaway	United States	664.8
7	Tesla	United States	647.7
8	Alibaba	China	610.8
9	Taiwan Semiconductor Mfg. Co. (TSMC)	Taiwan	605.9
10	Visa	United States	495.1

With the sophisticated use of enormous amounts of online data and high capability of analysing and utilizing them with latest AI technologies, the Big Tech companies now have dominant positions in the global digital economy. The fear against the Big Tech companies can be considered as the strongest factor for the Japanese government to establish a new regulatory framework over the Digital Platform operators.

33 Steiger, Paul E., "What a difference 25 years makes", CNBC, April 29, 2014, https://www.cnbc.com/2014/04/29/what-a-difference-25-years-makes.html.

34 *Dogs of the Dow*, s.v. "Largest Companies by Market Cap Today", accessed June 4, 2021, https://www.dogsofthedow.com/largest-companies-by-market-cap.htm.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2.2. *The formation process of the "Act on improving Transparency and Fairness of Digital Platform"*

The Government initiated the policy discussion on digital platform regulation in 2018. The first action that led to establish the new rules to regulate the digital platformers was called for by the "Investments for the Future Strategy 2018", that was formally adopted by the Cabinet under the leadership of Prime Minister Shinzo Abe in June 2018.[35] This strategy mandated the government to formulate the basic design rules that guide the implementation of the regulatory framework by December 2018. Under this mandate, three agencies were engaged to analyse and implement the proper legal instruments aimed to provide a fair and effective regulatory framework for the digital platformer operations. [36]

In Japan, when a new regulatory framework is proposed, it is almost standard to designate one government agency in charge in general. In the case of digital platform regulation however, three agencies were assembled to cooperate. This is highly unusual and illustrates how complex the issue could be.

Hence the Ministry of Internal Affairs and Communications (MIC), Ministry of Economy, Trade, and Industry (METI), and Japan Fair Trade Commission (JFTC) are formally engaged. A basic design rule for setting the regulatory framework for digital platform operators were agreed. Those basic rules consisted of the following seven elements:

1. Legal evaluation viewpoints of digital platform operators
2. Promotion of proper development of digital platform operators
3. Establish transparency to ensure the fairness of digital platform operators
4. Establish fair and free competition among digital platform operators
5. Consider the rules for data portability and openness
6. Implement the balanced, flexible, and effective rules
7. Consider the international enforcement and harmonization method

---

35 Prime Minister and his Cabinet, *Joint Meeting of the Council on Economic and Fiscal Policy and the Council on Investments for the Future* (Cabinet Public Relations Office, 2018), https://japan.kantei.go.jp/98_abe/actions/201806/_00039.html.

36 Prime Minister and his Cabinet, *Future Investment Strategy 2018 (Draft)*, (Cabinet Public Relations Office, 2015), http://www.kantei.go.jp/jp/singi/keizaisaisei/dai28/siryou1.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

In January 2019, the JFTC conducted a comprehensive research on the existing practices of the digital platform operators and came out with the Interim Report in April 2019 [37] and the Final Report in October 2019. [38] This Final Report first provided the overview of the "digital platform" in our socio-economic life, emphasizing their strong positive impacts with innovations, analysing their "double-sided market nature" and "network effect" as well as "low marginal costs" and "the economy of scale" in economic terms. It further points out that digital platforms could produce enormous benefits with highly efficient use of large data, while they may also offer potential over-concentration to a few platform operators and may lead to monopolies or oligopolies and result in lock-in effects due to the high switching costs.

The Report continued to share concerns around competition policies, such as abuse of dominant position, exclusion of other platform operators, exclusion of competitive business users, and unfair coupling of digital platform operators to stifle competition. Based on research, this report highlighted some of the unfair practices found, such as unilateral change of rules by the dominant platform operator, unfair treatments, and excessive burden of shipping costs imposed to small and medium business users by the platform operators, or exclusive restrictions over competitive services by app platform operators. It also pointed out the potential abuse of transaction data by the platform operators; unfair treatments of business users by the operators, unilateral enforcement of "Most Favoured Nation status" or product pricing.

They also addressed the need for new mechanisms in addition to the aggressive enforcement of the existing anti-trust legal framework. Adopting the anti-trust laws with ex-post enforcement such as an exclusion order or penalty would require strict due process that may not be able to provide timely, flexible, and effective relief required for regulating the business practices over the new digital platforms. As for the methodology of the regulation, a "co-regulation" approach was proposed that would allow the voluntary effort of private sector players which will be supplemented by abstract codes and principles set by the law.

---

37 Japan Fair Trade Commission, "Interim report regarding trade practices on digital platforms", Japan Fair Trade Commission, last modified 2019, https://www.jftc.go.jp/en/pressreleases/yearly-2019/April/190417.html.

38 Japan Fair Trade Commission, "Report regarding trade practices on digital platforms (Business-to-Business transactions on online retail platform and app store)", Japan Fair Trade Commission, last modified 2019, https://www.jftc.go.jp/en/pressreleases/yearly-2019/October/191031.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2.3. *Act on Improving Transparency and Fairness of Digital Platforms (AITFDP) enacted*

It took two years to pass the new law "Act on Improving Transparency and Fairness of Digital Platforms (AITFDP)."[39] Under this Act, digital platform providers that meet the criteria stipulated under the Cabinet Order are obliged to disclose terms and conditions of trading, secure fairness in operating digital platforms, submit a report on the current situation of business operation with self-assessment every fiscal year. The government under the Minister of Economy, Trade, and Industry then makes an assessment of this report and publicizes the results.

The Act obligates METI to establish a system in which METI should request the JFTC to exercise certain measures under the Antimonopoly Act if METI finds any cases violating the Antimonopoly Act. The new Act also requires specified digital platform providers to give prior notices of any change thereof to the platform users.[40] The new Act sets the annual revenue in Japan as the benchmark to designate these platform players under the regulatory subject as specified providers.

In February 2021, five such specified operators are announced by the government. The first group consists of Amazon Japan, Rakuten and Yahoo! who offer comprehensive online services such as e-commerce sales, travel, banking and security services, as well as other numerous online services, making more than 300-billion-yen (USD 3bn) revenue per year. The second group consists of Apple and Google as mobile application providers or app stores with more than 200-billion-yen (USD 2bn) annual turnover. [41]

The obligations for the specified operators seem light:

i) disclose terms and conditions of trading, secure fairness in operating digital platforms,

---

39 Ministry of Economy, Trade and Industry, *Cabinet Decision on the Bill for the Act on Improving Transparency and Fairness of Digital Platforms* (Tokyo, Ministry of Economy, Trade and Industry, 2020), https://www.meti.go.jp/english/press/2020/0218_002.html.

40 "Japan's new law regulating tech giants' commerce platforms takes effect", *the japan times*, February 1, 2021, https://www.japantimes.co.jp/news/2021/02/01/business/tech/tech-giant-law-takes-effect/.

41 "Summary of a Bill on Improving Transparency and Fairness of Specifies Digital Platforms", https://www.kantei.go.jp/jp/singi/digitalmarket/pdf_e/documents_200218.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ii) submit a report on the current situation of business operation with self-assessment,
iii) give prior notices of any change thereof to the platform users. However, since "fairness" is not explicitly defined in this Act, there is room for interpretation and evaluation by the government.

If the METI Minister finds the report and its assessment not fair and publicly announces this, the operator will have to be *voluntarily* forced to change their terms and conditions in their own languages.

In other words, the government would not say "do this or do that", but the operators themselves must judge how to satisfy the government, and the public. This could be more difficult sometimes than to follow the explicit rule.

As the language of the new Act indicates, there is little room for strong enforcements but mostly voluntary actions to meet rather vague terms of "disclose information" and "secure fairness." This is very much the same approach as other Acts on Platform regulations, like the Act for Elimination of the Hate Speech or Act on the Protection of Personal Information (APPI).

2.4. Privacy and Personal Data protection

The third area of the policy concerns is of privacy and personal data protection. One of the challenges of establishing proper protection of personal data in Japan has been that there was no single unified regulatory system at work. The Act on the Protection of Personal Information (APPI) was established in 2003, but its narrowly segmented sectoral approach had been problematic with a large part of its implementation in practice left to each industry sector and their corresponding ministries.

To overcome these shortcomings of APPI, the Personal Information Protection Commission (PPC) was established as a central agency to manage the regulatory system under APPI in 2016 to provide the protection of the rights and interests of individuals while taking into consideration proper and effective use of personal information including "My Number", a national ID system for citizens. The PPC is an "independent organ in the Japanese legal framework."[42] The PPC has been working to improve the

42 Personal Information Protection Commission, "Personal Information Protection Commission", last modified 2016, https://www.ppc.go.jp/en/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

regulatory system and several revisions of the APPI have been implemented.

In 2020, the APPI received a major revision to cope with the increased use of digital data especially by the digital platform operators applying highly sophisticated "big data" and AI related technologies.[43] This new trend has created challenges for citizens to grasp the way their own rights are protected/infringed in advance. Thus, the new revisions tried to enhance protection for the individual rights including information disclosure proceedings, added obligations for business operators to include short-term data as the subject to protect and preserve, and electromagnetic (digitized) data was added as the form of information disclosure.

The benefits for business operators were also considered and the new articles on anonymous and pseudonymous information were added to the APPI that allow anonymously processed data to be shared by the third party, but not the pseudonymously processed data in general.

The gap between the central government and the local municipalities in terms of regulatory harmonization was also a big problem. There are more than 1,700 local governments in Japan that all have different rules or ordinances for the personal data protection procedures.

Now, the passage of the new package of digital reform laws on May 12, 2021 included the APPI's revision to close that gap. [44] The government now claims that Japan's personal data protection procedures will be streamlined across national and local governments and will have much higher efficiency for the benefit of all. However, some consumer advocates fear that the respect for privacy and human rights, which are often given higher priorities in local ordinances, may be compromised in the interest of the business use of the personal data once they are all unified under the new national system.[45]

---

43 Personal Information Protection Commission, "Promulgation of the Amendment Act of the Act on the Protection of Personal Information, etc.", last modified 2020, https://www.ppc.go.jp/en/news/archives/2020/20200618/.

44 "Japan passes laws to set up digital policy agency in September", *Nikkei Asia*, May 12, 2021, https://asia.nikkei.com/Politics/Japan-passes-laws-to-set-up-digital-policy-agency-in-September.

45 "どうなる?"個人情報保護制度"「デジタル改革関連法」成立", NHK, May 12, 2021, https://www3.nhk.or.jp/news/html/20210512/k10013026561000.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2.5. Transfer of personal data to a foreign country

It has been very difficult to regulate the use and transfer of personal data outside the jurisdiction. There is a strong concern that the global Big Tech, Google, Amazon, Facebook and Apple, for example, are collecting huge amounts of personal data via transaction, posting, or various forms of information search and retrieval and utilize them with effective advertising and sales beyond national regulatory control.

To cope with these challenges, the revision of the APPI in 2020 also added new restrictions on transfer of personal data to a third party in a foreign country. Yet these revisions will only become effective in 2022 and the details of new rules were not yet announced from the PPC thus creating ambiguous reactions from both consumer groups and the business community.

2.6. Tentative Conclusion

Since the Act on Improving Transparency and Fairness of Digital Platforms (AITFDP) has just been enforced in February 2021, it remains to be seen how effective the new regulatory framework will be. Some are again sceptical as the language is vague and basic, and they doubt it has any real effect of bringing the Japanese players on par to the Big Tech, which is the original aim of the policy and the strategy of the government and industry.

Prof. Takanori Ida of Kyoto University who is also the Chair of the Cabinet Working Group on Digital Market Competition Council said that the AITFDP adopted the "co-regulation" approach where the government set the basic framework while the details were left to the creativity and wills of the private sector. They are now starting to discuss the possible co-regulation on the Digital Advertising market as the third area of digital platform regulation. [46]

The Author believes that while this soft approach will not bring an immediate effect of making Japanese corporations viable in the global digital platform marketplace, it may urge the companies and their management to become more serious and aggressive in executing their business innovations that may take longer but produce more concrete outcomes.

---

46 Takanori Ida, *"New competition law for the digital platformers"*, in: *Horitsu no Hiroba (Legale Square)*, Tokyo, Gyosei, May 1, 2021

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The protection is one thing, self-reliance and bold moves are another. Strong will and commitment to the excellence should be, regardless of the amount of time it might take, placed as the core of the economic and political strategy Japan should undertake.

If we stood for the citizens' benefits, should we look for strict regulations and explicit enforcement mechanisms including heavy penalties, once the actions of a business enterprise or of a xenophobic group are found illegal?

In the case of HSEA, the financial penalty in national law may not be the most effective way to eliminate the root cause of the problem. It is the responsibility of citizens, who find these hate actions destructive to our society, to start campaigns against them and the stronger their voices are, the more effectively they can stop the undesired actions. The law merely "allows" or encourages these voices to be heard, and it clearly indicates where justice may be found.

These "co-regulation" approaches may suit Japan's social structure in the most productive way. They can also be seen as taking the multi-stakeholder approach, largely exercised among the Internet Governance policy circles.

Bibliography

Ayako Hatano, Shinji Higaki and Yuji Nasu. *Hate Speech and International Law*. Cambridge. Cambridge University Press: 2021. https://doi.org/10.1017/9781108669559.004.

Center for the Promotion of Human Rights Education and Encouragement. *外国人住民調査報告書－訂正版.* 2017. http://www.moj.go.jp/content/001226182.pdf.

Dogs of the Dow, s.v. "Largest Companies by Market Cap Today". https://www.dogsofthedow.com/largest-companies-by-market-cap.htm.

Gang, Deogsang. "Kanto Daishinsai (Kanto Great Earthquake)". Tokyo: Chuo-Koronsha, 1975.

Human Rights Protection Committee of Daini Tokyo Bar Association. *Internet and Hate Speech*. Gendai Jimbunsha. 2019.

Japan Fair Trade Commission. *Interim report regarding trade practices on digital platforms* Japan Fair Trade Commission. https://www.jftc.go.jp/en/pressreleases/yearly-2019/April/190417.html.

Japan Fair Trade Commission. *Report regarding trade practices on digital platforms (Business-to-Business transactions on online retail platform and app store)*. Japan Fair Trade Commission, 2019. https://www.jftc.go.jp/en/pressreleases/yearly-2019/October/191031.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Jiji, Kyodo. "Japan's new law regulating tech giants' commerce platforms takes effect". *the japan times*, February 1, 2021. https://www.japantimes.co.jp/news/2021/02/01/business/tech/tech-giant-law-takes-effect/.

Jiji, Kyodo. "Tokyo court orders Oita man to pay ¥1.3 million in damages over 'vicious' racist comments against boy". *the japan times*. May 13, 2021. https://www.japantimes.co.jp/news/2021/05/13/national/crime-legal/tokyo-court-ruling-racist-comments/.

Joji Mochida. "ヘイトスピーチは止まったか：川崎市が全国初の罰則付き条例" *Nippon*.com, November 12, 2020. https://www.nippon.com/ja/in-depth/d00648/.

Katsuo Yakura, Shinji Higaki and Yuji Nasu. *The Legislative Process Leading to the Hate Speech Elimination Act*. Cambridge: Cambridge University Press, 2021. https://doi.org/10.1017/9781108669559.011.

Kazushi Ogura, Shinji Higaki and Yuji Nasu. *Hate Speech on the Internet*. Cambridge: Cambridge University Press, 2021. https://doi.org/10.1017/9781108669559.019.

Kyodo. "Japan passes laws to set up digital policy agency in September". *Nikkei Asia*. May 12, 2021. https://asia.nikkei.com/Politics/Japan-passes-laws-to-set-up-digital-policy-agency-in-September.

Malina Andreia Pal. "The Japanese invasions of Korea: who was the real winner of the Imjin war?" January 2020, Geneva. https://www.researchgate.net/publication/344072575_The_Japanese_invasions_of_Korea_who_was_the_real_winner_of_the_Imjin_war.

Ministry of Economy, Trade and Industry. *Cabinet Decision on the Bill for the Act on Improving Transparency and Fairness of Digital Platforms*. Tokyo: Ministry of Economy, Trade and Industry, 2020. https://www.meti.go.jp/english/press/2020/0218_002.html.

Park Eun-sik. "韓国独立運動の血史 "*The Bloody History of the Korean Independence Movement*". (Heibonshya, 1972).

Personal Information Protection Commission. "Personal Information Protection Commission". https://www.ppc.go.jp/en/.

Personal Interview with Mr. Chun Kang Heon. Secretary. Culture Center Arirang in Shin-Okubo district of Shinjuku on May 11, 2021.

Prime Minister and his Cabinet. *Future Investment Strategy 2018 (Draft)*. Tokio: Cabinet Public Relations Office. 2015. http://www.kantei.go.jp/jp/singi/keizaisaisei/dai28/siryou1.pdf.

Prime Minister and his Cabinet. *Joint Meeting of the Council on Economic and Fiscal Policy and the Council on Investments for the Future*. Tokio: Cabinet Public Relations Office, 2018. https://japan.kantei.go.jp/98_abe/actions/201806/_00039.html.

Ryangok Ku, Shinji Higaki and Yuji Nasu. *The Current Movement of Hate Speech*. Cambridge: Cambridge University Press, 2021. https://doi.org/10.1017/9781108669559.006.

Shimbun, Kanagawa. "ヘイトデモ、我が街に通さず　川崎・桜本". *Kanagawa Newspaper*, 2015. https://www.kanaloco.jp/news/social/entry-67417.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Shinji Higaki and Yuji Nasu, eds. *Hate Speech in Japan*. Cambridge: Cambridge University Press, 2021.

Steiger, Paul E. "What a difference 25 years makes". CNBC, April 29, 2014. https://www.cnbc.com/2014/04/29/what-a-difference-25-years-makes.html

Takanori Ida, *"New competition law for the digital platformers"* in *Horitsu no Hiroba (Legale Square)*, Tokyo, Gyosei, May 1, 2021.

The Headquarters of Buraku Liberation league. *"What is Buraku Discrimination?"* Last update: Dec 25, 2005 http://www.bll.gr.jp/en/index.html.

Toru Mori, Shinji Higaki and Yuji Nasu. *An Injunction Banning a Xenophobic Group from Demonstrating, Kawasaki Case*. Cambridge: Cambridge University Press, 2021. https://doi.org/10.1017/9781108669559.015.

Toshihide Yamamura. *A Chronology of Events and Legislation Related to Hate Speech in Japan*. eds. Shinji Higaki and Yuji Nasu (Cambridge: Cambridge University Press, 2021), Appendix A, 723, https://doi.org/10.1017/9781108669559.022.

Ooba, Yasunori. "*Zainichi Kankoku and Chosenjin* (South and North Korean residents in Japan". Chuokoron Shinsha. 1993.

"Japan's first 'Internet election". *the japan times*. July 10, 2013. https://www.japantimes.co.jp/opinion/2013/07/10/editorials/japans-first-internet-election/.

Shinji Higaki. *The Hat Speech Elimination Act*. eds. Shinji Higaki and Yuji Nasu. Cambridge: Cambridge University Press, 2021. Chapter, 11, 368, https://doi.org/10.1017/9781108669559.012.

"Summary of a Bill on Improving Transparency and Fairness of Specifies Digital Platforms". https://www.kantei.go.jp/jp/singi/digitalmarket/pdf_e/documents_200218.pdf.

"どうなる? "個人情報保護制度"「デジタル改革関連法」成立". *NHK*, May 12, 2021. https://www3.nhk.or.jp/news/html/20210512/k10013026561000.html.

Annex: Historical Relationship between Japan and Korea

Average Japanese people today have very little knowledge of the history between Japan and Korea. The history with neighbouring Asian countries has been largely excluded in the formal school education, especially that of the modern history.

The following is a very short summary of major topics that may help to understand some unfortunate and conflicting elements, as the basic factors that led the hate speech attitudes of some Japanese and the counter-reactions of many Korean residents in Japan.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

A.1. Ancient age to Middle Age

There is evidence that certain parts of the primitive Japanese culture and society were shaped by the people who migrated from Korean Peninsula to Japanese archipelago in the ancient age. Hence there are many similarities in both cultures.

The first hostile or discriminatory attitudes of Japanese people against Koreans can be found in the feudal era when the ruler *Toyotomi Hideyoshi* launched two military invasions to Korean Peninsula in 1592 and in 1598. Both battles resulted in an ultimate retreat of Japan's army, but the cruel acts of Japanese warriors to Korean civilians are well known and remembered among the Korean people.[47]

After the Meiji Restoration that put an end to Japan's feudal system in the late 19th century, the new government first requested to open a formal diplomatic and trade relationship with Korea. However, Korea declined the request and chose to remain within the sinocentric regime. Based on those cornerstones, a political debate has risen in Japan whether and how to force Korea to accept Japan's request, which was then expanded into Imperialism over neighboring Asian countries such as China and Russia. The First Sino-Japanese War (1894-96) and the Russo-Japanese War (1904-1905), both resulted in Japan's victory, were essentially the fight over the control of Korean peninsula.

After these victories, Imperial Japan began its colonial aggression to Korea and then to "Manchuria" (Northeast region of China) and finally to central China. In 1911, Imperial Japan "annexed" Korea, effectively colonized Korea under military force. Korean people started the protest against Japan, including "March 1st Movement" in 1919 with the proclamation of Independence in the center of the capital city of Seoul and demonstrations in many other locations, 7,500 were killed, 16,000 wounded, and 46,000 arrested by the Japanese ruler.[48]

Japan's aggression was finally terminated at the end of the World War II in 1945 and Korea reclaimed the independence, yet divided into North and South until today.

---

47 Malina Andreia Pal, *The Japanese invasions of Korea: who was the real winner of the Imjin war?* January 2020, Geneva. https://www.researchgate.net/publication/344072575_The_Japanese_invasions_of_Korea_who_was_the_real_winner_of_the_Imjin_war.

48 Park Eun-sik, "*韓国独立運動の血史 The Bloody History of the Korean Independence Movement*, (Heibonshya, 1972).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

It must also be noted that in the aftermath of the Kanto Great Earthquake in September 1923, more than 6,000 ethnic Koreans living in Tokyo metropolitan areas were horribly killed by Japanese militias.[49] With some instigating languages in the martial law degree from the Interior Ministry, in addition to the local police and military troops, a wide range of civil members under the activities of resident association in local communities played a significant role of actual killings.[50]

A.2. Post WW II situation of Korean residents in Japan

There are more than half a million ethnic *Zainichi* Koreans living in Japan now. Most of them belong to the second, third or fourth generation of Koreans whose parents or ancestors came to Japan before the World War II. Many chose to migrate to earn better living, many more were "forced" to come for economic or military reasons. They were treated as the second-class citizens and have not been awarded with an actual equal status to those of the Japanese.

After Japan was defeated by the Allies in 1945, a significant number of Koreans went back to their motherland, especially among those who were forced to come to Japan against their wills. However, some chose to remain in Japan or had gone back to Korea once but decided to return to Japan, in part due to the severe socio-economic situation of Korean Peninsula caused by the Korean War in the early 1950s. [51]

The political divide between south and north along the Military Demarcation Line (MDL) started in 1952 further added complication among the Korean residents and their communities in Japan. Many Korean families whose origins were in the northern part started to "return" to North Korea in the 1960s, as the Democratic People's Republic of Korea (DPRK), or North Korea, strongly demanded Japanese government to facilitate the return program. Many Korean residents remained skeptical to the propaganda made by the communist government and however decided not to move.

---

49 Ryangok Ku, Chapter 5 *The Current Movement of Hate Speech,* in *Hate Speech in Japan* (Cambridge: Cambridge University Press, 2021), 218.

50 Gang, Deogsang, "*Kanto Daishinsai (Kanto Great Earthquake)*", (Tokyo, Chuo-Koronsha, 1975).

51 Yasunori Ooba, "*Zainichi Kankoku and Chosenjin* (South and North Korean residents in Japan", (Chuokoron Shinsha, 1993).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Japanese society at large has kept continuous discriminatory attitudes against Korean residents who were mostly kept outside the Japanese socio-economic system; received discriminatory treatments in education, healthcare and social welfare, employment, business opportunities to name a few.

The diplomatic relation between Japan and Korea, both south and north, has been tense, or counter-productive for many years, even after the Republic of Korea, or South Korea, restored their democracy. It also remained painful between North Korea and Japan even after Japan's Prime Minister Koizumi made a sudden visit to North Korea and shook hands with North Korea Leader Kim Jong-Il in 2002 and 2004.

The "Comfort women" issue during wartime had been the subject of diplomatic negotiations between two governments and the Japanese government made an official reflection and apology in 1996. In Korea, some victims and their supporters have taken the issue to court to demand compensation from the government of Japan under the Korean court. Both cases received judgement in favor of the plaintiffs, but the final solution is still uncertain.

These historical contexts affect the complex relationship and bitter sentiments between Koreans and Japanese. With all these social, economic, and political complications, it is the author's persuasion that the Japanese society never embraced the Korean residents in a warm and civil manner in full.

It should be also noted that the relationship of both countries, especially between their citizens is not entirely negative. There have been many cases where they communicate and collaborate with and respect each other very well. There still is a good basis to build a better world in the East Asia.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Social Media Platform Regulation in India – A Special Reference to The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021

Siwal Ashwini

Abstract: India has experienced the potential of social media platforms and witnessed the far-reaching consequences which these platforms may pose. The current Indian legal framework on social media platforms (hereinafter: SMPs) tend towards a co-regulatory model relying both on statutory framework and self-regulation of SMPs. The chapter analyses the regulatory framework of SMPs in India and the contentious "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021" (hereinafter IM Rules, 2021) from the prism of fundamentals of free speech. The chapter discusses how the free speech may get affected by the imposition of additional responsibilities like; appointment of India based compliance officers, first originator traceability requirements, deployment of automated filtering software, identification of physical address from the users of accounts, and other restrictions on SMPs through the recently notified IM Rules, 2021. These rules are alleged to be flouting certain key legal principles and are argued to be the outcome of legislative overreaching. Therefore, the IM Rules, 2021 warrants scrutiny from the perspective of free speech in the backdrop of above raised concerns.

Keywords: Social Media Intermediary, Significant Social Media Intermediary, Digital News Portals, Social Media Regulation in India and Intermediary Rules, 2021

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Introduction

State and private investments in communication technologies have resulted in an increased access to the Internet across South Asia[1]. India has around 530 million WhatsApp users, 410 million Facebook users, 160 million Twitter users, 448 million YouTube users by January 2021.[2] India, therefore, is not an exception to the penetration of social media and its rising popularity and usage among the varied segments of the Indian society.[3] According to certain scholars, the growing number of social media users in South Asia in general and in India in particular will play a critical role in shaping the trajectory of digital platforms, cultures, and politics in the coming years.[4] The nature, modus operandi of social media platforms and their regulatory frameworks are scarcely being deliberated or debated among the Indian communication scholars to the desired extent despite the meteoric increase in the number of social media users. Given the influence that platforms like Facebook, Twitter, YouTube and Amazon now wield on a global stage and with the growing number of users on social media and episodes of its misuse to spread hate speech, misinformation and political propaganda etc.[5], it has become crucial to granularly traverse and outline the role, nature, modus, and the regulatory framework of the digital intermediaries in India. Though, I firmly concede to the argument

---

1 Aswin Punathambekar and Sriram Mohan, "Introduction" *Global Digital Cultures: Perspectives from South Asia*, eds., Aswin Punathambekar and Sriram Mohan (Ann Arbor, MI: University of Michigan Press, 2019), doi. 3, https://doi.org/10.3998/mpub.9561751.

2 "Framework and Guidelines for Use of Social Media for Government Organisations", Department of Electronics and Information Technology, Ministry of electronics and Information Technology, Government of India, https://www.meity.gov.in/writereaddata/files/Approved%20Social%20Media%20Framework%20and%20Guidelines%20_2_.pdf.

3 Ankita Chakravarti, "Government reveals stats on social media users, WhatsApp leads while YouTube beats Facebook, Instagram", India *Today*, February 25, 2021, https://www.indiatoday.in/technology/news/story/government-reveals-stats-on-social-media-users-whatsapp-leads-while-youtube-beats-facebook-instagram-1773021-2021-02-25.

4 Punathambekar and Mohan, *Global Digital Cultures: Perspectives from South Asia*, doi. 3.

5 Shakuntala Banaji and Ram Bhat, "*WhatsApp Vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India*", https://blogs.lse.ac.uk/medialse/2019/11/11/whatsapp-vigilantes-an-exploration-of-citizen-reception-and-circulation-of-whatsapp-misinformation-linked-to-mob-violence-in-india/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

advanced by the scholars in their seminal work *"Global Digital Cultures: Perspectives from South Asia"* that the platforms being capitalist and imperialistic in nature, will seldom allow for such granular probe into their experiential engagement with state, industry, and user practices coalescing on these platforms.[6]

Still an attempt to traverse at least the regulatory framework and its effectiveness is certainly timely. The present study is a small endeavour to cursorily understand the regulatory framework of SMPs in India and the recently notified "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules (hereafter: IM Rules, 2021) under the Information Technology (Amendment) Act, 2008 (hereinafter: IT Act, 2008). The IM Rules, 2021 are not duly promulgated legislation by parliament. These are brought by the central government by purview of the rule making power under section 87 (zg) of the IT Act, 2008 which enables central government to issue guidelines to be followed by intermediaries in order to enjoy immunity from liability. The striking feature of the new rules is the imposition of new responsibilities on intermediaries. The study is limited in scope and focusses specifically on SMPs regulations under the IT Act, 2008 (which is primarily a law to regulate e-commerce) and its corresponding IM Rules, 2021.

Social Media in India: A prolegomenon

India's experience with this medium of information and disinformation communication has been mixed. It has viscerally seen its potential and far-reaching consequences in terms of caste and religion-based polarisation and the consequent episodes of murders and mob lynching.[7] India has also witnessed the power of SMPs in upholding free speech. A study of social media landscape in a specific part of the country is indicative of the prevalence of the social media even in the hinterland of this vast and divergent country.[8]

Though the legislative and policy framework relating to social media platforms in India has started emerging recently, yet it has received due

---

6 Punathambekar and Mohan, *Global Digital Cultures: Perspectives from South Asia,* Chap. 1, doc. 3.

7 Banaji and Bhat, **"***WhatsApp Vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India".*

8 Shriram Venkatraman, *Social Media in South India* (London: UCL Press, 2017), chap. 2, doc. 25-55, doi:10.2307/j.ctt1qnw88r.8.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

attention and consideration at the judicial level on the one hand and at the regulatory level on the other. Both the courts and sectoral regulators like Competition Commission of India have been continually active in deciphering, determining and validating the regulatory principles and the broad contours of social media platforms in India. At the time of this writing, there are multiple petitions challenging the legality of the social media platform regulations as well as antitrust complaints are lying sub-judice before various high courts and the competition commission of India, respectively. In order to assess the effectiveness of the platform regulations, gauging merely adequacy of the legislative and policy framework of social media platforms may not be a realistic approach given the fact that these platforms are continually evolving and intrinsically dynamic in nature. The focus, therefore, should be on gauging the promptness of the executive, judiciary, and sectoral regulators in taking up the task of enacting and ameliorating the existing framework within their jurisdictional contours. This expected promptness from the important pillars of democracy has always been debatable. The interfering role of executive in social media can very well be discerned by the often-invoked internet shutdowns, usually without well documented reasons, and usually with the tacit acquiescence of ISPs in abiding all directions of the government.[9] The recent farmer's protest and the invocation of internet shutdown orders speaks volumes about the governmental control on social media in India. Oxford Internet Institute identified India as one of the ten major countries of organised social media manipulation.[10]

Seventy percent of the Indian population is young and have recently got access to portable computing devices like smart phones, tabs etc. with poor social media literacy[11] The smart phone market in India has seen tremendous surge in sales of these devices. All smart phone users have social media accounts across all age groups and all regions in country. The hi-speed data availability at very reasonable cost has further proliferated

9 Joshita Pai and Nakul Nayak, "Initial Inputs for The Project on Freedom of Expression and The Private Sector in The Digital Age", Centre for Communication Governance, National Law University, Delhi, https://www.ohchr.org/Documents/Issues/Expression/PrivateSector/CentreCommunicationGovernance.pdf.

10 Samantha Bradshaw and Philip N. Howard *"Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation"*, Project on Computational Propaganda, Oxford Internet Institute, University of Oxford (2018), https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf.

11 Mark Linscott and Anand Raghuraman, *"Aligning India's Data Governance Frameworks"*, Atlantic Council, September 1, 2020. http://www.jstor.org/stable/resrep25999.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the number of users over last few years. Very few from these users have appropriate digital information literacy to understand the consequences of what they write, read and spread on social media. Even the educated masses in India have very bleak understanding of the legal consequences of misuse of this sphere.[12]

The present situation of social media illiteracy clubbed with not so efficacious legislative framework on social media in India brings home the question raised by one of the noted communication scholars: "as to who is representing India's vast populace via these platforms where these data-based public platforms are misused through bots, web robots etc. to script content"?[13] In the political sphere, the rise of political bots has created a new layer of computational propaganda on social life, gaining a newfound influence on the shaping of public opinion.[14] Given the social media illiteracy of Indian populace, the deluge of misinformation can rightly be termed as global public health threat as propounded by noted anthropologist *Heidi Larson* quoted in the study "*Addressing Misinformation through Intermediary Liability Policy, Platform Design Modification, and Media Literacy*".[15]

In an environment of scepticism like this, it is natural for politicians, policy bureaucrats, and the social media companies to shape the present framework in their favour.[16] The most recent and relevant instances of this misuse of the information ecosystem during the COVID-19 outbreak in India are the deletion of tweets of several influential persons targeting the

---

12 Ben Medeiros and Pawan Singh, "Addressing Misinformation on WhatsApp in India Through Intermediary Liability Policy, Platform Design Modification, and Media Literacy," *Journal of Information Policy*, Vol. 10 (2020): 288, https://www.jstor.org/stable/pdf/10.5325/jinfopoli.10.2020.0276.pdf?refreqid=excelsior%3A470ea419086d439103e5a165185e48ff.

13 Payal Arora, "Politics of Algorithms, Indian Citizenship, and the Colonial Legacy" in *Mapping Global Digital Cultures, in Global Digital Cultures: Perspectives from South Asia, eds.,* Aswin Punathambekar and Sriram Mohan (Ann Arbor, MI: University of Michigan Press, 2019), chap.1, doc. 41, https://doi.org/10.3998/mpub.9561751.

14 Joyojeet Pal, "The Making of a Technocrat: Social Media and Narendra Modi" in Mapping *Global Digital Cultures: Perspectives from South Asia, eds.,* Aswin Punathambekar and Sriram Mohan (Ann Arbor, MI: University of Michigan Press, 2019), chap.7, doc.163-183, https://doi.org/10.3998/mpub.9561751.

15 Medeiros and Singh, "Addressing Misinformation on WhatsApp in India,"277.

16 Niranjan Sahoo, "Mounting Majoritarianism and Political Polarisation in India" in *Political Polarization in South and Southeast Asia: Old Divisions, New Dangers*, eds., Thomas Carothers and Andrew O'Donohue (Carnegie Endowment for International Peace, 2020), chap.1, doc.9. http://www.jstor.org/stable/resrep26920.7.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

government's mishandling of the second wave of COVID-19 outbreak in India[17] and the demand by the Indian government to remove references to the 'Indian Variant' of COVID-19 from all SMPs.[18] The deletion of individual's account or taking down protected expressions by a unilateral decision of SMPs calling it a "Bad Content" under their community standards, or at the behest of the government in the absence of clear and settled law violates the right to free expression which is guaranteed in the Constitution of India. Review mechanism of SMPs, their modus operandi and their impartiality also remains a matter of concern. SMPs are found to be favouring the ruling parties[19], contrary to their claim that their content review decisions are made in the best interest of the community and not for commercial political reasons.[20] Recently, SMPs also claimed to have constituted oversight groups to hear appeals and challenges on content deletion and moderation, but the formation of these oversight groups generally remains contentious across the globe and India is not an exception.[21]

The Legal Framework of Social Media Platforms in India

There is no duly promulgated legislation to govern SMPs in India. Talks to have an exclusive and omnibus law to cater to this unique and challenging platform of information communication have not yet started in India. Yet, there are a few statutes in force which do address the issues pertaining to SMPs directly and indirectly in India. Primarily, India relies on the Constitution of India for tenets on privacy and free speech on any medi-

17 Jen Patja Howell, The Lawfare Podcast;' India v. Platforms, June 3rd, 2021, https://www.lawfareblog.com/lawfare-podcast-india-v-platforms.

18 Billy Perrigo, "India Is Demanding Social Media Remove References to the 'Indian Variant' of COVID-19. But What Should It Be Called?," Time, May 26, 2021, https://time.com/6051039/indian-variant-social-media/.

19 Sangeeta Mahapatra and Johannes Plagemann, "*Polarisation and Politicisation: The Social Media Strategies of Indian Political Parties,*" (German Institute of Global and Area Studies (GIGA), 2019). http://www.jstor.org/stable/resrep24806.

20 Shubham Verma, "Facebook briefly hid posts calling for PM Modi's resignation by mistake, govt responds," *India Today*, April 29, 2021, https://www.indiatoday.in/technology/news/story/facebook-hid-posts-calling-for-pm-modi-s-resignation-briefly-says-it-was-a-mistake-1796123-2021-04-29.

21 Dipayan Ghosh, "Are We Entering a New Era of Social Media Regulation?", *Harward Business Review*, January 14, 2021, https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

um[22], the IT Act, 2008[23] and the IM Rules 2021.[24] Apart from the statutory regulations, the SMPs have their own self regulations. The current Indian legal framework on SMPs tend towards a co-regulatory model relying both on statutory framework and self-regulation of SMPs.

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021: A critical Analysis (Illustrative, not exhaustive)

As highlighted above, in the absence of an exclusive and omnibus legislation, SMPs are governed by the IT Act, 2008 and its corresponding IM Rules, 2021 in India which ought to be necessarily framed within the contours of the Constitution of India and its basic tenets.[25] IM Rules, 2021 [26] are the new and elaborate rules which have replaced and repealed the Information Technology (Intermediaries Guidelines) Rules, 2011 (hereinafter called IM Rules, 2011), which according to some noted scholars were also not in consonance with the international best practices so far as the issues of safe harbour to Internet service providers and intermediaries'

---

22 See Article 19 and Article 21 of the Indian Constitution envisaging certain inviolable rights to the citizens and any law, order, byelaw must be necessarily in consonance of the same.

23 The Information Technology Act, 2008, "*An Act to provide legal recognition for transactions carried out by means of electronic data interchange and other means of electronic communication, commonly referred to as — electronic commerce, which involve the use of alternatives to paper-based methods of communication and storage of information, to facilitate electronic filing of documents with the Government agencies and further to amend the Indian Penal Code, the Indian Evidence Act, 1872, the Banker's Books Evidence Act, 1891 and the Reserve Bank of India Act, 1934 and for matters connected therewith or incidental thereto*". Available at: https://www.meity.gov.in/content/information-technology-act-2000

24 The IT Rules, 2021 are purportedly made under Section 87(1) of the parent Act, more particularly Section 87(2) (y), (z), (zb) and (zg) of The Information Technology Act, 2008, https://www.meity.gov.in/content/information-technology-act-2000.

25 The Indian Constitution is the paramount source of law in the country. The Indian Constitution is the groundnorm which is ought to be obeyed and any law/rule/byelaw needs to abide by its basic tenets.

26 "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021," Ministry of Electronics and Information Technology, Government of India, February 25, 2021, https://www.meity.gov.in/writereaddata/files/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

liability are concerned.[27] Though, a robust jurisprudence on the same evolved in a plethora of cases which reached to the apex court of the country and were decided with finality.[28] The IM Rules, 2011 were completely silent on the aspect of SMPSs, OTT Platforms and Digital Media Platforms and there are no decided cases where the IM Rules 2011 were invoked by the apex court to regulate the SMPs in India.

Regulating Social Media Intermediaries and Digital Media together: An Incongruous Approach

As is evidently clear from the nomenclature of the IM Rules 2021, that these rules are not restricted in the application to internet intermediaries alone. Rather the rules are meant to provide guidelines for digital media also. This is, as is being contended in the Indian courts right now, out of the purview of the IT Act, 2008.[29] The rules are portrayed to be premised on a balanced approach and are meant to do predominantly two major things so far as SMPs are concerned: firstly, to provide the legal definition to the complex concept of social media and to classify social media platforms from significant social media platforms (the previous IM Rules 2011 were silent on it) and, secondly; to regulate the same by imposition of additional operational responsibilities with the hybrid model of self-regulation and governmental control.

While framing these rules the government claims to have taken into consideration the growing prominence of social media platforms in India and the relevant societal implications of the content being transmitted on these platforms on the one hand and the freedom of expression as

27 Risabh Bailey, "Censoring the Internet: The New Intermediary Guidelines." *Economic and Political Weekly* 47, no. 5 (2012): 15-19, http://www.jstor.org/stable/41419840.

28 Pritika Rai Advani, "Intermediary Liability in India," *Economic and Political Weekly* 48, no. 50 (2013), http://www.jstor.org/stable/24479053.

29 The Wire Staff, "Why the Wire Wants the New IT Rules Struck Down", *The Wire*, March 9, 2021, https://thewire.in/media/why-the-wire-wants-the-new-it-rules-struck-down.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

well as privacy on the other hand. [30] These rules are being portrayed as "progressive, liberal and contemporaneous.[31]

But even prior to the notification of new IM Rules on February 25th, 2021, the concerns relating to freedom of speech and privacy were raised by the different sections of the society.[32] Concerns also arose about the way the IM Rules, 2021 were framed by the concerned ministry/ies of the Government of India.[33]

The new rules also bring many entities, including curated-content platforms such as Netflix and Amazon Prime as well as digital news publications, into the definition of intermediaries which were out of the purview of the definition of intermediaries in the previous IM Rules 2011. Therefore, new rules are inviting lot of criticism from different sections of the stakeholders for being incongruous in approach.[34]

Flawed Assumptions

The IM Rules, 2021 seems to be based on the assumption that the social media intermediaries are no longer acting like mere conduits, rather they are accomplice with the publishers and content creators. This flawed assumption can also be inferred from the introductory remarks of the IM Rules, 2021 on the press information bureau website, where the language is aptly clear that SMPs owe accountability against its misuse and abuse

---

30 Shishir Gupta, "I and B ministry starts work on self-regulation law for OTT platforms, online news," January 16, 2021, *India News*, https://www.hindustantimes.com/india-news/ib-ministry-works-on-self-regulation-law-for-ott-platforms-and-digital-media-101610774195596.html.

31 "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021," Ministry of Electronics and Information Technology, Government of India, February 25, 2021, Press Information Bureau, https://pib.gov.in/PressReleseDetailm.aspx?PRID=1700749.

32 "India must resist the lure of the Chinese model of online surveillance and censorship #IntermediaryRules #RightToMeme #SaveOurPrivacy," *Internet Freedom Foundation*, December 24, 2018, https://internetfreedom.in/india-must-resist-the-lure-of-the-chinese-model-of-surveillance-and-censorship-intermediaryrules-righttomeme-saveourprivacy/.

33 "Latest Draft Intermediary Rules: Fixing big tech, by breaking our digital rights?," *Internet Freedom Foundation*, February 25, 2021, https://internetfreedom.in/latest-draft-intermediary-rules-fixing-big-tech-by-breaking-our-digital-rights/.

34 Torsha Sarkar, "New intermediary guidelines: The good and the bad", *Down to earth* February 26, 2021, https://www.downtoearth.org.in/blog/governance/new-intermediary-guidelines-the-good-and-the-bad-75693.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

by users.[35] Under this assumption, the intermediaries are over-burdened with the imposition of additional responsibilities like; appointment of India based compliance officers, first originator traceability requirements, deployment of automated filtering software and identification of physical address from the users of accounts, swift take down of content etc. This is not welcome by most of the social media giants in India.[36]

The above said obligations may undermine the right to free speech. The requirement of rapid removal and monitoring of user's content is one such obligation which may prompt the SMPs to over comply with take down requests to preclude any liability. The deployment of automated filtering software may also not prove effective because it is unlikely to identify the unlawful content in different cultural backgrounds through software. The obligation of originator traceability has its own privacy related ramifications (explained in more detail in the ensuing part).

In case of non-observance of these rules, intermediaries will lose the safe harbour provided under section 79 of the IT Act, 2008[37] and face criminal sanctions which may lead these SMPs to remove even lawful content as a precautionary measure.

Even the United Nations special rapporteurs have written to the government of India expressing concerns about the newly notified IM Rules, 2021, and asked the Indian government to carry out a detailed review and consult with all relevant stakeholders. They have shown serious concerns about due diligence obligations and expressed apprehensions of serious infringement of human rights because of the newly notified IM Rules, 2021.[38]

Prior to the notification of these new rules, intermediaries were provided with a comprehensive protection to any liability arising from third party publications or information made available by them. The purpose to provide exemption from liability helps the intermediary to operate

---

35 "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021," Ministry of Electronics and Information Technology, Government of India, February 25, 2021, Press Information Bureau, https://pib.gov.in/PressReleseDetailm.aspx?PRID=1700749.

36 Aashish Aryan, "Twitter interim resident grievance officer resigns", *The Indian Express*, July13th, 2021, https://indianexpress.com/article/india/twitter-interim-grievance-officer-for-india-quits-amid-row-over-new-it-rules-7378424/.

37 Rule 7 of the IM Rules, 2021.

38 Neha Alawadhi, "UN Special Rapporteurs write to govt against IT Rules, ask for review", Business standard, June 19th, 2021, https://www.business-standard.com/article/economy-policy/un-special-rapporteurs-write-to-govt-against-it-rules-ask-for-review-121061801338_1.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

without any intervention, but with the new IM Rules, 2021, there would be stringent norms to be adhered to by the intermediaries to avail of the safe harbour.[39] The IM Rules, 2021 in a marked departure from its predecessor IM Rules, 2011[40] delineates social media intermediaries (hereinafter referred to as SMIs)[41] and significant social media intermediaries (hereinafter referred to as SSMIs)[42] as a separate class within the ambit of definitional clause of IM Rules, 2021.[43] As discussed above, the IM Rules, 2021 brings in additional and onerous due diligence obligations to be followed by SSMIs.[44] The IM Rules also impose criminal sanctions for non-observance of the additional due diligence which appears as a disproportionate consequence, a restraint on free speech facilitated by intermediaries and may lead to chilling effect.[45]

---

39 Garima Jhunjhunwala and Prashant Kumar, "Developments in India—Website Owner and Service Provider Liability for User-Generated Content and User Misconduct," *The Business Lawyer* 70, no. 1 (2014): 307-12 http://www.jstor.org/stable/43665705.

40 "Intermediary" means an intermediary as defined in clause (w) of sub-section (1) of section 2 of the Information Technology Act, 2008; "intermediary", with respect to any particular electronic records, means any person who on behalf of another person receives, stores or transmits that record or provides any service with respect to that record and includes telecom service providers, network service providers, internet service providers, web-hosting service providers, search engines, online payment sites, online-auction sites, online-market places and cyber cafes.

41 Rule 2(1) (v) 'Social Media Intermediary' means an intermediary which primarily or solely enables online interaction between two or more users and allows them to create, upload, share, disseminate, modify or access information using its services.

42 Rule 2(1) (w) 'Significant Social Media Intermediary' means a social media intermediary having number of registered users in India above such threshold as notified by the Central Government. The present threshold for significant social media intermediary is five million, See notification: https://www.meity.gov.in/writereaddata/files/Gazette%20Significant%20social%20media%20threshold.pdf.

43 "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021," Ministry of Electronics and Information Technology, Government of India, February 25, 2021, Press Information Bureau, https://pib.gov.in/PressReleseDetailm.aspx?PRID=1700749.

44 Rule 4, "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021," https://pib.gov.in/PressReleseDetailm.aspx?PRID=1700749.

45 Rule 7, "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021," https://pib.gov.in/PressReleseDetailm.aspx?PRID=1700749.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The inclusion of SMIs and SSMIs is a welcome step in the IM Rules 2021, but it also raises several doubts on the intention of the government especially on the aspect of regulating the SSMIs like WhatsApp, Telegram and Signal who are primarily involved in providing messaging services. According to IM Rules 2021 these messaging SSMIs are now required to enable the identification of the first originator of the information on its computer resource on a judicial order passed by a court of competent jurisdiction or an order passed under section 69 of the IT Act, 2008[46] by the competent authority as per the Information Technology (Procedure and Safeguards for interception, monitoring and decryption of information) Rules, 2009[47], which shall be supported with a copy of such information in

---

46 Section -69- Information Technology Act, 2008.- Power to issue directions for interception or monitoring or decryption of any information through any computer resource. -
(1) Where the Central Government or a State Government or any of its officers specially authorised by the Central Government or the State Government, as the case may be, in this behalf may, if satisfied that it is necessary or expedient to do in the interest of the sovereignty or integrity of India, defence of India, security of the State, friendly relations with foreign States or public order or for preventing incitement to the commission of any cognizable offence relating to above or for investigation of any offence, it may, subject to the provisions of sub-section (2), for reasons to be recorded in writing, by order, direct any agency of the appropriate Government to intercept, monitor or decrypt or cause to be intercepted or monitored or decrypted any information generated, transmitted, received or stored in any computer resource.
(2) The procedure and safeguards subject to which such interception or monitoring or decryption may be carried out, shall be such as may be prescribed.
(3) The subscriber or intermediary or any person in-charge of the computer resource shall, when called upon by any agency referred to in sub-section (1), extend all facilities and technical assistance to-
(a) provide access to or secure access to the computer resource generating, transmitting, receiving or storing such information; or
(b) intercept, monitor, or decrypt the information, as the case may be; or
(c) provide information stored in computer resource.
(4) The subscriber or intermediary or any person who fails to assist the agency referred to in sub-section (3) shall be punished with imprisonment for a term which may extend to seven years and shall also be liable to fine.]

47 Another set of corresponding rules made by central government under IT Act, 2008 for blocking etc. Information Technology (Procedure and Safeguards for interception, monitoring and decryption of information) Rules, 2009, https://www.meity.gov.in/writereaddata/files/Information%20Technology%20%28Procedur

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

electronic form.[48] The requirement of identification of the first originator of the message under the new rules, though is subject to certain provisos appended to the said rule and of course subject to judicial decisions in this regard,[49] is still problematic and seems to be a threat to the privacy of users of messaging services. The issue of traceability of the originator of a message on messaging platforms and the ineffectiveness of these platforms to facilitate traceability already reached the apex court via public interest litigation relating to linking of Aadhar[50] with social media accounts even prior to notification of these IM Rules, 2021.[51] Some of the studies have suggested that these platforms are vulnerable to falsification of originator information by bad actors to frame an innocent person for sending the illegal message.[52] The concern of traceability as envisaged under rule 4(2) of the IM Rules, 2021 further deepens in the presence of reliable studies questioning the reliability of the end-to-end encrypted platforms.[53]

An Inchoate Attempt

The IM Rules, 2021 also invites criticism on account of being myopic, the rules focus solely on SMIs and have not attempted to define and differentiate other variants of intermediaries like telecom service providers, network service providers, internet service providers, web-hosting service providers, search engines, online payment sites, online-auction sites, online-market places and cyber cafes which form part of the definition of intermediary

---

e%20and%20Safeguards%20for%20Interception%2C%20Monitoring%20and%20Decryption%20of%20Information%29%20Rules%2C%202009.pdf.

48 Rule 4(2) of the IM Rules, 2021.

49 Sunil Abraham, "Shreya Singhal and 66A: A Cup Half Full and Half Empty," *Economic and Political Weekly 50*, no. 15 (2015), http://www.jstor.org/stable/24481877.

50 Aadhaar number is a 12-digit random number issued by the 'Unique Identification Authority of India' to the residents of India after taking the biometric and iris information of the applicant, "Unique Identification Authority of India," Government of India, https://uidai.gov.in/what-is-aadhaar.html.

51 *Antony Clement Rubin v. Union of India (T.C. Civil No.189 of 2020),* https://indiankanoon.org/doc/37202571/.

52 Manoj Prabhakaran, "On a Proposal for Originator Tracing on WhatsApp," *An Independent Expert Report*, https://drive.google.com/file/d/1vivciN8tNSbOrA9eZ8Ej0mCAUBzRWu5N/view.

53 Manoj Prabhakaran, "On a Proposal for Originator Tracing on WhatsApp," https://drive.google.com/file/d/1vivciN8tNSbOrA9eZ8Ej0mCAUBzRWu5N/view.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

as per section 2(1) (W) of the IT Act, 2008. In most of the advanced legal systems of the world the above-mentioned intermediaries are well defined which helps in regulating the same in a very systematic manner.

Regulating Overzealously

Regulating digital news portals, who are just publishers of news and current affairs content and are completely different from publishers of online curated content like OTT platforms, appears to be an overreach and beyond the scope of the IT Act, 2008. The IM Rules, 2021 classify digital news portals as "Digital Media" and seek to regulate these news portals by imposing government control and code of ethics.[54] Many of these digital news portals have approached the courts and challenged the constitutionality of the IM Rules, 2021 for overreaching the IT Act, 2008 which nowhere provides any provision for regulating the non-intermediary entities like digital news portals.[55] At the time of this writing, the Delhi high court has issued notice to the central government to file a reply to the present petition and further directed the government to give reasons as to why the operation of the rules should not be stayed. The government's reply is awaited in this regard. Another similar petition is also pending before the High Court of Delhi wherein Part III of the IM Rules, 2021 has been challenged for being ultra vires the IT Act, 2008 in as much as the classification of 'publishers of news and current affairs content' ("digital news portals") as part of 'digital media' is concerned.

The classification of digital news portals as digital media, which are integral to uphold the freedom of speech and expression in every democracy of the world, appears to be an overreach of the power vested with the central government under section 87 of the IT Act, 2008 for the reason that the objective of IT Act, 2008 is to facilitate e-commerce and validate electronic transactions only.[56] There does not lie any legislative backing in this move of the central government because these news portals are not intermediaries in strict sense under the IT Act, 2008. This makes these guidelines a camouflaged way of regulating online news portals through

54 Rule 2(i) of the IM Rules, 2021.

55 *Quint Digital Media Limited and Anr. V. Union of India and Anr, 2021*, https://www.livelaw.in/pdf_upload/the-quint-delhi-hc-petition-it-rules-390804.pdf.

56 Karen Kornbluh and Ellen P. Goodman,"Safeguarding Digital Democracy: Digital Innovation and Democracy Initiative Roadmap." *German Marshall Fund of the United States*, 2020. http://www.jstor.org/stable/resrep24545.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a delegated legislation by bringing these portals under the aegis of the IT Act, 2008 without following the due process of parliamentary scrutiny.[57]

Conclusion

Multiple petitions are pending on constitutionality and overreach of the IT Act, 2008 in framing these rules. WhatsApp has also reached the court challenging these rules on the ground of user's privacy.[58] Based on the illustrative analysis attempted in the preceding paragraphs, it appears that these rules are slightly disproportionate and lack the requisite democratic approach of SMPs governance and legislative backing. The IM Rules, 2021 trivialises the opportunity of bringing more comprehensive and realistic regulatory framework and then to providing a level playing field to the intermediaries especially the SMPs. Rather the rules have created an environment of fear and panic amongst the SSMIs.

Bibliography

Abraham, Sunil. "Shreya Singhal and 66A: A Cup Half Full and Half Empty." *Economic and Political Weekly 50*, no. 15 (2015): 12-16. http://www.jstor.org/stable/24481877.

Advani, Pritika Rai. "Intermediary Liability in India." *Economic and Political Weekly* 48, no. 50 (2013): 120-128. http://www.jstor.org/stable/24479053.

Antony Clement Rubin v. Union of India (T.C. Civil No.189 of 2020). https://indiankanoon.org/doc/37202571/.

Bailey, Risabh. "Censoring the Internet: The New Intermediary Guidelines." *Economic and Political Weekly* 47, no. 5 (2012): 15-19. http://www.jstor.org/stable/41419840.

Banaji, Shakuntala., and Ram Bhat. "WhatsApp Vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India", November 11, 2019, https://blogs.lse.ac.uk/medialse/2019/11/11/whatsapp-vigilantes-an-exploration-of-citizen-reception-and-circulation-of-whatsapp-misinformation-linked-to-mob-violence-in-india/.

57 Norms of Journalistic Conduct of the Press Council of India under the Press Council Act, 1978 regulates newspapers in India.

58 Joseph Menn, "WhatsApp sues Indian government over new privacy rules", May 26th, 2021, https://www.reuters.com/world/india/exclusive-whatsapp-sues-india-govt-says-new-media-rules-mean-end-privacy-sources-2021-05-26/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Bradshaw, Samantha and Philip N. Howard. *"Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation"*, Project on Computational Propaganda, Oxford Internet Institute, University of Oxford (2018). https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf.

Chakravarti, Ankita. "Government reveals stats on social media users, WhatsApp leads while YouTube beats Facebook, Instagram", *India Today*, February 25, 2021, https://www.indiatoday.in/technology/news/story/government-reveals-stats-on-social-media-users-whatsapp-leads-while-youtube-beats-facebook-instagram-1773021-2021-02-25.

Daskal, Jennifer. "Speech Across Borders." *Virginia Law Review 105, no. 8* (December 2019): 1605-1666. https://www.jstor.org/stable/26891057.

Department of Electronics and Information Technology, Ministry of Electronics and Information Technology, Government of India. "Framework and Guidelines for Use of Social Media for Government Organisations." https://www.meity.gov.in/writereaddata/files/Approved%20Social%20Media%20Framework%20and%20Guidelines%20_2_.pdf.

Foundation for Independent Journalism and Ors. V. Union of India and Anr, 2021. https://www.medianama.com/wp-content/uploads/2021/03/Foundation-for-independent-journalism-petition-Delhi-HC-redacted.pdf.

Ghosh, Dipayan. "Are We Entering a New Era of Social Media Regulation?." *Harvard Business Review*, January 14, 2021, https://hbr.org/2021/01/are-we-entering-a-new-era-of-social-media-regulation.

Gupta, Shishir. "I and B ministry starts work on self-regulation law for OTT platforms, online news." *India News*, January 16, 2021, https://www.hindustantimes.com/india-news/ib-ministry-works-on-self-regulation-law-for-ott-platforms-and-digital-media-101610774195596.html.

Jhunjhunwala, Garima. and Prashant Kumar. "Developments in India—Website Owner and Service Provider Liability for User-Generated Content and User Misconduct." *The Business Lawyer* 70, no. 1 (2014): 307-312. http://www.jstor.org/stable/43665705.

Kornbluh, Karen, Ellen P. Goodman. "Safeguarding Digital Democracy: Digital Innovation and Democracy Initiative Roadmap." Report. *German Marshall Fund of the United States*, March 1, 2020. http://www.jstor.org/stable/resrep24545.

Mahapatra, Sangeeta and Johannes Plagemann. *"Polarisation and Politicisation: The Social Media Strategies of Indian Political Parties,"* (German Institute of Global and Area Studies (GIGA), 2019. http://www.jstor.org/stable/resrep24806.

Medeiros, Ben and Pawan Singh. "Addressing Misinformation on WhatsApp in India Through Intermediary Liability Policy, Platform Design Modification, and Media Literacy." *Journal of Information Policy*, Vol. 10 (2020): 276-298. https://www.jstor.org/stable/pdf/10.5325/jinfopoli.10.2020.0276.pdf?refreqid=excelsior%3A470ea419086d439103e5a165185e48ff.

Menn, Joseph. "WhatsApp sues Indian government over new privacy rules". May 26th, 2021. https://www.reuters.com/world/india/exclusive-whatsapp-sues-india-govt-says-new-media-rules-mean-end-privacy-sources-2021-05-26/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Ministry of Electronics and Information Technology, Government of India. "Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021." February 25, 2021. https://www.meity.gov.in/writereaddata/files/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf.

Pai, Joshita and Nakul Nayak, "*Initial Inputs for The Project on Freedom of Expression and The Private Sector in The Digital Age*", Centre for Communication Governance, National Law University, Delhi. https://www.ohchr.org/Documents/Issues/Expression/PrivateSector/CentreCommunicationGovernance.pdf.

Perrigo, Billy. "India Is Demanding Social Media Remove References to the 'Indian Variant' of COVID-19. But What Should It Be Called?," *Time*, May 26, 2021, https://time.com/6051039/indian-variant-social-media/.

Prabhakaran, Manoj. "On a Proposal for Originator Tracing on WhatsApp." *An Independent Expert Report*. https://drive.google.com/file/d/1vivciN8tNSbOrA9eZ8Ej0mCAUBzRWu5N/view.

Punathambekar, Aswin and Sriram Mohan, eds. *Global Digital Cultures: Perspectives From South Asia*. MI: University of Michigan Press, 2019. https://doi.org/10.3998/mpub.9561751.

Quint Digital Media Limited and Anr. V. Union of India and Anr, 2021. https://www.livelaw.in/pdf_upload/the-quint-delhi-hc-petition-it-rules-390804.pdf.

Sahoo, Niranjan. "Mounting Majoritarianism and Political Polarisation in India." In *Political Polarization in South and Southeast Asia: Old Divisions, New Dangers*, edited by Thomas Carothers and Andrew O'Donohue, 9-24. Carnegie Endowment for International Peace, 2020. http://www.jstor.org/stable/resrep26920.7.

Sarkar, Torsha. "New intermediary guidelines: The good and the bad." *Down to earth,* February 26, 2021, https://www.downtoearth.org.in/blog/governance/new-intermediary-guidelines-the-good-and-the-bad-75693.

Unique Identification Authority of India. Government of India. https://uidai.gov.in/what-is-aadhaar.html.

Venkatraman, Shriram. *Social Media in South India.* London: UCL Press, 2017. https://www.jstor.org/stable/j.ctt1qnw88r.

Verma, Shubham. "Facebook briefly hid posts calling for PM Modi's resignation by mistake, govt responds," *India Today*, April 29, 2021, https://www.indiatoday.in/technology/news/story/facebook-hid-posts-calling-for-pm-modi-s-resignation-briefly-says-it-was-a-mistake-1796123-2021-04-29.

Why the Wire Wants the New IT Rules Struck Down. *The Wire*, March 9, 2021, https://thewire.in/media/why-the-wire-wants-the-new-it-rules-struck-down.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Thoughts on the Regulation of Content on Social Media in Latin America: Authors' Rights, Limitations and Content Filtering

Maria L. Vazquez, Maria Carolina Herrera Rubio, Alejandro Aréchiga Morales

Abstract: The array of issues involved in the regulation of social media, and their cardinal importance in Latin America - from privacy, freedom of expression, liabilities, disinformation, right to erasure, and copyright - is compelling, yet staggering in its expanse and substance. This review is not meant to be exhaustive, but it does provide thoughts and selected examples on the regulatory framework of social media content in Latin America. The first part of this paper aims to give an overview of some of the more noteworthy developments regarding internet and social media regulations in certain countries. Then, the analysis will focus on authors' rights in Latin America, analyzing how successfully – or not – traditional exceptions and limitations of international copyright law regulate copyrighted content on social media.

Keywords: Social Media, Latin America, Copyright, Limitations, Content, Filtering, Regulation

Chapter 1. Introduction

Determining what the regulation of social media content is in Latin America and how it is enforced is a constant challenge. The array of relevant issues, and their cardinal importance in the region - from privacy, freedom of expression, liabilities, disinformation, right to erasure, and copyright - is compelling, yet staggering in its expanse and substance.

With the noteworthy exception of Brazil's Marco Civil of the Internet passed in 2014, a pioneer law regarding the protection of fundamental rights and principles on the Internet, Latin American countries do not

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

have specific regulations for social networks,[1] yet it could be said that there is a perception among policymakers in the region that the online space remains under-regulated.[2] The lack of understanding as to how content moderation works in Latin America, among other reasons, has led to increasing calls from both governments and civil society, for regulation against major platforms.[3]

Nevertheless, with over 410 million users of social media[4] in Latin America, legal controversies arise and courts have relied on classic civil law rules to decide on matters regarding defamation lawsuits and invasion of privacy.

This review was not meant to be, and clearly could not be, exhaustive. The first part of this paper aims to give an overview of selected developments regarding the regulation of social media content in some countries of Latin America, while mentioning the type of traditional norms that are applied to the digital environment. Then, the analysis will narrow in on copyright in Latin America, and how statutory exceptions and limitations regulate copyrighted content on social media. The analysis seeks to address whether the present copyright regulatory framework is suitable to the digital environment. Finally, content moderation on social networks in Latin America shall be briefly discussed, with specific attention to the new notice and takedown procedures introduced in the 2020 Mexican Federal Copyright Law.

Chapter 2. The Latin American landscape of social media governance: A brief overview of regulations in México, Colombia, Argentina and Chile

Latin America is a region with deep economic and social inequalities. In this context, access to information and communication technologies serves as a possible tool to attain social inclusion and achieve the region's

1 Andrés Calderón, "Moderate Globally Impact Locally: Content Moderation in Social Media in Latin America: A promise to consumers", *Yale Law School Information Society Project*, October 27, 2020, https://law.yale.edu/isp/initiatives/wikimedia-initiative-intermediaries-and-information/wiii-blog/moderate-globally-impact-locally-content-moderation-social-media-latin-america-promise-consumers.

2 Augustina del Campo, "Social media in Latin America: Caught between a rock and a hard place", *Global Voices*, September 17, 2020, https://globalvoices.org/2020/09/17/social-media-in-latin-america-caught-between-a-rock-and-a-hard-place/.

3 *N.N.*, "Social media in Latin America."

4 Simon Kemp, "Digital 2020 July Global Statshot", *we are social*. July 21, 2020, https://wearesocial.com/blog/2020/07/digital-use-around-the-world-in-july-2020.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

development goals. However, though there are important differences from country to country, according to data from the Development Bank of Latin America (CAF), 38% of the region's population does not have internet access.[5] Thus, the inequalities of the region are also present in its digital ecosystem.

Adding to that, "*Internet use in much of Latin American households is limited to communication tools and social networks*."[6] This falls in line with the report that Facebook and Whatsapp are the principal news sources in Argentina, Brazil, Chile, and Mexico.[7] In 2019, Latin American users top the global rank of most time spent in social media, with an average of 212 minutes a day.[8]

Evidently, a comprehensive legal framework to regulate social media is due. However most Latin American countries have not included provisions that directly regulate social media platforms into their domestic legal systems.

Before providing a general layout of the different rules that inform the different domestic legal systems of Latin American countries in this subject, it is important to make an approximation of how social networks are defined in Latin America. Currently, there is a bill in Mexico, proposed by Senator Ricardo Monreal Avila, that defines "social media services" as the "internet services that have the main function of sharing content published by users, in the form of texts, data, voice notes, images, video, music, sounds or a combination of these, with the purpose of informing, entertaining or educating audiences."[9]

---

5 "Transformación digital para la América Latina del S. XXI", Banco de Desarrollo de América Latina, accessed June 22, 2021, https://www.caf.com/es/conocimiento/visiones/2020/02/transformacion-digital-para-la-america-latina-del-s21/.

6 Carlos I. Ortuño, "COVID-19 and digital inclusion in Latin America and the Caribbean: A connectivity and access problem", *SELA. Latin American And Carribean Economic System*, June 04, 2020, http://www.sela.org/en/press/articles/a/64488/covid-19-digital-inclusion-in-latin-america-and-the-caribbean.

7 Observacom, "Redes sociales son las principales vías de acceso a la información en América Latina", *Observatorio Latinoamericano de Regulación de Medios y Convergencia*, June 22, 2020, https://www.observacom.org/redes-sociales-son-las-principales-vias-de-acceso-a-la-informacion-en-america-latina/

8 Fernando Duarte, "Los países en los que la gente pasa más tiempo en las redes sociales (y los líderes en América Latina)", *BBC News Mundo*, September 9, 2019, https://www.bbc.com/mundo/noticias-49634612.

9 Iniciativa con proyecto de Decreto por el que se REFORMAN y ADICIONAN diversas disposiciones de la Ley Federal de Telecomunicaciones y Radiodifusión, Senador Dr. Ricardo Monreal Ávila, Morena Political Party, Mexico.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

It is important to point out that social networks bring together a diversity of users, from the private and public sector, and these actors communicate a variety of social, political and/or commercial interests. Thus, legislation drafted to intervene these spaces must be in balance with this breadth of protagonists and functionalities.

This section shall point out advances made in the regulatory frameworks of México, Colombia, Argentina and Chile regarding some of the major issues which arise in relation to social media platforms. These issues are: (a) Liability of Internet Intermediaries, (b) Protection of personal data and privacy, (c) Right to honor and reputation, and (d) Legal trends related to hate speech and influencers. Regulation of copyrighted content in social media will be addressed in Section 3.

Chapter 2.a. Intermediary Liability

As far as online intermediaries are concerned, liability standards have been set in the region, though none comparable to Section 230 of the U.S. Communications Decency Act. Before going over the laws and rulings that have informed the intermediary liability systems in some Latin American countries, it is convenient to briefly define the terms "strict liability" and "subjective liability". Strict liability arises from the damages caused to a third party in the exercise of an activity considered to be risky, regardless of whether the conduct was carried out negligently or with harmful intent. Instead, when subjective liability rules are applied, the focus is on the accused's intention, knowledge or awareness to determine, on a case-by-case basis, if the acts were originated in ignorance or negligence. Thus, a strict liability regime entails a greater demand on the conduct of internet intermediaries; even when unaware of the commission of the reproached behaviors, they shall, nonetheless, be responsible by virtue of their activities as providers of internet services.

a) MEXICO

In Mexico, there is not a specific law regulating intermediary liability. However, last year's modification to the Federal Copyright Law, proposes a limitation to their responsibility in relation with the circulation of content protected by authors' rights. This Law exonerates the "online service providers" from liability for the infringement of intellectual property

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

rights, under the condition that they proceed to withdraw this content once they have been notified of the existence of the infringing content by the owner of the protected works or an authorized representant or ordered to remove it by a competent authority[10].

b) COLOMBIA

Colombia's current legal framework provides no specific law concerning exclusively with intermediary liability. Nonetheless, Law 679 of 2.001 set liability standards for intermediaries, in relation to the exhibition of child pornography through global networks of communication. Following are the main points of the aforementioned law:

- The national government must promote the adoption of autoregulation systems[11] for persons and enterprises located in Colombia, whose main commercial activity is the trade of goods and services using global networks of information[12].
- The law prohibits providers, administrators, and users of global networks of information, from: (i) hosting pornographic material of minors on their own websites; (ii) hosting explicit material in which the participants could be believed to be minors and, (iii) hosting links to websites that distribute such material[13].
- Failure of an intermediary to denounce, contend and take down such content or using networks of communication in the manner prohibited by the Law[14] will generate fines (up to 100 minimum legal wages) and the cancelation or suspension of infringing websites[15], plus prison charges.

Given the gaps in Colombia's regulatory framework, courts have filled in the blanks (a repeated pattern in the region). Following are the points

10 Article 114, Octies, Ley Federal Del Derecho De Autor, Congreso De Los Estados Unidos Mexicanos, 2020.
11 Article 6, Law 679, 2001, Congress Republic of Colombia.
12 Article 3, Law 679, 2001, Congress Republic of Colombia.
13 Article 7, Law 679, 2001, Congress Republic of Colombia.
14 Article o 8, Law 679, 2001, Congress Republic of Colombia.
15 Article 10, Law 679, 2001, Congress Republic of Colombia.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of analysis observed by Colombia's Constitutional Court in two rulings relevant to the subject at hand, cases T-277/15[16] and SU420/19[17]:

- In case T-277/15, the Court refers to the Joint Declaration on Freedom of Expression and the Internet, adhering to its principle of "mere transmission", meaning that no person or enterprise who exclusively provides technical Internet services shall be held responsible for the content created by third parties. After recognizing the role of the Intermediaries as catalysts for the free traffic of ideas, the Court concludes that holding the Intermediaries responsible for the illicit doings in user-generated content would affect the communicative exercise in the digital space "because it would give them the power to regulate the flow of information in the Web".
- In case SU420/19, the Court confirms its position regarding the absence of liability of intermediaries for content created by platform users. However, it considers that during the legal proceedings carried out in defense of the rights of honor and reputation, if the user who has uttered the reproved expressions is absent, the intermediary must participate as a third party as, eventually, it will have to follow the judges content removal order.

c) ARGENTINA

Once again, given the lack of specific legislation, when addressing issues related to intermediary liability, Argentine judges have drawn upon the general principles of civil and criminal liability.

One of the most relevant cases in Argentina's jurisprudence is that of "*Rodriguez Maria Belen c/ Google. Inc*".[18] María Belén Rodríguez, an Argentine entertainment personality, brought a civil case against Google, seeking damages as a result of having her name and images associated to websites with explicit content. The first ruling found Google responsible for having infringed the rights of the plaintiff, awarding the latter compensation for

16 Sentencia de tutela Radicado No. T-277/15, 2015, Corte Constitucional, Colombia. https://www.corteconstitucional.gov.co/relatoria/2015/t-277-15.htm

17 Sentencia de Unificación Radicado No. SU420/19, 2019, Corte Constitucional, Colombia. https://www.corteconstitucional.gov.co/relatoria/2019/SU420-19.htm

18 Rodríguez, María Belén c. Google Inc. s. daños y perjuicios, 2014, Corte Suprema de Justicia de la Nación, Argentina. https://cdh.defensoria.org.ar/normativa/rodriguez-maria-belen-c-google-inc-s-danos-y-perjuicios/

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

damages and ordering the removal of the links. During the appeal, the previous judgment was partially annulled, in this instance, the National Appeals Chamber of Argentina found that there was no evidence that the defendant refused to take down the offensive content after being notified of its existence.

Finally, the case reached the Supreme Court, which displayed a comparative analysis of legal precedents from several countries from which it generated the following conclusions:

- When applying civil rules to cases where a fundamental right is at risk, these laws are to be interpreted in the way that better adapts to the National Constitution.
- If intermediaries do not have a general duty to monitor the content, they cannot be held responsible for the content generated by users, thus rejecting a strict liability standard on the basis of the threat it would pose to free expression rights.
- When the conducts of intermediaries are examined in a judicial procedure, the judge must use the rules of subjective liability. Intermediaries shall be accountable for the damages caused to third parties only when they had “effective knowledge” of the commission of the illicit behavior and did not respond accordingly.
- When the conduct is not clearly transgressive to the rights of honor and image of the user who is denouncing it, that is when it is not a case of "gross and manifest harm”, the court held that search engines could not be liable for unlawful content upon notification unless a public authority had adjudicated the material as unlawful.
- Google Image thumbnails were considered links and not Google’s own content.

A recent case involving former president, and current vice-president Cristina Fernandez de Kirchner, also received much news coverage. The case to be decided is the lawsuit filed by the vice-president against Google for defamation and tarnishing of her image, name, and honor, based on the fact that when entering her name into Google's search engine, instead of mentioning her position in the government, there appeared an epithet reading "Thief of the Argentine Nation". The result did not refer to any third-party website but was under the sole responsibility of Google.

As a preventive measure, the plaintiff filed a petition with the court demanding that Google preserve the data related to her name for inspection in the proceedings. This request was granted by both the Civil Judge and the appellate court after the defendant appealed the first decision. Google presented a complaint to Argentina's Supreme Court against the appellate

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

court's ruling; however, this was overruled in March 2021.[19] This case may set an important precedent as to the way evidence is handled in cases relating to acts of defamation on the internet.

d) CHILE

In Chile, the intellectual property regime was modified by Law 20,435, which includes a chapter on the liability of intermediaries. Intermediaries are not forced to monitor the content generated by users, thus releasing them from liability in this regard, on the condition, the intermediary abides by the rules of article 85Ñ of said law.

The mentioned article applies to providers of search and linking services and providers that, at the request of the user, host data in their systems, stating that these subjects are freed of liability if they:

- Do not have effective knowledge of the illicit data;
- Do not profit from the infringing conduct;
- Appoint an agent to receive the judicial notices of the existence of the illicit content;
- Expeditiously remove the material considered to be infringing.[20]

On its part article 85U of the law in question, devises a "notice to notice to notice"[21] system, meaning that when an intermediary has received the notices of the allegedly infringing content, they must inform the creator or owner of said content, briefing them on the facts of the notice.

Chapter 2.b. Personal Data Protection

Regulating social media policies regarding personal data is essential, being that users expose their personal and domestic life, revealing information that can be easily exploited against their interests, thereby undermining their rights to privacy. Latin American countries have been mindful of this

19 Patricia Blanco, "La Corte Suprema falló a favor de Christian Kirchner en la causa que inició contra Google", *infobae*, March 19, 2021, https://www.infobae.com/politica/2021/03/19/la-corte-suprema-fallo-a-favor-de-cristina-kirchner-en-la-causa-que-inicio-contra-google/.

20 85Ñ, Ley 20.435, 2010, Congreso Nacional, Chile.

21 85U, Ley 20.435, 2010, Congreso Nacional, Chile.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

need and Data Protection Laws in the region have proliferated through the last decade.

a) MEXICO

The Mexican Law on Protection of Personal Data in the Private Sector was passed in 2010, and the General Law on the Protection of Personal Data Possessed by Obligated Subjects was passed in 2017.[22]

b) COLOMBIA

In 2012, the Colombian congress passed Statutory Law 1581, which regulates the treatment and protection of personal data collected in the Colombian territory and the data collected elsewhere, by a person who is obligated to comply with Colombian law by virtue of international treaties[23].

c) ARGENTINA AND CHILE

Argentina and Chile have the oldest laws in the continent and currently, both are pending updating. For Argentina, this is Law 25.326 of 2000 and for Chile, Law 19.628 of 1999.[24]

Chapter 2.c. Rights to Honor and Reputation

Defamation is a criminal offense in the penal codes of certain Latin American countries.On the international stage, the Inter-American Human Rights Court has established that the need to repair a defamed person's right is not a justification to restrict freedom of expression *prima facie.* This

22 Paulina Bojalil, "Despuntan las reformas en materia de protección de datos en América Latina", *ABIERTA al público* (blog), February 12, 2019. https://blogs.iadb.org/conocimiento-abierto/es/proteccion-de-datos-gdpr-america-latina/.

23 Artículo 2, *Ley Estatutaria 1581, 2012,* Congreso de República de Colombia.

24 Valentina Hernández Bauzá, *Sucesos regulatorios en materias de privacidad e internet en Latinoamérica* (Derechos Digitales América Latina, 2020), https://www.derechosdigitales.org/wp-content/uploads/tendencias-privacidad-latam.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

means that to define which right should prevail between the right to honor or the right to free speech, the court must take into account the specifics of each case. This exercise implies engaging in a proportionality judgment, in which prior censorship must be prevented, this was expressed by the mentioned tribunal in the Kimel case, which will be further explained later.

a) MEXICO

Mexico is an example of decriminalization of defamatory offenses within the Latin American context. Since the reform of the Federal Criminal Law in 2007, insult, slander, and defamation are considered illegal acts, instead of criminal offenses, generating liability via civil law.[25]

b) COLOMBIA

In Colombia's legal system, defamation is still penalized by criminal law. As of July 2020, a new bill,[26] which regulates defamation and other related offenses against honor, reputation, privacy and image, gives the victims of such acts the prerogative of filing for reparation through in civil courts as well.

c) ARGENTINA

Argentina's Law № 26.551 allows all forms of expression when they concern matters of public interest[27]. This legislation was prompted by the case Kimel vs Argentina,[28] a relevant precedent for the entire human rights system in Latin America, and a milestone in defamation cases in Argentina,

25 La Relatoría Especial Para La Libertad De Expresión. (2013, noviembre 11). Comunicado de Prensa R 85/13, https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=934&lID=2

26 *L. 48, 2020*, Gaceta No. 593 del Congreso de Colombia.

27 Artículo 5, *Ley 26.551, 2009*. Congreso de la Nación Argentina.

28 *Kimel V. Argentina, 2008. Sentencia de Fondo, Reparaciones y Costas*. Corte Interamericana de Derechos Humanos.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

which began progressively removing the prison penalties for slander and defamation from its penal code[29].

Kimel, a journalist, writer, and investigative historian published "The San Patricio Massacre ("La Masacre de San Patricio), a book on his investigation of the murder of five people from a religious order during Argentina's military dictatorship, criticizing how the authorities handled the judicial procedures that followed. In 1991, the State brought criminal proceedings against Kimel for defamation of a judge criticized in the book. Upon the conclusion of the criminal proceedings, he was convicted of libel and sentenced to one-year imprisonment and payment of a large sum in damages.

The Inter-American Court of Human Rights found that the State violated the American Convention on Human Rights. The importance of this decision lies in its very precise restatement establishing that speech regarding public officials acting in the course of their duties, and the public interest, enjoys a greater degree of protection. The ruling provides a proportionality analysis between the judge's right to reputation and Kimel's right to free speech, with a three-part test regarding the degree to which each right was affected, the importance of each right, and the existing justifications to restrict one right and satisfy the other. It also emphasized the need to scrutinize very carefully when using criminal law to restrict freedom of expression.

d) CHILE

The victim of defamation and slander may file for reparations in civil or criminal courts, as these offenses are also part of the Chilean penal code, punishable by prison sentences and fines. However, according to Law 19.733, personal opinions relating to political, literary, historic, artistic, scientific, technical and sports subjects" expressed in "social communication media" are not considered slander, as long as there is not intent to insult.[30]

---

29 La Relatoría Especial Para La Libertad De Expresión. (2013, noviembre 11). Comunicado de Prensa R 85/13, https://www.oas.org/es/cidh/expresion/showarticle.asp?artID=934&lID=2.

30 Article 29, Law 19.733, Congreso de Chile.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 2.d. Other Regulatory Trends

a) Hate speech

The exercise of free speech is the cornerstone of social and political interactions on the internet. Due to this dynamic, trying to regulate hate speech may cause unintended negative effects on freedom of expression.

At the regional level, the American Convention on Human Rights (The Pact of San José), protects freedom of expression in its article 13,[31] refusing to consider hateful national, racial or religious speech, that incites to commit a violent act, as a legitimate manifestation of this right. In the Latin American region, with the exception of Bolivia and Venezuela[32], there are no domestic laws that directly prohibit or regulate hate speech in the digital sphere.

According to Human Rights Watch[33], the political climate around freedom of expression in Mexico is going through a worrying situation due to the previously mentioned draft bill of legislation, authored by Senator Ricardo Monreal Avila, that seeks to reform the federal law on telecommunications and broadcasting. One of the most criticized points of this bill is the ample faculties given to the Federal Telecommunications Institute (IFT). In order to operate in Mexico, digital platforms must present their terms and conditions before the IFT, agreeing to limit the dissemination of hate speech. The bill, however, does not define what should be considered as a hateful message. Said bill grants the IFT the ultimate decision powers in the challenges presented by users regarding the decisions to cancel accounts and remove content made by the platforms.

Argentine law on hate crimes is based mainly on the Anti-Discrimination Act, Law No. 23592 of 1988. Currently in Argentina there is a bill[34] referring to this subject, in which hate speech is defined as the messages or expressions that "intimidate, discriminate or incite hatred or violence against based on motives of race, religion, nationality, gender, sexual orien-

---

31 Convencion Americana Sobre Derechos Humanos (Pacto De San José), 1969. Organización de Estados Americanos. Artículo 13.

32 Rodrigo Vargas Acosta, *Sucesos regulatorios en materias de libertad de expresión e internet en Latinoamérica* (Derechos Digital América Latina, 2020), https://www.derechosdigitales.org/wp-content/uploads/tendencias-regulacion-digitales.pdf.

33 Human Rights Watch, "Mexico: Online Free Speech at Risk," April 14, 2021, https://www.hrw.org/news/2021/04/14/mexico-online-free-speech-risk

34 Proyecto de Ley 848/20. Senado De La Nación. Argentina. https://www.senado.gob.ar/parlamentario/comisiones/verExp/848.20/S/PL

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tation, disability, among others". The same law urges platforms to follow a procedure for receiving complaints in which the denounced publications are temporarily withdrawn.

b) The legal regulation of influencers

Given that "influencers", protagonists of social media, tend to play a role in consumer choices and behaviors, many countries have tried to regulate the exercise of this activity.

Mexico[35] and Colombia have not dictated laws to control influencers. However, in Colombia[36], the Superintendency of Industry and Commerce, issued a guide of best advertising practices for influencers, seeking to create transparency between them, their sponsors, and consumers. Argentina[37], on its part, has a new bill proposing to regulate influencers as "digital advertisers", with mandatory disclosure of their sponsoring contracts with sponsors, and consumer protection regulations as to the disclosure of information and contraindications of the advertised products. Finally, although Chile does not have a law directed to influencers, they are regulated through tax laws.[38]

---

35 Luis Mario Lemus Rivero, "Influencers, aspectos legales a considerar", *Foro Jurídico*, October 8, 2020, https://forojuridico.mx/influencers-aspectos-legales-a-considerar/.

36 Superintendencia de Industria y Comercio de Colombia. (2020, October 1). *Superindustria expide "Guía de buenas prácticas en la publicidad a través de influenciadores."* Sic.Gov.Co. https://www.sic.gov.co/slider/superindustria-expide-"gu%C3%ADa-de-buenas-prácticas-en-la-publicidad-través-de-influenciadores

37 Paula Fernandes Pfizenmaier, "Influencers' Regulation In Argentina: When No Law Is Better Than A Bad Law", Mondaq, 16 July, 2020, https://www.mondaq.com/unitedstates/socialmedia/965950/influencers39-regulation-in-argentina-when-no-law-is-better-than-a-bad-law.

38 Marcela Gómez, Matías Bobadilla, "Influencers deben pagar impuestos por ganancias en redes sociales", pauta.cl, March 28, 2021, https://www.pauta.cl/economia/influencers-tributacion-impuestos-chile-redes-sociales.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3. Intellectual Property Laws in the context of Social Media Platforms in Latin America: Regulating Copyrighted Content in Latin America

As discussed, the regulation of social media platforms is, as of yet, a matter of unfinished public policy in Latin America. Although there are a few *ad hoc* norms in force in some countries, the regulatory framework for social media platforms is still emerging, with the region struggling to apply current regulations to channel the legal questions arising from interactions on social networks.[39]

As far as the regulation of copyright in Latin America is concerned, each country has its particular legislations and guidelines to prevent the unauthorized use of works. Most of the provisions have been harmonised with the copyright protection criteria established in international treaties, as shall be addressed herein.

Furthermore, all Latin American countries provide for limitations and exceptions within their copyright frameworks to allow certain unlicensed uses of copyrighted materials. To ensure that the legitimate interests of rights holders are respected, laws typically include limitations restricting such content from being used for commercial purposes or from interfering in the original work's market.

Two things to keep in mind: First, limitations and exceptions do not waive the author's moral rights (such as the right of authorship, the right of integrity of work and the right of divulgation). Second, continuous technological progress keeps creating new possibilities for uses of copyrighted works, yet the same legislations are applied to regulate the new uses. Very few countries have adapted their copyright laws specifically to the digital environment.

Chapter 3.a. Overview of the Copyright System in Latin America and its Exceptions and Limitations

Although the term "copyright" is often used in reference to authors' rights in Latin America, it is important to point out that Latin American countries follow the model of the continental legal system, rooted particularly

39 Moisés Sánchez, *Informe sobre control estatal de redes sociales* (Alianza Regional por la Libre Expresión e Información, 2016), http://www.alianzaregional.net/wp-content/uploads/Informe-Arti%CC%81culo-XIII-2016-GF-SR-DM.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

in French law. As such, the essence is that, in addition to the economic rights that the law grants to authors of literary, artistic or scientific works, the *droit d'auteur* legal system grants authors moral rights, related to the "paternity", integrity and disclosure of the works. These two sets of rights -moral and economic- are characteristic of the "continental" vision (*droit d'auteur*), in contrast to the Anglo-Saxon vision (*copyright*), where the moral component has not been incorporated until rather recently, and perhaps with little enthusiasm.[40]

As mentioned, authors' rights in Latin America have been harmonised, thanks to the international treaties and multilateral conventions. The Berne Convention is the oldest of these conventions, dating from 1886.[41] Therein, the region adheres to the principle of automatic protection of works, which establishes that works will be protected from their creation[42], without the need for registration or any formality. The only condition is that the work is captured in a fixed medium and has a minimum of originality.

a) Background: Berne and the Three-Step Rule

The main concern of authors when the Berne Convention was adopted in the late 19th century,[43] was to avoid the improper appropriation or reproduction of their works by third parties. In the absence of a harmonized or uniform international system for the recognition of copyright, plagiarism

40 J. Carlos Fernández-Molina and Eduardo Peis, "The moral rights of authors in the age of digital information", *Journal of the American Society for Information Science and Technology* 52, issue 2, (2001): 109-117, https://doi.org/10.1002/1097-4571(2000)9999:9999%3C::AID-ASI1060%3E3.0.CO;2-B.

41 In addition to this, there are other international instruments in those that have established standards for the protection of copyright and related rights and that have contributed to the consolidation of a uniform international system, such as the Rome Convention 1961, the Phonograms Convention of 1971, the Convention on the distribution of signals satellite channels of 1974, the World Intellectual Property Organization (WIPO) Copyright Treaty of 1996 (WCT) and the WIPO Performances and Phonograms Treaty of 1996 (WPPT).

42 See article 5, paragraph 2 of the Berne Convention.

43 "Reseña del Convenio de Berna para la Protección de las Obras Literarias y Artísticas", Organización Mundial de la Propiedad Intelectual, accessed June 22, 2021, https://www.wipo.int/treaties/es/ip/berne/summary_berne.html#:~:text=El%20Convenio%20de%20Berna%20trata,que%20quieran%20valerse%20de%20ellas.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

or unauthorized use of works was a constant risk.[44] Berne gave authors, musicians, poets, painters, among others, the means to control who used their works, how and under what conditions. At the same time, it served to establish the minimum standards of international protection for literary and artistic works.

When setting these international rules for the recognition and protection of copyright, it was also made clear that countries could limit the protection of a work or allow the exceptional use of literary or artistic works without the consent of their author. These provisions are known as **copyright limitations and exceptions.**[45]

Among the most common limitations set by the Berne Convention are the following:

(i) the limitation on the protection of official texts[46],
(ii) the limitation on protection of daily news and press information[47],
(iii) the limit on protection of political speeches and those in legal proceedings [48].

On the other hand, in relation to exceptions, we have:

(i) the right to use citations or the right to quote (in educational and other particular circumstances)[49];
(ii) the use for teaching purposes[50];
(iii) the use of articles in newspapers and periodical collections[51];

---

44 Prior to the adoption of the Berne Convention, there were national laws historical relevance that recognized and protected copyright, such as, the "Statute of Anne", the original title of which is *" An act for the encouragement of learning, by vesting the copies for printed books in the authors or purchasers of such copies, during times there in mentioned*", passed in 1709. Other countries joined this wave of copyright protection; in 1790 the United States enacted its first copyright law; In 1791 and 1793 France approved Decrees 13 and 19 on the protection of literary works, and finally, in Spain, 1847 (Christian Schmitz Vaccaro, "Evolución de la regulación internacional de la propiedad intelectual," (Concepción, Chile: Universidad Católica de la Santísima Concepción, 2013) https://revistas.uexternado.edu.co/index.php/propin/article/view/3580/3661).

45 Limitations and exceptions are based on Article 9 (2) of the Berne Convention.

46 See article 2.4 Berne Convention.

47 See article 2.8 Berne Convention.

48 See article 2 bis 1 Berne Convention.

49 See article 10.1 Berne Convention.

50 See article 10.2 Berne Convention..

51 See article 10bis 1 Berne Convention..

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

(iv) the use of works in information relating to current events;[52] and
(v) the use of information from conferences, speeches and other similar events.[53]

The interpretation criteria for these limitations and exceptions are based on what is commonly known as the **three-step rule** introduced in article 9.2 of the Berne Convention: "It shall be a matter for legislation in the countries of the Union to permit the reproduction of such works in certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author."

These three factors, which has become central in international conventions relating to copyright, is the basis of interpretation establishing the limits of permissible uses of works of third parties. Though the Berne Convention established this rule referring only to the right of reproduction, through the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS)[54], this rule was extended to any of the exclusive rights related to copyright.[55]

Exceptions and limitations are an important part of any copyright system, allowing creators to access and continue creating, using the knowledge generated by others. Without exceptions and limitations, the authors' rights system would not be able to achieve at least one of its fundamental purposes, which is to stimulate creation and promote innovation for the benefit of humanity.[56] The background described shows that there is a broad regulatory system related to the protection of copyright, not only in Latin America, but in the world, that serves as a fundamental framework for the unauthorized use of content on the internet.

---

52 See article 10bis 2 Berne Convention.
53 See article 2bis.2 Berne Convention..
54 See article 13 TRIPS.
55 "La regla de los tres pasos," Electronic Frontier Foundation, accessed June 22, 2021, https://www.eff.org/sites/default/files/filenode/tpp_3pasos.pdf.
56 Gwen Hinze, "Hacer que el conocimiento sea accesible a través de las fronteras: excepciones mínimas obligatorias de derechos de autor internacionales para la educación," *Electronic Frontier Foundation*, October 30, 2008, https://www.eff.org/wp/making-knowledge-accessible-across-borders-case-ma.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

b) Authors' rights in copyrighted content from the user's standpoint: Are everyday practices of social media content-sharing illegal in Latin America? Is copyright affecting essential tasks on the internet and limiting social practices of democracy, such as access to culture and information?

As is commonly known, the use of the internet in general, and social networks in particular, has generated practices that are based on the creation of content, or the reproduction, re-use or transformation of third -party content. Sometimes these dynamics imply obtaining a direct or indirect profit. Some examples are the creation of memes, sharing GIFs, those unavoidable loops of animation, or retweeting the status of another user, uploading stories to social networks such as Instagram or Twitter in which third-party songs are incorporated and the streaming of e-sport games on platforms such as Twitch or YouTube.

Thus, while a user of social media platforms is potentially the author of copyright--protected content, he or she, in turn, is a possible infringer of copyrights that belong to third parties.

In this context, one considers whether these dynamics carried out on the internet are adequately regulated by the same principles that apply to activities carried out in our analog, offline environment. Are our Latin American *droit d'auteur* laws outdated?

Most scholars in Latin America agree that local copyright laws can be applied to content in social networks, although not specifically mentioned in the norms, when such content fulfills copyrightability standards, such as originality.

To cite an example, several scholars have maintained that though Law 11,723, the main legal provision on authors' rights in Argentina, is almost one hundred years old, its broad wording, complemented with the international treaties to which Argentina is a party, allow it to achieve a comprehensive copyright protection applicable in the digital age.[57]

Other authors like Busaniche, consider that Law 11,723 regulates copyright through a highly restrictive model that, consequently, curtails circulation and makes common practices of socialization of culture illegal, affecting essential tasks and social practices of democracy, such as access to culture and information, the work of teachers and students and their

57 Carlos A. Villalba and Delia Lipszyc, *El derecho de autor en la Argentina* (Ciudad Autónoma de Buenos Aires, Argentina: La Ley 2009); Ariel Alberto Neuman, "Derechos de autor y era digital", *El Cronista*, June 6, 2018, https://www.cronista.com/legales/Derechos-de-autor-y-era-digital-20180606-0011.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

access to educational materials, and the work of libraries. She argues that in today's digital environments, this system, in which the conditions that gave meaning to copyright are completely modified, needs a structural transformation.[58]

It is important to clarify that using or reproducing third-party content is not always illegal. As mentioned, limitations and exceptions are a fundamental part of the copyright system, and allow the use and disclosure of content, provided the use is deemed permissible under the three-step rule.

Yet, in order to permit the aforementioned socialization of culture, when thinking about copyright rules in the digital era, we must pay as much attention to addressing limitations and exceptions, as to enhancing copyright protection. Are the existing limitations and exceptions to copyright in Latin America suitable to permit the regular interactions taking place on today's social media? More importantly, what about the access to culture and circulation of information in the digital environment? It does not seem clear that many of these acts will be permissible under the present limitations and exceptions system, which has a very narrow and limited scope.

As user-generated content flourishes, Elkin-Koren argues that users play a critical role in copyright law, and makes a fascinating case for the "user rights approach".[59] Observing the user's interests only through the spectacles of limitations and exceptions, is far too narrow, and overlooks the vital role users play in the copyright system. Elkin-Koren makes the thought-provoking suggestion that permissible uses under copyright law should be articulated and treated as rights.

---

58 Beatriz Busaniche, "Argentina Copyleft. La crisis del modelo de derechos de autor y las prácticas para democratizar la cultura", Fundación Via Libre, September 10, 2010, https://www.vialibre.org.ar/argentina-copyleft-la-crisis-del-modelo-de-derechos-de-autor-y-las-practicas-para-democratizar-la-cultura/.

59 Niva Elkin-Koren, "Copyright in a Digital Ecosystem: A User-Rights Approach", Forthcoming in RUTH OKEDIJI, COPYRIGHT IN AN AGE OF LIMITATIONS AND EXCEPTIONS, July 28, 2015, https://ssrn.com/abstract=2637027.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

c) Should exceptions and limitations in Latin America be reformed in order to adapt to the common practices in the digital environment? Is there a possibility of incorporating broader criteria, such as the Copyright Fair Use factors?

Questions arise as to how to adapt our Latin American authors' rights frameworks to better suit the digital environment.

Limitations to authors' rights in the laws of Latin American countries are lists of very specifically defined, and narrowly constructed exceptions to the exclusivity granted to authors by law. These are exhaustive, closed lists; if a use of copyrighted content does not fall into one of these very specific categories, it will be considered an infringement. In contrast, the four statutory factors of fair use in U.S, copyright law, and the fifth factor of "transformative use" introduced by courts, provide more flexible criteria that can be used by courts to decide whether a specific use is permissible on a case-by-case basis.

The idea of adopting the Fair Use interpretation criteria has not, as of yet, been a viable alternative for Latin American countries. There is a certain fear of the unpredictability associated with the application of the fair use criteria, and the traditional rule in countries, including those of the European and Latin America, is that copyright limitations and exceptions must be narrowly defined.

Nevertheless, it is surprising that the international three-step test, which has been incorporated into Latin American legislations through the adoption of the Berne Convention as well as other treaties, is in fact rooted in the Anglo-American copyright tradition.[60]

It has often been considered that the three-step test in international copyright law is is an obstacle to the adoption of more flexible criteria at the national level, yet, Geiger, Gervais & Senftleben have considered that the test was actually intended to serve as a more flexible balancing tool, offering national policy makers the possibility to adopt a flexible system of limitations and exceptions. [61]

---

60 Christophe Geiger, Daniel J. Gervais and Martin Senftleben, "The Three-Step-Test Revisited: How to Use the Test's Flexibility in National Copyright Law", (November 18, 2013) *American University International Law Review*, Vol. 29, No. 3 (2014), pp. 581-626, https://ssrn.com/abstract=2356619 or http://dx.doi.org/10.2139/ssrn.2356619

61 Christophe Geiger, Daniel J. Gervais and Martin Senftleben. "The Three-Step-Test Revisited: How to Use the Test's Flexibility in National Copyright Law", *American University International Law Review* 29 no. 3 (2014):581-626.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Among the treaties signed by many Latin American nations are the WIPO Copyright Treaty (WCT) and the WIPO Performances and Phonograms Treaty (WPPT). In order to consider the possibility of Latin American countries to reform their legislations regarding limitations and exceptions, it is particularly noteworthy to highlight certain provisions of the WIPO Copyright Treaty and its "Agreed Statements". The details of these treaties is beyond the scope of this overview, but suffice it to mention that some provisions therein, indicate quite specifically that the function of the three-step test is to serve as a flexible framework for the adoption of limitations and exceptions at the national level.

Article 10(1) of the WCT is a direct application of the three-step test to WCT rights:

> "Contracting Parties may, in their national legislation, provide for limitations of or exceptions to the **rights granted to authors of literary and artistic works under this Treaty** in **certain special cases** that do **not conflict with a normal exploitation** of the work and **do not unreasonably prejudice the legitimate interests of the author**."

The Agreed Statement Concerning Article 10 WCT states: [62]

> "It is understood that the provisions of Article 10 permit Contracting Parties to carry forward and appropriately extent into the digital environment limitations and exceptions in their national laws which have been considered acceptable under the Berne Convention. Similarly, these provisions should be understood to permit Contracting Parties to devise new exceptions and limitations that are appropriate in the digital network environment. It is also understood that Article 10(2) neither reduces nor extends the scope of applicability of the limitations and exceptions permitted by the Berne Convention."

This statement is between all parties to the WCT, prepared at the same time as the Convention, which makes it an important context for the interpretation of article 10.

62 Sam Ricketson, "WIPO Study on Limitations and Exceptions of Copyright and Related Rights in the Digital Environment", March 3, 2008.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3.b. Notice and Take-Down: Content Filtering in Latin America

In order to protect copyright online, notice and take-down procedures and content filtering systems work in tandem, allowing the copyright owner to request the removal of the infringing content, while the filtering systems serve to ensure that the offending content does not re-upload on the internet. Systems such as these are provided for in the U.S. DMCA,[63] implemented in 1998 and the Directive of the European Union[64] of 2000.

In Latin America, many have argued that in practice these systems have been used to generate acts of censorship of opponents, as tools to control reputation and public image, as well as to manipulate or modify the publicly accessible information that is hosted on the internet.[65] All of the above, based on the filing of claims for alleged invasions of copyright online.[66] For example, in Ecuador, some years back, copyright laws were used to remove content criticizing the government.[67]

The creation of these systems has close ties to the creation of the WIPO internet treaties (WCT) and (WPPT) in 1996. As mentioned, most countries in Latin America are party to the WIPO treaties (WCT and WPPT), which establish the obligation to have effective measures to avoid the execution of infringing actions on the internet.

In Chile these mechanisms were established through Law 20,435 of 2010.[68] In Costa Rica through Regulation 36880-COMEX-JP of 2011 and in Paraguay through Law 4,868 of Electronic Commerce of 2013[69] and more recently in Mexico with the reforms to the Federal Copyright Law of 2020.

---

63 See https://www.law.cornell.edu/uscode/text/17/512

64 «Directiva sobre el comercio electrónico 2000/31/CE.» 2000. https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32000L0031&from=ET (accessed: 2021).

65 Reseña del Tratado de la OMPI sobre Derecho de Autor (WCT) (1996)," Organización Mundial de la Propiedad Intelectual", accessed June 22, 2021, https://www.wipo.int/treaties/es/ip/wct/summary_wct.html.

66 Alejandro Aréchiga Morales, "Sistema de notificación y retirada en México: los derechos en juego", Centro de Política Digital para América Latina, 2021 .

67 Derechos Digitales, "The various paths of Internet censorship in Latin America", IFEX, November 14, 2014, accessed 10 July, 2021, https://ifex.org/the-various-paths-of-internet-censorship-in-latin-america/.

68 See Article 14 WCT and Article 23 WPPT, Provisions on Enforcement of Rights.

69 Rodrigo Vargas Acosta, "Responsabilidad de intermediarios por infracciones a los derechos de autor en Chile, Paraguay y Costa Rica: Un análisis desde la libertad de expresión", *Revista chilena de derecho y tecnología*, Vol. 5, no. 1 (2016), https://doi.org/10.5354/0719-2584.2016.41782.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a) The particular case of the new law in Mexico

A particular case study in Latin America is that of the new Mexican Federal Copyright Law. In response to the United States-Mexico-Canada Agreement (USMCA in the US, T-MEC in Mexico) which took effect on July 1, 2020, the Mexican government reformed its intellectual property legislation, including amendments to its Federal Copyright Law, effective July 2, 2020.

Among the provisions that were amended, is the addition of a notification and withdrawal system that enables Mexican Internet users to file claims when they consider that their copyright is affected by a third party.[70]

The implementation of the notification and withdrawal system generated concern among interested parties, with several civil organizations publicly arguing that these measures affect the exercise of other rights on the internet, for example, freedom of expression, access to culture or information.[71]

Among the main criticisms that the notification and withdrawal system received are the speed with which the reform was approved, due to the pressure to comply in time with the commitments and negotiations derived from the T-MEC. Likewise, the effects that it can generate on the exercise of other rights on the Internet and the errors with which it was incorporated into the LFDA were criticized.

Naturally, the reform was defended by the government and other interested parties. Discussions on the legality of the system by the academic sector and media became more compelling when the National Commission of Human Rights (CNDH), an autonomous constitutional body of Mexico, filed a claim of unconstitutionality before the Supreme Court of Justice of the Nation (SCJN) on the grounds that the notification and withdrawal system may affect the exercise of fundamental rights on the internet.[72]

---

70 Consult articles 114 septies and octies of the Federal Law of Copyright of Mexico http://www.diputados.gob.mx/LeyesBiblio/pdf_mov/Ley_Federal_de_Derechos.pdf

71 "Red de defensa de los derechos digitales. Ni censura ni candados", R3D, accessed June 22, 2021 https://participa.nicensuranicandados.org/.

72 Consult Action of Unconstitutionality 217/2020, https://www.cndh.org.mx/tipo/209/accion-de-inconstitucionalidad?field_fecha_creacion_value%5Bmin%5D=&field_fecha_creacion_value%5Bmax%5D=&keys=217%2F2020&items_per_page=10.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

In addition, in the action of unconstitutionality, the Supreme Court pointed out deficiencies, errors and omissions with which the system was incorporated into the Federal Copyright Law. This claim has not yet been resolved.

Chapter 4. Conclusion

In a region with glaring economic and political inequalities such as Latin America, digital transformation will have a strong impact on inclusiveness, and social media can provide empowerment and, help shape users to progress socially, economically, educationally, and politically.

When the state gives companies more faculties to moderate content, greater control of the public debate falls on the private sector. Yet, is it counterproductive that content can only be moderated if the state allows it? There are models of moderation in online communities, where users decide what kind of content should be filtered, based on the interest of maintaining a healthy dialogue. Ideally, companies could maintain a certain flexibility to decide what content to allow on their platform, but with clear and transparent rules. A social media platform should have to report to the user the reasons a certain expression is being restricted, and there should be appeal mechanisms. Regulation would have to focus on making the exercise of that power accountable.

The challenge for Latin America will be to provide regulation for these activities, whilst crafting rules that safeguard freedom of expression, appropriate to each country's particular domestic social, legal and political contexts, while securing privacy and facilitating civic and social engagement.

Bibliography

"Reseña del Convenio de Berna para la Protección de las Obras Literarias y Artísticas" Organización Mundial de la Propiedad Intelectual, accessed June 22, 2021, https://www.wipo.int/treaties/es/ip/berne/summary_berne.html#:~:text=El%20Convenio%20de%20Berna%20trata,que%20quieran%20valerse%20de%20ellas.

«Directiva sobre el comercio electrónico 2000/31/CE.» 2000. https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:32000L0031&from=ET (último acceso: 2021).

Alejandro Aréchiga Morales. "Sistema de notificación y retirada en México: los derechos en juego," Centro de Política Digital para América Latina, 2021 .

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Banco de Desarrollo de América Latina. "Transformación digital para la América Latina del S. XXI." Accessed June 22, 2021. https://www.caf.com/es/conocimiento/visiones/2020/02/transformacion-digital-para-la-america-latina-del-s21/.

Bojalil, Paulina. "Despuntan las reformas en materia de protección de datos en América Latina." *ABIERTA al público* (blog), February 12, 2019. https://blogs.iadb.org/conocimiento-abierto/es/proteccion-de-datos-gdpr-america-latina/.

Convencion Americana Sobre Derechos Humanos (Pacto de San José), 1969. Organización de Estados Americanos. Articulo 13.

Duarte, Fernando. "Los países en los que la gente pasa más tiempo en las redes sociales (y los líderes en América Latina)." *BBC News Mundo*, September 9, 2019. https://www.bbc.com/mundo/noticias-49634612.

Electronic Frontier Foundation. "La regla de los tres pasos." Accessed June 22, 2021. https://www.eff.org/sites/default/files/filenode/tpp_3pasos.pdf.

Fernández-Molina, J. Carlos and Peis, Eduardo. "The moral rights of authors in the age of digital information." *Journal of the American Society for Information Science and Technology* 52, issue 2, (2001): 109-117. https://doi.org/10.1002/1097-4571(2000)9999:9999%3C::AID-ASI1060%3E3.0.CO;2-B.

Gómez, Marcela and Bobadilla, Matías. "Influencers deben pagar impuestos por ganancias en redes sociales." pauta.cl, March 28, 2021. Pfizenmaier, P. F. (2020, July 16). *Influencers' Regulation In Argentina: When No Law Is Better Than A Bad Law*. Mondaq.Com. https://www.mondaq.com/unitedstates/social-media/965950/influencers39-regulation-in-argentina-when-no-law-is-better-than-a-bad-law

Hernández Bauzá, Valentina. *Sucesos regulatorios en materias de privacidad e internet en Latinoamérica*. Derechos Digitales América Latina, 2020. https://www.derechosdigitales.org/wp-content/uploads/tendencias-privacidad-latam.pdf.

Hinze, Gwen. "Hacer que el conocimiento sea accesible a través de las fronteras: excepciones mínimas obligatorias de derechos de autor internacionales para la educación." *Electronic Frontier Foundation,* October 30, 2008. https://www.eff.org/wp/making-knowledge-accessible-across-borders-case-ma.

Human Rights Watch. "Mexico: Online Free Speech at Risk." April 14, 2021. https://www.hrw.org/news/2021/04/14/mexico-online-free-speech-risk

LA RELATORÍA ESPECIAL PARA LA LIBERTAD DE EXPRESIÓN. (2013, November 11). *Comunicado de Prensa R 85/13*.

Lemus Rivero, Luis Mario. "Influencers, aspectos legales a considerer." *Foro Jurídico*, October 8, 2020. https://orojuridico.mx/influencers-aspectos-legales-a-considerar/#:~:text=Para%20comenzar%2C%20es%20importante%20hacer,Ley%20Federal%20de%20Radio%20y.

Ley 20.435, 2010. Congreso Nacional, Chile.

Ley 26.551, 2009. Congreso de la Nación Argentina, Articulo 5.

Ley 48, 2020. Gaceta del Congreso No. 593, Colombia.

Ley 679, 2001. Congreso de República de Colombia, Capitulo II: Del uso de redes globales de información en relación con menores, Colombia.

Ley Estatutaria 1581, 2012. Congreso de República de Colombia. Artículo 2.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Ley Federal Del Derecho De Autor, 2020, Congreso De Los Estados Unidos Mexicanos, Artículo 114 Octies, México.

Ley Federal Del Derecho De Autor, 2020, Congreso De Los Estados Unidos Mexicanos, Artículo 114 Octies, México.

Moisés Sánchez. *Informe sobre control estatal de redes sociales* (Alianza Regional por la Libre Expresión e Información, 2016), http://www.alianzaregional.net/wp-content/uploads/Informe-Arti%CC%81culo-XIII-2016-GF-SR-DM.pdf.

Observacom. "Redes sociales son las principales vías de acceso a la información en América Latina." *Observatorio Latinoamericano de Regulación de Medios y Convergencia*, June 22, 2020. https://www.observacom.org/redes-sociales-son-las-principales-vias-de-acceso-a-la-informacion-en-america-latina/

OMPI. GUIA SOBRE LOS TRATADOS DE DERECHO DE AUTOR Y DERECHOS CONEXOS. s.f.

Organización Mundial de la Propiedad Intelectual. Reseña del Tratado de la OMPI sobre Derecho de Autor (WCT) (1996). s.f. https://www.wipo.int/treaties/es/ip/wct/summary_wct.html (último acceso: 10 de 04 de 2021).

Ortuño, Carlos I. "COVID-19 and digital inclusion in Latin America and the Caribbean: A connectivity and access problem." *SELA. Latin American And Carribean Economic System*, June 04, 2020. http://www.sela.org/en/press/articles/a/64488/covid-19-digital-inclusion-in-latin-america-and-the-caribbean.

Proyecto de Ley 848/20. Senado De La Nación. Argentina.

R3D. "Red de defensa de los derechos digitales. Ni censura ni candados." Accessed June 22, 2021. https://participa.nicensuranicandados.org/.

Rodrigo Vargas Acosta. "Responsabilidad de intermediarios por infracciones a los derechos de autor en Chile, Paraguay y Costa Rica: Un análisis desde la libertad de expression." Revista chilena de derecho y tecnología, Vol. 5, no. 1 (2016). https://doi.org/10.5354/0719-2584.2016.41782.

Rodríguez María Belén c. Google Inc. s. daños y perjuicios, 2014, Corte Suprema de Justicia de la Nación, Argentina.

Sam Ricketson. "WIPO Study on Limitations and Exceptions of Copyright and Related Rights in the Digital Environment," March 3, 2008. Accessed June 22, 2021. https://www.wipo.int/edocs/mdocs/copyright/en/sccr_9/sccr_9_7.pdf.

Schmitz Vaccaro, Christian. "Evolución de la regulación internacional de la propiedad intellectual." (Concepción, Chile: Universidad Católica de la Santísima Concepción, 2013). https://revistas.uexternado.edu.co/index.php/propin/article/view/3580/3661.

Senador R. M. Ávila, *Iniciativa con proyecto de Decreto por el que se REFORMAN y ADICIONAN diversas disposiciones de la Ley Federal de Telecomunicaciones y Radiodifusión*, Partido Morena, México. https://ricardomonrealavila.com/wp-content/uploads/2021/02/REDES-SOCIALES-Propuesta-Iniciativa-29.01.21.pdf

Sentencia de tutela Radicado No. T-277, 2015, Corte Constitucional, Colombia. https://www.corteconstitucional.gov.co/relatoria/2015/t-277-15.htm

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Sentencia de Unificación Radicado No. SU420/19, 2019, Corte Constitucional, Colombia. https://www.corteconstitucional.gov.co/relatoria/2019/SU420-19.htm

Superintendencia de Industria y Comercio de Colombia. (2020, October 1). *Superindustria expide "Guía de buenas prácticas en la publicidad a través de influenciadores."* Sic.gov.co. https://www.sic.gov.co/slider/superindustria-expide-%E2%80%9Cgu%C3%ADa-de-buenas-pr%C3%A1cticas-en-la-publicidad-trav%C3%A9s-de-influenciadores%E2%80%9D

Vargas Acosta, Rodrigo. *Sucesos regulatorios en materias de libertad de expresión e internet en Latinoamérica*. Derechos Digital América Latina, 2020. https://www.derechosdigitales.org/wp-content/uploads/tendencias-regulacion-digitales.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Topic-based Regulation: Media Law and Data Protection

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Media Law Regulation of Social Networks - Country Report: Germany

Bernd Holznagel, Jan Christopher Kalbhenn

Abstract: In 2018, the German Federal Constitutional Court identified dangers in digitalisation of the media. This leads to "increased difficulty in the separation of fact from opinion, content from advertisement, as well as to new uncertainties regarding the credibility of sources and assessments. Individual users themselves must now process and assess the information provided by the mass media, which would traditionally have passed through the filter of professional selection in the spirit of responsible journalism." German lawmakers have responded to this with the Interstate Media Treaty. For the first time, this treaty sets requirements for content providers on social networks and addresses platforms as content distributors. Supplementary requirements result from the recent case law of the civil courts. The legislature is placing high demands on digital offerings by public broadcasters, who are no less required to provide a counterweight to the dangers of the network and platform economy.

Keywords: content moderation; due diligence; media law; disinformation; hate speech; algorithmic transparency; filter systems; recommendation systems; public service broadcaster; social media; public European space; design specifications

Chapter 1. Increased need for truthful information on the Internet

Initially, there was little knowledge about the novel Corona virus. At the same time, strategies to contain the pandemic required the cooperation of citizens and affected everyone's daily lives. This triggered an increased need for information. Without the filtering function of professional journalism, individuals would have been lost in the flood of news and information. Consequently, the first lockdown was also accompanied by an increase in media usage. The internet had the highest gains (19 percentage points) in informative daily reach. This includes informative use of

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

algorithmic-driven online platforms. One in two people used Google, Facebook and the like to obtain information during the Corona pandemic (increase: 22 percentage points). In Germany, public information and news services are also represented there and generate high demand figures.[1] Since the outbreak of the pandemic, the podcast charts have frequently been topped by the "Coronavirus Update." In it, virologist Christian Drosten regularly explains the latest scientific findings. At the same time, however, offers that spread misinformation also gained in reach. For the German-speaking world, these include the YouTube channels of CompactTV, SCHRANG TV and Games of Truth, which attempted to prove connections between the Corona outbreak and the expansion of the new 5G mobile communications standard. False information spread via social networks, such as that Corona immunity could be obtained injecting disinfectant. Internationally active is the conspiracy theorist network QAnon, which spreads the theory that Bill Gates, the Rothschild family and others have invented Corona as a bioweapon.

Chapter 2. State duty to protect the democratic discourse

The discourse model of the German Constitution (*Grundgesetz*) can only function if political will is formed on the basis of arguments. Social, economic and cultural challenges can only be solved on the basis of facts. In a battle of opinions, the better arguments should win. The German constitution also assumes that public opinion is formed through speech and counter-speech. This takes place through argumentative disputes in the public sphere. The state must stay out of it. However, it has a constitutional duty to protect. [2] Thus, it must shape a positive media order in such a way that the "diversity of existing opinions finds expression in the broadest possible range and completeness." The word 'diversity' in this context is by no means to be equated with the word 'multiplicity'. [3]

It is questionable whether current media law legislation is able to ensure that the public is also supplied with factually correct and credible information on the internet and in social media. It is true that the possibilities of digital distribution channels such as social networks have led

1 Data for Germany in Bernd Holznagel and Jan Kalbhenn, *Monitoring Media Pluralism in the digital Era – Country Report Germany* (2021).

2 Steinebach et al., *Desinformation aufdecken und bekämpfen*, (Baden-Baden: Nomos, 2020).

3 Horst Röper, *Konzentration und Vielfalt im deutschen Rundfunk*, UVK Medien, 1997.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to a differentiation and multiplication of offerings. In this regard, the *Bundesverfassungsgericht* (German Federal Constitutional Court) also states that offerings are often not aimed at diversity of opinion, but are "determined by one-sided interests or the economic rationality of a business model, namely to maximise the time users spend on pages as much as possible and thereby increase the advertising value of the platform for customers." The highest German court also sees a danger in the fact that content is specifically tailored to the interests and inclinations of users by algorithmic means. In this respect, results in search engines are also pre-filtered and partly financed by advertising, partly dependent on "click numbers".[4]

As a result, the Federal Constitutional Court states that digitisation of the media leads "to increased difficulty in the separation of fact from opinion, content from advertisement, as well as to new uncertainties regarding the credibility of sources and assessments. Individual users themselves must now process and assess the information provided by the mass media, which would traditionally have passed through the filter of professional selection in the spirit of responsible journalism."

When the highest court points out such dangers to democratic discourse, it is tantamount to a mandate for legislators to address the problems and consider whether they must fulfil their duty to protect. At the end of 2020, the German states, which are responsible for media law, adopted the Medienstaatsvertrag (Interstate Media Treaty, in short: MStV).[5] For the first time, this sets specifications for content providers on social networks (III.). Platforms as content distributors are also addressed, with additional requirements arising from the recent case law of the civil courts (IV.). The legislator places high demands on the digital offerings of public service broadcasters, who are expected to do no less than act as a counterweight to the dangers of the network and platform economy (V.).

---

4 *Bundesverfassungsgericht*, Decision from 18 July 2018 – BVerfGE 149, 222 („Rundfunkbeitrag“).

5 English version of the Interstate Media Treaty under https://www.die-medienanstalten.de/fileadmin/user_upload/Rechtsgrundlagen/Gesetze_Staatsvertraege/Interstate_Media_Treaty_en.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3. Media law requirements for content on social networks

Chapter 3.a. Journalistic standards of due diligence

Due diligence obligations for the press and broadcasting have long been known in German media law. These providers must check news for content, origin and truth with the due diligence required by the circumstances before disseminating it. The German Press Code serves as a benchmark in this regard. [6] In it, the German Press Council sets out guidelines aimed at truthfully informing the public and enhancing the credibility and reputation of the media. The Press Code also includes the requirement to treat facts as such with care and the requirement to separate editorial and advertising publications. In connection with reporting on a study on the Corona infectivity of children, the Press Council reprimanded the biggest national (boulevard) newspaper BILD in 2020 for breaches of the duty of care. The paper had suppressed several facts important for understanding the study and had obscured a study result that was unpopular in some political circles, namely that children also transmit the Corona virus.

These journalistic standards of care have so far applied on the internet only to the offerings of radio stations and press publishers. Other providers, such as offerings distributed as podcasts or via YouTube by digital native news providers, were not bound by standards and appropriate oversight. Misrepresenting a study here would not be reprimanded on the basis of a journalistic due diligence violation. However, much false information is disseminated via channels whose presentation can hardly be distinguished at first glance from news sources operating according to high standards. Such offerings are widely disseminated via platforms such as YouTube and Instagram. The Interstate Media Treaty extended the obligation to observe journalistic standards to this area as well.[7] Now the rules also apply to "other journalistic-editorial telemedia which regularly contain news or political information".

Things now get complicated when it comes to supervising compliance with these standards. In contrast to the other information services, the online offerings of the press are generally exempt from supervision by the state media authorities. Self-regulation is given priority for information services. They are given the option of joining the Press Council or a volun-

---

6 Press code of the German Press Council https://www.presserat.de/pressekodex.html.

7 § 19 MStV.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tary self-regulation institution. Nevertheless, supervision by the competent state media authority will take place alongside. Voluntary self-regulatory bodies require approval by the state media authorities, and their decisions are subject to review and objection by those authorities. For services not affiliated with voluntary self-regulation, the state media authorities are directly responsible.[8]

Chapter 3.b. Labelling of social bots

In the debate about disinformation, the role of social bots is regularly emphasised. This refers to computer programs that are used on social networks to produce automated content and messages and that appear to originate from a human. To be sure, opinions still differ on the extent to which such computer programs are already being used effectively. However, a new study on media education in Germany shows that digital media education is in a poor state and that many digital phenomena cannot be classified.[9] Social bot programs, for example, are in danger of jeopardizing democratic discourse.[10] Bots can be used to distort public opinion by pushing certain content *en masse*. Bots can also simply disseminate false information and support the virality of certain harmful content. The Interstate Media Treaty now introduces mandatory labelling for providers of social bots on social networks such as Facebook, TikTok, Twitter and YouTube. If such accounts are operated there, the account holders must make the fact of automation known. This is intended to take account of the fundamental potential of these programs to influence individual and public opinion-forming, without completely banning the use of such services. A complete ban on social bots can probably not be justified, if only because they can also be used for harmless and non-political purposes, such as customer advice. In the implementation of these new rules, it will be important that the labels are made in such a way that they are effective and that the users of social networks can classify the accounts accordingly. In practice, this will require the expertise of media designers and media

8 Bernd Holznagel and Jan Christopher Kalbhenn, "Journalistische Sorgfaltspflichten auf YouTube und Instagram", in *Festschrift für Jürgen Taeger,* ed. Specht-Riemenschneider et al. (Frankfurt: R&W, 2020), 589-608.

9 https://www.stiftung-nv.de/sites/default/files/studie_quelleinternet.pdf (last accessed: 15 April 2021).

10 Christian Grimme et al., "Demystifying Social Bots: On the Intelligence of Automated Social Media Actors", *Social Media & Society* Vol. 6, Nr. 3 (2020) 1-14.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

psychologists. Through statutes and case law, concrete requirements for the interface design of social networks can develop from this (design requirements).

The addressees of these obligations are also the social networks. They must ensure that service providers comply with the labelling obligation. How they do this is up to the networks to decide. The only rule is that they must do so carefully.[11]

Chapter 3.c. Labelling of political advertising

The business model of many open platforms, such as social networks, is based substantially to exclusively on advertising. This business model allows the reach of content to be scaled in return for monetary payments. Micro-targeting is considered particularly effective in this context. The advertiser can then have its advertising message displayed to a target group that it can precisely determine in advance on the basis of individual criteria. This is possible because the platforms create incredibly detailed profiles of their users. To advertise baby food, it then makes sense for the advertiser to target young mothers with a certain income. In the analog world, this would require advertising in an appropriate magazine or in the context of an appropriate TV show without the involvement of personal data.[12]

A further risk situation that must be taken into account first arises in the case of political advertising.[13] For example, micro-targeting can be used by a political party to make different 'election promises' to certain selected groups of voters. Young families can be promised more child benefit, while the same party promises to cut child benefit for another voter group and to do more for the care of dogs and cats (dark ads).

Classic media law recognises that political advertising on TV and radio has long been a particularly sensitive category of advertising. It is only permitted under strict conditions in a short period before elections. Otherwise, a strict ban applies. The possibility of banning political advertising on social networks also presented itself with the Interstate Media Treaty, which for the first time contains rules for political advertising in the digital

11 § 18 Sec. 3 MStV.

12 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019).

13 Judit Bayer, "Double harm to voters: data-driven micro-targeting and democratic public discourse", *Internet Policy Review 9(1)* (2020).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

realm.[14] To avert danger, however, the initial focus is on transparency. A labelling obligation for political, ideological and religious advertising will be introduced for all telemedia. This transparency obligation is intended to provide better information about the origin and financing of such advertising. For this reason, the relevant advertisers or clients must be clearly indicated in an appropriate manner. The state media authorities are responsible in each case. The specific dangers of microtargeting are not addressed. The regulation applies to this type of advertising just as it does to such advertising that is displayed to all users of a platform or website. To be fruitful as a tool in the fight against disinformation and contradictory campaign promises, the concept of advertising in this context must be interpreted widely. It must cover all political content whose relevance is promoted by payments to the networks. Otherwise, difficult demarcations may arise. For example, parties and politicians are increasingly using layouted messages that cannot initially be distinguished from paid advertising in purely visual terms.

Chapter 3.d. Interim conclusion

The regulations presented are a clear step forward in combating the spread of disinformation. In the 2016 US election campaign and the Brexit referendum in the same year, so-called dark ads, non-transparent micro-targeting, and social bots in particular were identified as influencing public opinion. With labelling requirements for social bots and political advertising, transparency rules for these problems are now available for the first time. However, the regulations found by the legislature are quite challenging in terms of oversight and enforcement. Many demarcation issues will arise, so that an intensive learning process lies ahead for those involved. This becomes particularly clear in the area of supervision of compliance with journalistic due diligence obligations. Here, the circle of obligated parties is greatly expanded with the information services in the Interstate Media Treaty.[15] But in supervision, too, the state media authorities and institutions of voluntary self-regulation join the Press Council. The regula-

14 § 22 Sec. 3 MStV.
15 § 19 Sec. 1 MStV.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tions proposed by the European Commission in the Digital Services Act go much further, especially in the area of advertising. [16]

Chapter 4. Media law requirements for content moderation on social networks

More and more people are turning to platforms such as Facebook and YouTube for information or news. Just under one in two adult Germans (48 %) has also obtained information about the virus from social media. In the 18- to 24-year-old group, the figure is 72 percent. [17] Journalistic editorial content is also displayed in the Facebook newsfeed or on YouTube. The success or dissemination of disinformation, but also of truthful information, on the internet depends in many cases on platform content moderation. Central to the architecture of the platforms are the filtering and recommendation systems that decide which content comes onto the platforms and which users are shown which content. The programming and the underlying relevance criteria of algorithms thus determine the visibility and reach of content. Content that is recommended by algorithms and also shared by people has a strong chance of becoming viral and achieving an extraordinary reach. Filtering systems that suppress, devalue or delete illegal and harmful content have the effect of curbing reach. For the first time, the new Interstate Media Treaty lays down regulations regarding these systems and thus also regarding content moderation on social networks, as well as flanking design specifications. Particularly harmful and criminal content must be deleted quickly under the rules of the Netzwerkdurchsetzungsgesetz (Network Enforcement Act, in short: NetzDG). Initial legal guidelines for the measures used by platforms to combat disinformation on the basis of their general terms and conditions and community standards can be taken from civil court case law. For the first time, the Interstate Media Treaty also formulates positive findability rules for public value content on digital platforms.

16 Jan Kalbhenn and Maximilian Hemmert-Halswick, "EU-weite Vorgaben zur Content-Moderation auf sozialen Netzwerken", *ZUM – Zeitschrift für Urheber- und Medienrecht* No. 3 (2021), 184.

17 Sascha Hölig and Uwe Hasebrink, *Digital News Report – Germany* (20220) https://www.hans-bredow-institut.de/uploads/media/default/cms/media/66q2yde_AP50_RIDNR20_Deutschland.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 4.a. Specifications for recommendation and filtering systems

a) Transparent recommendation algorithms

For the first time, the new Interstate Media Treaty takes a look at content moderation by recommendation systems. It requires certain platforms to make the central criteria of aggregation, selection, and presentation of content visible.[18] However, this obligation under media law applies to those platforms "which also aggregate, sort and publicly disseminate third-party editorial content".[19] Journalistically editorial offers, i.e. those that contribute to the formation of opinion through a planned activity with the aim of producing and promptly passing on an offer, are disseminated by all common social networks. These must disclose the sorting criteria. This transparency obligation is limited by protection of trade and business secrets, which is why, according to the explanatory memorandum to the law, the algorithm itself does not have to be published. The precise formulation of the transparency rule is left to the state media authorities. They must specify the requirements in statutes that apply nationwide.[20] They will require that the relative weighting of the individual criteria be described. The platforms' optimisation goals should also be transparent. Social networks must then also specify how exactly the findability of content can be influenced by monetary payments and what role profiling plays in this. Information on the architecture and design of the platforms will also be necessary, namely to disclose what influence the functions available to users (sharing, liking, and the like) have.

b) Transparent filter algorithms

Before content can even be captured by recommendation systems, it must first be posted on the platforms. During this process, they are already checked by the social network filter systems and, if necessary, not published at all. These filter systems are particularly advanced in the area of copyright. YouTube's content management system has long been con-

18 § 93 MStV.
19 § 2 MStV.
20 „Satzung der Landesmedienanstalten über die Regulierung von Medienintermediären gemäß § 96 Medienstaatsvertrag“ Draft of the Statue notified at the Commission can be found in the database https://ec.europa.eu/growth/tools-databases/tris/en/search/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

sidered particularly sophisticated and detects possible copyright infringements within a few seconds by comparing uploaded content with content already known in the archive. Systems for detecting "harmful" content that is not tolerated according to platform community standards also filter out content that constitutes hate speech, terrorism, or similar content on a large scale. The Interstate Media Treaty also responds to the use of such filtering systems with a transparency requirement. The social networks must disclose the criteria that determine whether a piece of content can be accessed and remain on the platform. These might, for example, be technical, economic, provider-related, user-related and content-related requirements. Information on content categories, the purpose of the measures, and possibilities of influence through payment must also be made transparent.

c) Prohibition of discrimination of journalistic content

Both the recommendation systems and the filtering systems are subject to a new prohibition of discrimination in favour of journalistic and editorial offerings in the Interstate Media Treaty. [21] This is intended to prevent certain offerings from being over- or under-represented in comparison to other editorial offerings, for example, due to their political orientation or organisational form (private or public) of the provider, and to directly or indirectly impair access or findability.

However, the threshold for the existence of discrimination under the Interstate Media Treaty is quite high; only discrimination of a systematic nature is prohibited. The duration, regularity, possible repetition and systematic nature of the discrimination must be taken into account. A distinction is drawn between two groups of cases of discrimination. Firstly, if the criteria to be published in accordance with the transparency requirements are deviated from without objective reason in favour of or to the detriment of a specific offer. Secondly, if bids are directly or indirectly unfairly systematically impeded by these criteria.

Discrimination may be justified in individual cases. In addition to technical reasons (e.g., displayability of the offering on mobile devices), justification can also be based on lawful conduct. For example, copyright assessments.

---

21 § 94 MStV.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

d) Design specifications

The Interstate Media Treaty stipulates that information on recommendation systems and filtering systems must be "easily perceivable, immediately accessible and constantly available". How this is to be understood will become clear in the medium term. If criteria for easy perceptibility emerge through statutes and case law, this will result in concrete requirements for the design of social networks. The NetzDG makes similar design specifications for the discoverability of complaint options.[22] Here, experience shows that some networks in practice tend to hide the possibility to use the Network-Enforcement-Act-complaint-tools in their design of the user interface. As a consequence, the legislator could feel compelled to set stricter criteria for "easy noticeability", which would then have to be implemented by (product) designers.

Chapter 4.b. Expeditious deletion of certain criminal content

Legal concretisations for the rapid deletion of content are contained in the NetzDG. The law is intended to safeguard rational discourse online and combat the spread of false news via social networks. However, the specified deadlines for deleting illegal content only apply to the catalogue of 22 criminal offences defined in the NetzDG, only a few of which in turn relate to the potential spread of false news (e.g. defamation under Section 187 of the Criminal Code).

Deletions under the NetzDG have so far been of little consequence for Facebook. In the first half of 2020, Facebook deleted 3,913 pieces of content, Twitter 122,302 on the basis of the NetzDG.[23] With the amendment to the NetzDG, the legislator is attempting to push back deletions according to community standards by making it easier to find the complaint option under the NetzDG and expanding reporting obligations. A "put-back" procedure is introduced to safeguard the rights of data subjects under Article 5 (1) of the *Grundgesetz*. In the future, a position will also be taken

22 Jan Kalbhenn and Maximilian Hemmert-Halswick, „Netzwerkdurchsetzungsgesetz“in: *Handbuch Multimedia-Recht*, ed. Hoeren/Holznagel/Sieber (München: C. H. Beck, 2021), part 21.3.

23 “NetzDG Transparenzbericht July 2020“ Facebook https://about.fb.com/wp-content/uploads/2020/07/facebook_netzdg_July_2020_German.pdf; “Community Standards Enforcement Report, Third Quarter 2020,” Facebook, https://transparency.facebook.com/community-standards-enforcement.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

on the use of automated procedures for finding content in transparency reports. Another new way of monitoring content is for those affected to go before a conciliation body. These bodies, which have yet to be established, are to mediate as a low-threshold service for those affected and open up the possibility of involving civil society. This regulatory regime is becoming more differentiated. [24]

a) Establishment of voluntary self-regulation

For example, the first recognised institution of voluntary self-regulation, Voluntary Self-Regulation of Media Service Providers (FSM), has been available since March 2020 for the area of content monitoring under the NetzDG. FSM is a non-profit association and membership is open to companies from the Online Media Sector. In the case of illegality that is difficult to assess, network providers are to be able to consult the FSM. For example, in cases of satire and political opinion campaigns, the limits of freedom of expression are traditionally more difficult to fathom. It is therefore not surprising that this group of cases is frequently represented among the FSM's first cases. Most recently, the panel of 50 lawyers had to judge a comment on Facebook explaining how a standard text can be used to circumvent the mask requirement when shopping. Here, the FSM denied the offence of public solicitation to commit a crime. All decisions are available online.[25] This is how a canon for content control on social networks can be created.

b) Case Study "Liberation of Germany from the Merkel Regime"

A post on Facebook was calling for action against the "Merkel Regime".[26] It reads as follows: "Half a million to a million Germans plan to demonstrate on August 1 in Berlin, a great opportunity to liberate Germany from the Merkel regime (Freemasons' puppet). Tens of thousands of Germans must storm into the Chancellery, occupy entire buildings, and the Committee to Rescue Germany takes over government. (…) The parliamentary

24 Kalbhenn and Hemmert-Halswick, *Netzwerkdurchsetzungsgesetz*.
25 Online Archive of FSM to be accessed under https://www.fsm.de/de.
26 FSM „Entscheidung Aktenzeichen NetzDG0092020" https://www.fsm.de/de, (translation by the authors).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

party system is past and finished, now the people, parliament and the Committee to rescue Germany decide on a future Germany. Merkel and the whole cabinet and colleagues, all party functionaries of the CDU, CSU, FDP, SPD, Greens and the LEFT, all constitutional judges, ARD ZDF directors and moderators, all the lying press (newspapers) owners and reporters, fascist terrorist Antifa groups, all must be arrested immediately for high treason, and a military court must decide on the fate of these traitors. (...) destroy all Masonic Lodges with their members – Soros - Bill Gates-Antifa-Greta-EU-NATO, destroy all Anti-Christian Party politicians, Islamists, Anti-Christian leaders (Cardinals, Bishops) in the Vatican official church who are Freemasons' puppets (...)".

This post could qualify as public provocation to commit crimes under the German Criminal Code. Then the content would be unlawful in the sense of the NetzDG and would have to be deleted. But this decision is not easy. Facebook also did not find the assessment easy and made use of the option to submit the case to the FSM. In order to be recognised as a self-regulatory body, institutions must meet a number of requirements. The installation of such self-regulated bodies is provided for in the NetzDG. These bodies must meet certain requirements prescribed by law to be recognised by the Federal Office of Justice. These lawyers are paid by the FSM. For example, they have to secure proper equipment for the examiners, guarantee a rapid testing within seven days and provide transparent rules of procedure. According to its procedural rules the FSM works in several "Audit Committees". These Audit Committees are composed of three persons, who appoint a chairman from among themselves. They work in line with a schedule of responsibilities. The Committee members are installed for at least one year, and must have the qualification for the office of a judge. Incompatibility rules define who is not allowed for the task, for example lawyers working for the same law firm representing the company, lawyers appointed by social networks, and employees of media authorities. Only social networks that are members of the FSM are entitled to submit applications. The procedure to be followed is prescribed as well. Social networks can request the FSM to decide on a case by email. These applications must meet formal requirements, for example in regard to their completeness. The FSM administrative office forwards the case to the competent audit committee, which has to decide within seven days. The committee can decide by telephone or in writing. The final decision must be submitted in writing and contain facts and reasons. It must state whether submitted content is unlawful in the sense of the NetzDG. All decisions must be published. If the submitted content is unlawful, the social network must take immediate action. The FSM procedure also has

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

complaint options. The uploader can request a review of the FSM decision. The deadline for this is two weeks and it may lead to a new decision. FSM decisions must be based on current German jurisdiction. The examination is limited to the question whether content is unlawful in the sense of NetzDG.

It was according to these rules that this self-regulated body dealt with the "The Merkel Regime" case. The FSM considered human rights and the jurisprudence of Federal Constitutional Court.

In this case the FSM considered and weighted the fundamental right of freedom of expression (Art. 5 Grundgesetz) and used leading judgements of the Federal Constitutional Court as a benchmark for its examination. It recalled the Federal Constitutional Court and said: "when interpreting expressions of opinion that aim to influence the opinion-forming process and are subject to freedom of opinion, the content of the declaration must also be determined against the background of social and political events". The FSM decided in favour of the User, namely that the post on Facebook does not constitute unlawful content in the sense of the NetzDG.

Chapter 4.c. Civil court requirements for content moderation according to community standards

According to its community standards, Facebook took action against 22.5 million pieces of content globally for hate speech from April to June 2020.[27] This may also affect content that is still protected by freedom of expression, such as certain forms of conspiracy theories or fake news. This raises the question of the extent to which private platforms are bound by freedom of expression when providing a public communication space. With regard to Facebook, the Federal Constitutional Court has already stated that it is "precisely for the dissemination of political programmes and ideas [...] a medium of paramount importance that is not readily replaceable" and that exclusion from the platform denies an essential opportunity to disseminate political messages and actively engage in discourse with users. [28]

---

27 "NetzDG Transparenzbericht July 2020" Facebook https://about.fb.com/wp-content/uploads/2020/07/facebook_netzdg_July_2020_German.pdf; "Community Standards Enforcement Report, Third Quarter 2020," Facebook, https://transparency.facebook.com/community-standards-enforcement.

28 *Bundesverfassungsgericht*, Decision from 22 May 2019 – 1 BvQ 42/19 („III. Weg").

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a) Hate speech

Community standards and platform content deletions based on them are fully reviewable in the German civil courts. There is already a broad canon of case law on paragraph 12 of community standards (hate speech). When examining the legality of the deletion decision in civil court, the courts first deal with the effectiveness of the platform Terms of Service. In Clause 12, Facebook reserves the right to delete hate speech, which it defines "as a direct attack on individuals based on protected characteristics: ethnicity, national origin, religious affiliation, sexual orientation, caste, gender, gender identity, serious illness or serious disability." The majority of courts consider Clause 12 (hate speech) to be a permissible contractual clause. Under German constitutional law, fundamental rights also apply between private actors. The courts then weigh the fundamental rights positions of the users (freedom of expression and the principle of equality) against those of the platforms (fundamental economic rights). Emphasis is placed on Facebook's interest in structuring its terms of use in its own business interests in such a way that people with different backgrounds and different values and moral concepts feel as unaffected and comfortable as possible. The restriction on users' freedom of expression is mitigated by the fact that, in principle, "humour and social criticism," among other things, are permitted in connection with topics covered by hate speech. Occasionally, courts recognise in the community standards an unreasonable disadvantage contrary to good faith, because operators of a public marketplace for information such as Facebook must ensure that a lawful expression of opinion is not removed. [29] Deletion can then only be considered if content is illegal, for example, if it violates one of the provisions of Section 1 (3) NetzDG.

b) Fact-checking

So-called fact-checking in particular is sometimes seen as a proven antidote to disinformation. Paragraph 21 of Facebook's Community Standards states: "We want to help people stay informed without hindering productive public discussion." Following a partially automated process established by Facebook, potential fake reports are identified and submitted to

29 Kalbhenn and Hemmert-Halswick, *Netzwerkdurchsetzungsgesetz* for an overview on Jurisprudence of German Courts.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

service providers for review. They can then flag the report accordingly, triggering a significant reach restriction. By default, the fact-checker's website is also linked to a call for donations.

The Karlsruhe Higher Regional Court (Oberlandesgericht) had to decide on a specific case of fact-checking on Facebook. An article by a medium called Tichys Einblick linked on Facebook had the headline "500 scientists declare: 'There is no climate emergency'." Facebook's fact-checking service provider 'Correctiv' inextricably linked the corresponding post with the note "fact-check" and "assertion partly false." The reasoning for that labelling was that not all signatories of the declaration mentioned in the article were "scientists".

The Oberlandesgericht Karlsruhe assumed an anti-competitive business act here since both Tichys Einblick and Correctiv were media outlets competing for attention.[30] Because freedom of opinion also applies between private actors under German constitutional law, the court consequently sets standards for fact-checking that would apply accordingly to a state actor. Thus, in journalistic competition, the duty of neutrality must be observed: Certain opinions or tendencies may not be favoured or disadvantaged by promotion. The principle of equal opportunity in communication must be observed: If the credibility of a particular participant in journalistic competition is particularly emphasised or publications would always include a reference to the opposing view or even all competing views, this would require special justification. The fact-checker must be particularly careful to avoid any misunderstanding as to which statement their criticism refers to, who made the statement, and whether the criticism is primarily evaluative or factual in nature. Incorrect information does not constitute an asset worthy of protection from the point of view of opinion formation.

Chapter 4.d. Findability of truthful content in user interfaces

In view of the flood of information, it is important that socially significant information and news offerings in particular can be found by the user community at all. Media policy has long called for making it easier to find public value offerings. The Interstate Media Treaty now introduces findability rules, but only for smart TV devices, streaming sticks and smart

30 *Oberlandesgericht Karlsruhe*, Decision from 09 September 2020, - 6 U 38/19 („Tichys Einblick").

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

speakers. [31] Public value offerings must be made easy to find through a highlighted presentation. These are primarily offerings by public broadcasters. Commercial offerings can also be specified by the state media authorities. This is intended to take into account the increasing importance of findability and positively ensure diversity.

Netflix and Amazon Prime, which are in particularly high demand among young people, are not subject to any obligations to make certain content easy to find or to grant access to the offering in the first place. This recognises the editorial sovereignty and programming of these services, as is familiar with other final media products (newspapers, TV programming). Facebook Newsfeed, Google Search or YouTube are also not subject to any discoverability rules as so-called media intermediaries. Corresponding demands failed in the countries where these companies have their German headquarters. Here, trust is still placed in self-regulation by providers. However, architecture specifications for the platforms are also possible. Under certain conditions, they could be required to programme their algorithms for diversity.

Chapter 5. Interim conclusion

Digital platforms are the main channel for dissemination of fake news, because the effects of communication based on algorithms can increase the spread enormously ("viral effects"). Regulation in this area is just becoming more differentiated in Germany. On the one hand, this applies to the Interstate Media Treaty, which provides transparency and non-discrimination rules for recommendation systems and filtering systems, but which only protects journalistic and editorial content. For content that is found to be criminal, the NetzDG requires expeditious deletion, but the decision-making process for this is distributed among various pillars and all parties involved are protected by procedural rules. The Digital Services Act also follows these regulatory approaches, even in some cases going beyond them.[32] Furthermore, the platforms have plenty of room to shape their content moderation and also align it with their economic goals. The case law of the civil courts provides some initial guidelines in this regard. In July 2021, the Federal Court of Justice ruled that it is necessary for Facebook to undertake in its terms and conditions to inform the user

31 § 84 MStV.

32 Kalbhenn and Hemmert-Halswick, *EU-weite Vorgaben zur Content-Moderation*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

concerned about the removal of a post at least retrospectively and about an intended blocking of his user account in advance, to inform him of the reason for this and to give him an opportunity to respond, followed by a new decision.[33] What is sorely lacking for digital platforms is, above all, discoverability rules for high-quality content, such as that of public broadcasters.

Chapter 6. High requirements as to content of public service media

The services provided by public broadcasters continue to enjoy high priority in Germany. In addition to a nationwide TV service (ZDF) and radio service (Deutschlandradio), public broadcasting is organised on a federal basis. Nine broadcasters distribute TV and radio programmes. Since 2019, broadcasters have also been increasingly active on the internet. The broadcasters' Corona coverage has been well received and approval ratings are at a high level. [34] However, rapid developments in the area of digital platform competition are causing problems. The development of public broadcasting is regularly driven by the case law of the Federal Constitutional Court. In 2018, the court defined the mission of public broadcasting in the digital world and also positioned it against disinformation and other threats fuelled by the network and platform economy.

Chapter 6.a. Public service broadcaster as "counterweight"

In 2018, the Federal Constitutional Court had to rule on a completely different question, but did not miss the opportunity to define the role of contribution-financed public broadcasting in the digital media world. The network and platform economy of the internet, including social networks, leads to "increasingly difficult separability between facts and opinion, content and advertising, as well as to new uncertainties regarding the credibility of sources and evaluations. The individual user must take over the processing and mass media evaluation that traditionally takes place

33 Bundesgerichtshof Decision from 29 July 2021 - III ZR 179/20 und III ZR 192/20, press release unter https://www.bundesgerichtshof.de/SharedDocs/Pressemitteilungen/DE/2021/2021149.html.

34 Nationwide survey Infratest Dimap, "Glaubwürdigkeit der Medien 2020„ *Auftraggeber: WDR* https://www1.wdr.de/unternehmen/der-wdr/unternehmen/studie-deutsche-medien-glaubwuerdig-106.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

through the filter of professional selections and responsible journalistic action."[35] For the court, it follows that in view of this development, the importance of the task incumbent on contribution-financed public broadcasting grows "through authentic, carefully researched information that keeps facts and opinions apart, does not present reality in a distorted way and does not put the sensational in the foreground, but rather forms a counterweight that ensures diversity and offers guidance."[36]

Chapter 6.b. Expansion of entitlements for online program

For a long time, the public broadcaster was only allowed to post in its media libraries the programmes already broadcast linearly, limited to seven days. In order to be able to form the counterweight demanded by the Federal Constitutional Court, a significant expansion of the public broadcaster's scope for action was created in 2019.[37] The core point of the reform was to mandate broadcasters to produce and distribute content that is oriented to the specifics of the internet and social media (i.e., online only). Thus, the mandatory reference of online offerings to a previously linear broadcast was abandoned. Content in media libraries can now also be available for longer than a week. In addition, broadcasters are authorised to network their content with each other. They are also to do this with the digital offerings of public cultural and educational services. Furthermore, the online offerings of broadcasters are to provide guidance, enable all population groups to participate in the information society, offer interactive communication, and promote the technical and content-related media competence of all generations and minorities. Information, education and advice are among the legal core tasks of public service telemedia as well. In this context, the principles of objectivity and impartiality of reporting, diversity of opinion, and balance of offerings must be taken into account.

Nevertheless, it is questionable whether public broadcasting can hold its own against international streaming platforms from Hollywood or Silicon

35 *Bundesverfassungsgericht*, Decision July 18 2018 – BVerfGE 149, 222 („Rundfunkbeitrag“).

36 *Bundesverfassungsgericht*, Decision July 18 2018 – BVerfGE 149, 222 („Rundfunkbeitrag“).

37 Jan Kalbhenn and Christian Schepers, “Öffentlich-rechtliche Telemedien und digitale Kommunikationsplattformen – Die digitalen Angebote von ARD, ZDF und Deutschlandradio auf Instagram, Netflix und Spotify“, *K&R – Kommunikation und Recht*, No. 5 (2021), 316-322.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Valley despite these possibilities. [38] In the medium term, this would probably require the development of a platform of its own.

Chapter 6.c. Further development into a public interest-oriented platform

Even before Netflix launched its service in Germany, ARD and ZDF had plans for a joint VOD streaming portal by 2014. These public broadcasters wanted to build a joint platform for accessing movies, series and other programmes. To this end, their own pool of content was to be supplemented by third-party content. A subscription model and an advertising-financed variant were planned. However, the German Federal Cartel Office was sceptical about the concept and expressed concern that the planned platform would have prevented other alternative platforms from entering the market. Just a few years later, this legal opinion would have been dismissed as absurd. As a result, there are increasing calls for a platform under public law that is to be oriented toward the common good. In the European context, plans are afoot for a digital platform for quality content.[39] In essence, this is an alternative to the existing monopoly providers Facebook and Google. The platform is to bring together the media libraries of public and private broadcasters, portals of publishers and cultural institutions such as universities, museums, and archives. In addition to this curated part, the platform is also to include various aggregating functions: in addition to search engines, these could also be 'citizen accounts' for mutual exchange. It should promote social cohesion and be committed to a citizen-friendly approach to Big Data. On the content side, competition is to prevail. A concept for this is now available. According to this, the "European Public Sphere" is to be designated as an open digital ecosystem divided into different levels and components. The basic technology is a cloud infrastructure as the foundation of the ecosystem. On top of this, on a second level, technology platforms are to provide applications such as "video player," "search," "translation," and "identity" as building blocks. Levels one and two form central elements for an open and digital ecosystem. Thus, a third

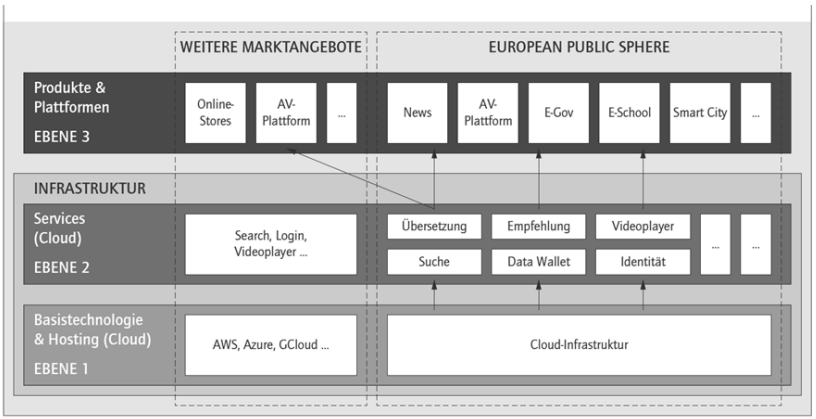
38 Hennig-Thurau et al., *Angriff aus Hollywood. Was es für den deutschen Streaming- und Fernsehmarkt bedeutet, wenn Hollywood-Studios zu Konkurrenten werden*, 2021, 26, https://www.marketingcenter.de/sites/mcm/files/downloads/news/2021/lmm_angriff_aus_hollywood.pdf.

39 Henning Kagermann and Ulrich Wilhelm (publisher), "European Public Sphere. Gestaltung der digitalen Souveränität Europas", *acatech* (2020).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

level should be able to provide a variety of offerings. Public services can use the infrastructure to offer smart city or e-school applications.[40]

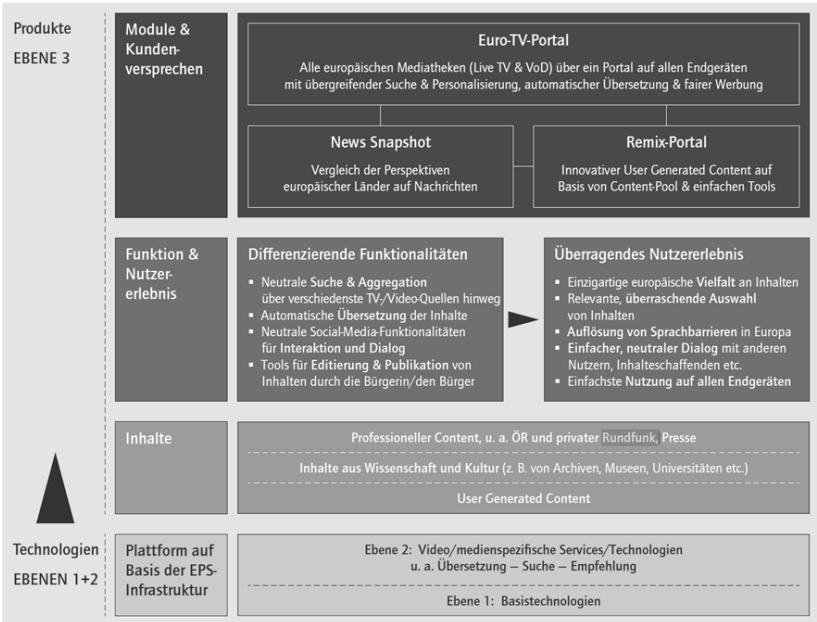
1. Figure: Graphics from Henning Kagermann and Ulrich Wilhelm (publisher), 'European Public Sphere. Gestaltung der digitalen Souveränität Europas', acatech (2020).



One focus of the model is on digital media. The concept paper states that the "European Public Sphere" is characterised in the area of digital media by offerings whose content and functionality can keep pace with today's content offerings. In addition, users should be offered new opportunities to form their own and public opinions. International diversity through European content should enable citizens to develop broader perspectives on diverse topics. Finally, the concept states, "Transparent rules of conduct and control mechanisms will prevent fake news and filter bubbles and enable open, democratic discourse. At all times, there is trust in the protection of one's own data."

40 Graphics taken from Kagermann and Wilhelm, "European Public Sphere. Gestaltung der digitalen Souveränität Europas".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA



The state must lead the way in such a solution. The digital infrastructure can only be created via a state effort. But the political will must be there. At the very least, existing public offerings at national and European levels should be linked to each other and to offerings from the fields of culture and science. These are not platform solutions, but they are better than nothing in view of rapid developments. How such a solution is to be financed has not yet been clarified. At present, the financing of the conventional offering is being put to the test.

Chapter 6.d. Funding of public service content

In Germany, public broadcasters have a constitutional right to funding in line with their needs. The structural and programming decisions of broadcasters in connection with provision of telemedia services and the associated cost requirements therefore also enjoy constitutional protection. For

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the years 2021 to 2024, this amounts to a fixed EUR 1371.1 million. [41] The independent expert panel of the KEF (Commission to Determine the Financial Requirements of Public Service Broadcasting) has recommended raising the broadcasting contribution by 86 cents from EUR 17.50 to a total of EUR 18.36 per household per month from January 1, 2021. The state governments of all the federal states must agree to this. At the end of 2020, this failed in the state parliament of Saxony-Anhalt. The broadcasters initiated proceedings before the Federal Constitutional Court, in which many legal questions will be raised but also the future impact of digital public value offerings as a counterweight will be decided.

Chapter 7. Overview of instruments

	Interstate Media Treaty	NetzDG
Transparency reporting		■
Timely Deletion of severe criminal content		■
Cooperation with national authorities following orders		■
Points of contact and, where necessary, legal representative	■	■
Notice and action and obligation to provide information to users		■
Complaint and redress mechanism and out of court dispute settlement		■
Trusted flaggers		
Findability of Public Service Content	■	
Labelling of Social Bots / Chatbots	■	
Labelling of political advertising	■	
Reporting criminal offences		■

41 See 22. KEF-Bericht, Rn. 56., https://kef-online.de/fileadmin/KEF/Dateien/Berichte/22._Bericht.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

	Interstate Media Treaty	NetzDG
Risk management obligations and compliance officer		
Transparency of recommender systems	■	
Data sharing with authorities and researchers		■
Non-Discrimination clause for journalistic content	■	
Crisis response cooperation		

Chapter 8. Conclusion

False information poses a risk to society as well as to the individual. The media legislator has recognised this and consequently introduced initial instruments. However, blanket solutions are not offered. The regulations presented concern a differentiated group of obligated parties and, in addition to transparency and due diligence obligations, also strive for discoverability privileges for certain content. A valuable asset in the fight against misinformation is Germany's strong public broadcasting system. It is mandated and empowered to act as a counterweight on the internet. It is impossible to predict whether the rules in the Interstate Media Treaty will be effective in ensuring credible information, especially since enforceability will be challenging. Initial experience with the instruments can be fruitful in the discussions about the Digital Services Act and the Artificial Intelligence Act,[42] whose draft contains similar instruments, some of which go further.[43]

The synopsis of the instruments and measures presented should be understood as a learning system. A wide variety of factors play a role in the spread of misinformation. In addition to a wide variety of actors (states, organisations, individuals) with different interests (economic, racial, polit-

42 Jan Kalbhenn, "Designvorgaben für Chatbots, Deepfakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung", ZUM – *Zeitschrift für Urheber- und Medienrecht,* No. 8/9 (2021).

43 See Chapter of Jan Kalbhenn "European legislative initiative for very large communication platforms".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ical), technology is the most important factor. This, however, is subject to continuous change. According to Moore's Law, the complexity of integrated circuits doubles every 12 months. The platform-dominated and algorithm-driven world of communication also comes up with new technologies and tools at ever shorter intervals. Platforms are continuously changing their design and their architecture, drawing on the insights of leading cognitive psychologists. In the Netflix documentary "The Social Media Dilemma," design ethicist Tristan Harris, formerly of Google, explains that many design features of social media are borrowed directly from the gambling industry's Las Vegas experience. This does not make the difficult undertaking of guaranteeing supply of trustworthy information any easier. Rather, it should be an incentive to defend the public debate space on the internet against commercial interests with a digital platform oriented toward the common good under the strong leadership of public broadcasting. This is an inter- and transdisciplinary task with computer scientists, designers, cognitive scientists and economists to be involved.

Bibliography

Assenmacher, Dennis, Lena Clever, Lena Frischlich, Thorsten Quandt, Heike Trautmann, and Christian Grimme. 'Demystifying Social Bots: On the Intelligence of Automated Social Media Actors.' Social media + society 6, no. 3 (July-September 2020): 1-14. https://doi.org/10.1177/2056305120939264.

Bayer, Judit. 'Double harm to voters: data-driven micro-targeting and democratic public discourse' Internet Policy Review 9(1) (2020).

Deutscher Presserat. 'Publizistische Grundsätze (Pressekodex). Richtlinien für die publizistische Arbeit nach den Empfehlungen des Deutschen Presserats. Beschwerdeordnung.' Accessed 20 April 2021. https://www.presserat.de/pressekodex.html.

Facebook. 'Community Standards Enforcement Report.' Accessed 20 April 2021. https://transparency.facebook.com/community-standards-enforcement.

Facebook. 'NetzDG Transparenzbericht.' 2020. Accessed 20 April 2021. https://about.fb.com/wp-content/uploads/2020/07/facebook_netzdg_July_2020_German.pdf.

Hennig-Thurau, Thorsten, Ricarda Schauerte, Niko Herborg, Veronika Schneid, and Nico Wiegand. 'Angriff aus Hollywood. Was es für den deutschen Streaming und Fernsehmarkt bedeutet, wenn Hollywood Studios zu Konkurrenten werden.' 2021. Accessed 20 April 2021. https://www.marketingcenter.de/sites/mcm/files/downloads/news/2021/lmm_angriff_aus_hollywood.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Hoeren, Thomas, Ulrich Sieber, und Bernd Holznagel. Handbuch Multimedia-Recht. Rechtsfragen des elektronische Geschäftsverkehrs. München: C. H. Beck, 2021.

Hölig, Sacha and Uwe Hasebrink. 'Reuters Institute Digital News Report 2020. Ergebnisse für Deutschland. Unter Mitarbeit von Julia Behre.' Arbeitspapiere des Hans-Bredow-Instituts. Projektergebnisse Nr. 50. Hamburg: Verlag Hans-Bredow-Institut, 2020. Accessed 20 April 2021. https://www.hans-bredow-institut.de/uploads/media/default/cms/media/66q2yde_AP50_RIDNR20_Deutschland.pdf.

Holznagel, Bernd and Jan Christopher Kalbhenn. ‚Journalistische Sorgfaltspflichten auf YouTube und Instagram', in Festschrift für Jürgen Taeger, ed. Specht-Riemenschneider et al. (Frankfurt: R&W, 2020): 589-608.

Holznagel, Bernd and Kalbhenn, Jan. ‚Monitoring Media Pluralism in the digital Era – Country Report Germany 2021' Florence (2021). Accessed 20 August 2021. https://cadmus.eui.eu/bitstream/handle/1814/71947/germany_results_mpm_2021_cmpf.pdf?sequence=1&isAllowed=y

Kagermann, Henning und Ulrich Wilhelm. 'European Public Sphere. Towards Digital Sovereignty for Europe', acatech, (2020). Accessed 20 April 2021. https://www.acatech.de/publikation/european-public-sphere/.

Kalbhenn, Jan Christopher and Christian Schepers. 'Öffentlich-rechtliche Telemedien und digitale Kommunikationsplattformen – Die digitalen Angebote von ARD, ZDF und Deutschlandradio auf Instagram, Netflix und Spotify' Kommunikation und Recht, no. 5 (2021).

Kalbhenn, Jan Christopher and Maximilian Hemmert-Halswick. 'EU-weite Vorgaben zur Content-Moderation auf sozialen Netzwerken', ZUM, no. 3 (2021): 184-194.

Kalbhenn, Jan and Maximilian Hemmert-Halswick, ‚Netzwerkdurchsetzungsgesetz', in: Handbuch Multimedia-Recht, ed. Hoeren/Holznagel/Sieber (München: C. H. Beck, 2021), part 21.3.

Kalbhenn, Jan, Designvorgaben für Chatbots, Deepfakes und Emotionserkennungssysteme: Der Vorschlag der Europäischen Kommission zu einer KI-VO als Erweiterung der medienrechtlichen Plattformregulierung', *ZUM – Zeitschrift für Urheber- und Medienrecht*, *No. 8/9* (2021) 663-674.

Kohl, Helmut. Vielfalt im Rundfunk – Interdisziplinäre und internationale Annäherungen, UVK Medien, 1997.

Kommission zur Ermittlung des Finanzbedarfs der Rundfunkanstalten. '22. Bericht.' Februar 2020. Accessed 20 April 2021. https://kef-online.de/fileadmin/KEF/Dateien/Berichte/22._Bericht.pdf.

Martin Steinebach, Katarina Bader, Lars Rinsdorf, Nicole Krämer, and Alexander Roßnagel. Desinformation aufdecken und bekämpfen. Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität. Baden-Baden: Nomos, 2020.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Meßmer, Anna-Katharina, Alexander Sängerlaub, and Leonie Schulz. '"Quelle Internet?" Digitale Nachrichten- und Informationskompetenzen der deutschen Bevölkerung im Test.' Accessed 20 April 2021. https://www.stiftung-nv.de/sites/default/files/studie_quelleinternet.pdf.

Specht-Riemenschneider, Louisa, Benedikt Buchner, Christian Heinze, and Oliver Thomsen. Festschrift für Jürgen Taeger. Frankfurt am Main: Deutscher Fachverlag GmbH, Fachmedien Recht und Wirtschaft, 2020.

WDR. 'Mehr Menschen halten Medien in Deutschland für glaubwürdig.' 2020. Accessed 20 April 2021. https://www1.wdr.de/unternehmen/der-wdr/unternehmen/studie-deutsche-medien-glaubwuerdig-106.html.

Zuboff, Shoshana. 'The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.' New York: PublicAffairs, (2019).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Reshaping Canada's Broadcasting Act: Solutions in Search of a Problem?

Michael Geist

Abstract: The Canadian government introduced much-anticipated reforms to Canada's national broadcasting legislation in November 2020. The legislation – Bill C-10 – was framed as a long-needed update to rules that had not been updated in decades and failed to account for the emergence of Internet streaming services such as Netflix and Disney+ that had become enormously popular with Canadian subscribers. The bill emerged as one of the most controversial political issues in Canada, highlighting the challenges of regulating Internet services. While the public is broadly supportive of new regulations to address concerns about powerful Internet companies, the Canadian experience suggests there are concerns about the implications for freedom of expression, over-broad regulation, competition, and consumer costs.

Keywords: Broadcasting, Internet, Internet streaming, freedom of expression, culture, Netflix, Canada, Bill C-10, Social Media

Chapter 1. Introduction

The Canadian government introduced much-anticipated reforms to Canada's national broadcasting legislation in November 2020.[1] The legislation – Bill C-10 – was framed as a long-needed update to rules that had not been updated in decades and failed to account for the emergence of Internet streaming services such as Netflix and Disney+ that had become enormously popular with Canadian subscribers. The initial response to the bill was largely positive, with support from Canadian creator groups and

1 House of Commons of Canada, *An Act to amend the Broadcasting Act and to make consequential amendments to other Acts*, Bill C-10, 2nd sess, introduced in House November 3, 2020, https://www.parl.ca/LegisInfo/BillDetails.aspx?Language=E&billId=10926636.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

indifference from much of the Canadian public. Yet months later, the bill was among the most controversial political issues in Canada, with political parties devoting days to debate it in the House of Commons, thousands signing petitions to protest the bill, and the government resorting to rarely used parliamentary maneuvers in order to shepherd the bill toward a final vote.

The battle over Bill C-10 highlights the challenges of regulating Internet services. While the public is broadly supportive of new regulations to address concerns about powerful Internet companies,[2] there are clearly concerns about the implications for freedom of expression, over-broad regulation, competition, and consumer costs. As a result, government proposals that enjoy support within the narrow confines of the cultural community may face noisy opposition from the broader public.

This chapter examines the Canadian legislative experience with crafting Internet rules that rely primarily on broadcasting reform. The chapter begins by highlighting the decades-long policy process that included an initial exemption for Internet services, the consistent demands that those early policies be revisited, and the foundation for Bill C-10. The chapter critiques many aspects of the bill, noting that many policy issues are more complex than simply pointing to the need for a "level playing field" or bringing Internet companies into a national regulatory framework. Indeed, the Canadian experience provides a cautionary tale on Internet regulation and points to the need to rethink whether conventional broadcast legislation is the optimal regulatory model.

Chapter 2. The Long Road to Internet Regulation

The Canadian legal community became one of the first to consider the transformative effects of the Internet when the Canadian Radio-Television and Telecommunications Commission ("CRTC"), the country's lead regulator on broadcast and telecommunications matters, launched its new media hearings in the summer of 1998.[3] Although the Canadian legal

2 See, e.g, The Strategic Counsel, *A Report to CIRA* (Toronto, Ottawa, Calgary, Houston: The Strategic Counsel, 2021), https://www.cira.ca/sites/default/files/2021-05/CIRA_better-internet-report-2021.pdf.

3 Canadian Radio-television and Telecommunications Commission, Public Notice CRTC 1998-82 - New Media - Call for Comments (Ottawa: Canadian Radio-television and Telecommunications Commission, 1998), https://crtc.gc.ca/eng/archive/1998/PB98-82.htm.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

and media communities expressed concern that the CRTC would use the hearings to establish new regulations to police the Internet, the final report yielded the opposite approach.[4] In fact, the CRTC heeded the barrage of submissions from organizations imploring it to refrain from establishing new regulations. At that time, it accepted arguments regarding the perceived futility of traditional regulatory approaches and the benefits of providing new media companies with the regulatory space to develop unhindered.

After reviewing current Internet activity and the definition of "broadcasting," the CRTC held that the majority of services then-available on the Internet consisted predominantly of alphanumeric text, and therefore fell outside the scope of the Broadcasting Act and outside the Commission's jurisdiction. Moreover, new media services where the potential for user customization was significant (as with end-users who create their own uniquely tailored content) was also deemed not to be transmission of programs for reception by the public, and therefore fell outside the scope of the Broadcasting Act.[5]

The CRTC did conclude that some new media services fall under the Broadcasting Act's definitions of "program" and "broadcasting", however. Included was Internet content that consists only of "audio, video, a combination of audio and video, or other visual images including still images that do not consist predominantly of alphanumeric text."[6]

Notwithstanding the application of the Broadcasting Act to certain forms of Internet broadcasting, the CRTC concluded that, for new media which falls under the definition of "broadcasting," regulation "will not contribute in a material manner to the implementation of the policy objectives set out in section 3(1) of the Act."[7] Accordingly, pursuant to section 9(4) of the Broadcasting Act, an exemption order was proposed with respect to all new media undertakings that are providing broadcasting services over the Internet, in whole or in part, in Canada.[8]

---

4 Canadian Radio-television and Telecommunications Commission, "Public Notice CRTC 1999-84 - Report on New Media," 1999, https://crtc.gc.ca/eng/archive/1999/PB99-84.htm.

5 Canadian Radio-television and Telecommunications Commission, "Public Notice".

6 Canadian Radio-television and Telecommunications Commission, "Public Notice".

7 Canadian Radio-television and Telecommunications Commission, "Public Notice".

8 *Broadcasting Act*, SC 1991, c 11, s 9(4).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

In 2009, 10 years after issuing its original exemption order, the Commission revisited the issue.[9] After days of hearings and thousands of pages of submissions, the Commission again side-stepped the pressure to "do something," by maintaining its hands-off approach. It concluded that regulatory intervention would impede innovation. Indeed, the decision noted that "the Commission is of the view that parties advocating repeal of the exemption orders did not establish that licensing undertakings in the new media environment would contribute in a material manner to the implementation of the broadcasting policy set out in the Act."[10]

There was at least one very noteworthy change to the new media exemption, however. The CRTC was clearly troubled by allegations of undue preferences being granted by wireless providers and proposed amendments prohibiting such practices.[11] Looking into the future, the Commission planned to review the decision within five years, initiate a reference at the Federal Court to sort out the status of ISPs within the Broadcasting Act, and extend the scope of new media monitoring by requiring "new media broadcasting undertakings to report details of their new media broadcasting activities, which may include broadcasting content usage and offerings, revenues and expenditures, at such time and in such form, as requested by the Commission."[12]

At the time, Commissioner Tim Denton raised concerns about the content provisions of the Broadcasting Act in a powerful concurring opinion, concluding "the rights of Canadians to talk and communicate across the Internet are vastly too important to be subjected to a scheme of

9 Canadian Radio-television and Telecommunications Commission, *Broadcasting Regulatory Policy CRTC 2009-329* (Ottawa: Canadian Radio-television and Telecommunications Commission, 2009), https://crtc.gc.ca/eng/archive/2009/2009-329.pdf.

10 *Ibid*, para 23.

11 *Ibid. See, e.g.*, "The Commission proposes amendments to the New Media Exemption Order, prohibiting new media broadcasting undertakings from conferring an undue preference on themselves or another person, or subjecting any person to undue disadvantage. To provide guidance on the type of situation that could give rise to an undue preference in the new media environment, the Commission offers the example of a new media broadcasting undertaking engaged in programming distribution that acquires content from an affiliated programming undertaking either to the exclusion of non-affiliated programming undertakings or on more favourable terms or conditions than those applicable to non-affiliated programming undertakings."

12 Canadian Radio-television and Telecommunications Commission, *Broadcasting Regulatory Policy*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

government licensing."[13] Denton's comments foreshadowed much of the controversy over Bill C-10, which focused on the free speech implications of the bill.

In the years following, the rise of over-the-top (OTT) video providers such as Netflix was the cause of much consternation in the legacy broadcasting community. In 2011, a coalition of broadcasters, broadcast distributors (cable and satellite companies), and creators groups wrote to the CRTC to ask for a public consultation on foreign over-the-top services operating in Canada.[14]

The battle had been brewing for some time and what was particularly striking was how badly Canadian broadcasters and broadcast distributors understood the future impact of the Internet on their businesses. The prospect of the Internet becoming a substitute for conventional broadcast was not exactly a secret at the new media hearing in 2009. Yet, at the time, a representative from Shaw, a leading Canadian cable company, told the CRTC that the Internet was primarily for "self-generated content" and that it posed little threat to traditional cable broadcasters.[15] Similarly, in 2009 Bell Media, Canada's largest communications company, told the Commission that OTT providers "*may never become a substitute*" to cable offerings.[16] Despite their views that the Internet was no threat to smart business operators, Canadian broadcasters and broadcast distributors unanimously adopted the position that the CRTC should not establish new regulations for Internet-based broadcasting.[17]

---

13 Canadian Radio-television and Telecommunications Commission, *Broadcasting Regulatory Policy*, 12.

14 Canadian Radio-television and Telecommunications Commission, *Broadcasting and Telecom Notice of Consultation CRTC 2011-344* (Ottawa: Canadian Radio-television and Telecommunications Commission, 2011), https://crtc.gc.ca/eng/archive/2011/2011-344.pdf.

15 *Ibid.*

16 Mirko Bibic, *Transcript of Proceedings* (Quebec: Canadian Radio-Television and Telecommunications Commission, 2009), https://crtc.gc.ca/eng/transcripts/2009/tb0311.html.

17 Canadian Radio-Television and Telecommunications Commission, *Transcript of Proceedings*, (Quebec, Canadian Radio-Television and Telecommunications Commission, February 26, 2009), https://crtc.gc.ca/eng/transcripts/2009/tb0226.htm; *Transcript of Proceedings* (Quebec: Canadian Radio-Television and Telecommunications Commission, March 10, 2009), https://crtc.gc.ca/eng/transcripts/2009/tb0310.html; *Transcript of Proceedings* (Quebec: Canadian Radio-Television and Telecommunications Commission, March 11, 2009), https://crtc.gc.ca/eng/transcripts/2009/tb0311.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

A mere two years later, the perspective of the broadcasters shifted enormously. At a 2011 CRTC hearing, a representative from Bell Media called OTT providers "formidable competitors", and warned that Netflix would soon be able to "outbid Canadian broadcasters for exclusive program rights, both online and on television."[18] For its part, Shaw testified at a 2011 hearing, revising its 2009 assessment of the threat posed by OTT providers calling it "alarming" and the result of "major structural shifts in technology and rights exploitation that are permanently reshaping the global broadcast landscape".[19] In the span of two years, legacy broadcasters had gone from the position of minimizing the potential effects of OTT providers on their businesses and calling on the Commission to refrain from regulating Internet broadcasting, to demanding immediate action on Netflix.

Chapter 3. Change in Government, Change in Policy

A change in governments in 2015 heralded a different approach to digital policy under the Liberal government. In September 2016, newly-appointed Canadian Heritage Minister Mélanie Joly launched a consultation on supporting Canadian content in a digital world. In the "pre-consultation" phase – an online poll of the public and stakeholders – there were hints at the policy challenges that would be faced by the new government. The poll received more than 10,000 responses with participants asked to identify the major barriers and challenges for Canadian content. The perspective of the public and industry stakeholders were strikingly different, with the public citing the challenges in finding and promoting content and the stakeholders seeking more money.[20]

Once the consultation started in earnest, it sparked renewed demands from industry stakeholders for more money from two main sources: unregulated Internet companies such as Netflix and the government. As

18 Kevin Crull, *Transcript of Proceedings* (Quebec: Canadian Radio-Television and Telecommunications Commission, April 4, 2011), https://crtc.gc.ca/eng/transcripts/2011/tb0404.html, para 136.

19 Paul Robertson, *Transcript of Proceedings* (Quebec: Canadian Radio-Television and Telecommunications Commission, April 6, 2011), https://crtc.gc.ca/eng/transcripts/2011/tb0604.html, para 2543.

20 Ipsos, *What we heard across Canada: Canadian Culture in a Digital World,* (Ottawa: Ipsos Public Affairs, 2017), https://qcgn.ca/wp-content/uploads/2017/03/PCH-DigiCanCon-Consultation_Report-EN_low.pdf

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a starting position, the new consultation paper made it clear that not everything would be on the table. In fact, the consultation adopted several notable policies and sent some signals about future funding sources. First, it left little doubt that the government opposed new regulations for online video providers.[21] Strong support for net neutrality and the avoidance of Internet regulation meant that proposals to exempt Canadian content from data caps or mandate certain rules for online providers were non-starters. Second, the government used the consultation to suggest where more money may come from and it was not Canadian tax dollars.[22] By framing the consultation as an initiative that sat alongside already-announced funding, it seemed unlikely that more funding would be viewed as the answer. Indeed, the government was pretty clear about where it thought the money would come from: foreign markets.

Despite the direction provided in the consultation document, the government was less than clear in its communication on the issue of new taxes and regulations for online video providers. Joly appeared on a national television program after the policy launch and though she started by clearly stating that "there will be no new Netflix tax", the remainder of the interview was spent making the case for one.[23] From the interview, it seemed that Joly subscribed to the view that there was a parallel between conventional broadcast and the Internet that invited a similar regulatory approach. Part of the rationale for broadcast regulation is that broadcast spectrum is scarce, therefore requiring licensing and regulation. By indicating that Internet services used a "large part of our spectrum", Joly made the case for treating Internet services as equivalent to broadcast.[24]

Following the completion of the consultation, the government announced in its 2017 budget that it planned to "review and modernize"

---

21 Canada, Department of Canadian Heritage, *Canadian Content in a Digital World: Focusing the Conversation,* (Ottawa: Department of Canadian Heritage, 2016), 7. *"To respect how Canadians want to consume and interact with digital content, we are committed to net neutrality – the idea that a public information network like the internet is most useful if all content, sites, and platforms are treated equally. The way forward is not attempting to regulate content on the Internet, but focusing on how to best support Canada's creators and cultural entrepreneurs in creating great content and in competing globally for both Canadian and international audiences."*

22 Ibid, 4.

23 "GST on Netflix still a possibility as Liberals review cultural production," *CTV News*, October 16, 2016, https://www.ctvnews.ca/politics/gst-on-netflix-still-a-possibility-as-liberals-review-cultural-production-1.3115996.

24 Ibid.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the Broadcasting Act and Telecommunications Act.[25] The consultation had revealed there was a strong appetite within the traditional Canadian culture lobby for bringing policies such as cultural taxes and mandated Canadian content requirements to the Internet, with groups claiming the Internet was rapidly replacing the conventional broadcast system as a means of distributing cultural content and that the longstanding analog rules should be shifted into the digital environment. Revisiting Canada's twin communications laws was regarded by the cultural lobby as the opening to treat telecommunications regulation as a matter of cultural policy in what would amount to the Broadcasting Act taking over the Telecommunications Act.

In order to support the upcoming review of the Broadcasting Act and Telecommunications Act, the government asked the CRTC to become involved in developing policy. Through an Order-in-Council the government requested that the CRTC conduct a study on programming distribution models and their impact on maintaining a "vibrant domestic market."[26] The Commission was asked to address three main issues in its report: (1) the distribution model or models of programming that were likely to exist in the future; (2) how and through whom Canadians would access that programming and (3) the extent to which those models would ensure a vibrant domestic market capable of supporting the continued creation, production and distribution of Canadian programming, in both official languages, including original entertainment and information programming.

Joly formally unveiled her digital Canadian content strategy in September 2017, delivering a wide ranging plan that included a commitment from Netflix to spend $500 million over five years on production in Canada.[27] The Netflix commitment was the headline of the day, and represented a major long-term commitment to the Canadian market. However, since Canada was already one of the company's top three countries for pro-

25 Michael Geist, "Budget 2017: Why Canada's Digital Policy Future Is Up For Grabs," *Michael Geist* (blog), March 22, 2017, https://www.michaelgeist.ca/2017/03/budget-2017-canadas-digital-policy-future-grabs/.

26 Canadian Radio-television and Telecommunications Commission, *Broadcasting Notice of Consultation CRTC 2017-359*, (Ottawa: Canadian Radio-television and Telecommunications Commission, 2017), https://crtc.gc.ca/eng/archive/2017/2017-359.htm.

27 Canadian Heritage, *Minister Joly Announces Creative Canada: A Vision for Canada's Creative Industries in the Digital Age* (Ottawa: Canadian Heritage, 2017), https://www.canada.ca/en/canadian-heritage/news/2017/09/minister_joly_announcescreativecanadaavisionforcanadascreativein.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

duction, it was unlikely the announcement would result in a significant increase in funding.

While the Netflix commitment attracted attention, the more important story was that the government had rejected pressures to levy new Internet or Netflix taxes, impose regulatory requirements on Internet services, or depart from its commitment to net neutrality. Indeed, Joly's comments on the importance of affordable Internet access and support for net neutrality effectively slammed the door shut on those proposals. Joly started the consultation by indicating that everything was on the table, which many cultural lobby groups hoped would lead to new Internet taxes and regulation. The decision to reject those proposals confirmed that the government's digital focus emphasized competition, a strong domestic market, as well as export and promotion of Canadian content.

Chapter 4. A Shift in Approach: Harnessing Change

The government's approach to regulating online video providers began to change after the release of the CRTC's report on programming distribution. In June 2018, the Commission released "Harnessing Change: The Future of Programming Distribution in Canada",[28] in which it jumped into the Internet regulation and taxation game with both feet. Work that had preceded the Commission's report, including Joly's Digital Cancon strategy[29] as well as the Commission's own Let's Talk TV report[30] had emphasized the benefits of the Internet and sided primarily with an export-oriented, competition focused strategy in which Canadian content and broadcasters would succeed based on the quality of their programming, not regulatory schemes designed to provide millions of dollars in support.

In *Harnessing Change*, the CRTC reversed that approach with a regulation-first strategy that envisioned new fees attached to virtually anything

---

28 Canadian Radio-television and Telecommunications Commission, *Harnessing Change: The Future of Programming Distribution in Canada* (Ottawa: Canadian Radio-television and Telecommunications Commission, 2018), https://crtc.gc.ca/eng/publications/s15/.

29 Daniel Leblanc, "Everything's on the table," *Globe and Mail*, April 23, 2016, https://www.theglobeandmail.com/news/national/exclusive-canadian-heritage-announces-sweeping-canconreview/article29722581/.

30 Canada Radio-television and Telecommunications Commission, *Let's Talk TV: A Conversation with Canadians* (Ottawa: Canadian-Radio-television and Telecommunications Commission, 2018), https://crtc.gc.ca/eng/talktv-parlonstele.htm.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

related to the Internet: Internet service providers, Internet video services, and Internet audio services (wherever located) to name a few. The CRTC's report was provided to the government, but was accompanied by the feeling of theatre, with a review of telecom and broadcast legislation set to get underway that was to be led by a panel that included several proponents of an Internet regulation strategy.

Following the CRTC's report, the Minister of Innovation, Science and Economic Development tasked the Broadcasting and Telecommunications Legislative Review (BTLR) panel with a review of Canada's communications legislative framework.[31] In September 2018, the panel opened a call for comments, through which industry stakeholders, civil society, academics and individuals provided their perspectives on the future of Canada's Broadcasting and Telecommunications acts. When the consultation closed in January 2019 thousands of submissions had been made.[32] In June 2019, when the interim report was released alongside the written submissions, it began to look increasingly likely that the government had already decided what direction it intended to take.[33]

Canadian Heritage Minister Pablo Rodriguez, who had taken over the file from Joly, signalled that the government's position on the major broadcasting and Canadian cultural issue was already set. For months, government officials had been arguing that large Internet companies needed to contribute to Canadian content creation, though it had avoided specifying precisely how. With an election weeks away, the government position seemed to be shifting. Soon after the release of the BTLR interim report, Rodriguez tweeted that the government was ready to legislate once receiving the panel's final recommendations, but followed the statement by saying that "[e]veryone has to contribute to our culture. That's why

31 Innovation, Science and Economic Development Canada and Canadian Heritage, *Government of Canada launches review of Telecommunications and Broadcasting Acts* (Ottawa: Canadian Heritage, 2018), https://www.canada.ca/en/canadian-heritage/news/2018/06/government-of-canada-launches-review-of-telecommunications-and-broadcasting-acts.html.

32 Michael Geist, "Sunlight on the Submissions: Why the Broadcasting and Telecommunications Legislative Review Panel Should Reverse Its Secretive Approach," *Michael Geist* (blog), January 18, 2019, https://www.michaelgeist.ca/2019/01/sunlight-on-the-submissions-why-the-btlr-should-reverse-its-secretive-approach/.

33 Innovation, Science and Economic Development Canada, *What we Heard Report* (Ottawa: Canada-Radio-television and Telecommunications Commission, 2019), http://www.ic.gc.ca/eic/site/110.nsf/eng/00011.html#s8

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

we'll require web giants to create Canadian content and promote it on their platforms."[34]

By suggesting that the Liberals were ready to commit to legislative reform that would require Internet companies to create and promote Canadian content, the government had seemingly shifted its policy approach well ahead of the final BTLR report.

Chapter 5. BTLR report

In January 2020, the Broadcast and Telecommunications Legislative Review Panel released its much anticipated report with a vision of a highly regulated Internet in which an expanded CRTC (or a renamed Canadian Communications Commission) would aggressively assert its jurisdictional power over Internet sites and services worldwide with the power to levy penalties for failure to comply with its regulatory edicts.[35]

The foundation of the content section of the report was the decision to regulate all media content, which includes audio, audiovisual, and news content delivered by telecom. In doing so, the report envisioned unprecedented government and regulatory intervention into the delivery of news services. It argued that there are three types of services that provide this content that require regulation where they access the Canadian market:

- **Curators** – services that disseminate media content with editorial control (broadcasters and streaming services such as Netflix, Spotify, and Amazon Prime)
- **Aggregators** – cable companies, news aggregators such as Yahoo News
- **Platforms for Sharing** – services that allow users to share amateur and professional content such as YouTube, Facebook and other platforms

The panel recommended that all of these kinds of companies be regulated (either by way of licence or registration), be required to contribute to Canadian content through spending percentages or levies, and comply

---

34 Pablo Rodriguez (@pablorodriguez), "Thanks to @JanetYale1 & panel for their work. We will be ready to legislate once we receive their recommendations," Twitter, June 26, 2019, 11:38 a.m., https://twitter.com/pablorodriguez/status/1143906301002620928.

35 Canada, Broadcasting and Telecommunications Legislative Review, *Canada's Communications Future: Time to Act* (Ottawa: Innovation, Science and Economic Development Canada, 2020), https://www.ic.gc.ca/eic/site/110.nsf/vwapj/BTLR_Eng-V3.pdf/$file/BTLR_Eng-V3.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

with CRTC regulations on discoverability that would include regulatory rules on how prominently Canadian content is displayed within the service. The CRTC would be empowered to decide whether to exempt services from regulation with the power to levy penalties for failure to comply with its decisions (penalties described as "high enough to create a deterrent foreign undertakings").

Services would also be required to disclose consumption data to the CRTC, so that the regulator would know what Canadians are watching or reading online. The regulator would be entitled to establish binding codes of conduct that cover resolution mechanisms, transparency, privacy and accessibility. It would also govern the commercial relationship between services and content producers, with the panel noting "it is essential that the CRTC be given the explicit jurisdiction to regulate the economic relationships between media content undertakings and content producers, as well as between media content undertakings."[36]

The basis for the most sweeping reforms were framed as a matter of cultural sovereignty, with the panel arguing for the need for Canada "to continue to assert its cultural sovereignty and for Canadians to continue to express their identity and culture through content."[37] However, at a press conference following the report's release, both chair Janet Yale and panelist Monique Simard instead emphasized the need to support Canadian jobs when asked to reconcile the industry data that confirms record amounts of film and television production in Canada.[38]

Alternatively, the panel argued that it was simply a matter of those that benefit from the "system", being required to contribute to it.[39] However, broadcasters and broadcast distributors already enjoyed a wide range of regulatory benefits in the system and their contributions were essentially a regulatory quid pro quo.

With respect to Canadian content, the panel acknowledged that "Canadians create and consume more types of content than ever before," indicat-

36 Broadcasting and Telecommunications Legislative Review, *Canada's Communications*, 144.

37 Ibid. at 117.

38 Headline Politics, "Broadcasting & Telecommunications Legislative Review Panel Releases Final Report," *CPAC*, January 29, 2020, https://www.youtube.com/watch?v=pYNpa04S4C0

39 Tony Wong, "Netflix should pay sales taxes, CBC should be ad free, communications panel recommends," *Toronto Star*, January 30, 2020, thestar.com/entertainment/2020/01/29/netflix-and-other-online-streaming-sites-must-contribute-to-canadian-culture-and-content-says-legislative-review-panel.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ing that its recommendations weren't about incentivizing the creation of Canadian content.[40] Rather, the report seemed focused on certain professions creating content and imposing a massive regulatory infrastructure in order to support that policy goal. The problem with this approach was that ticking the right boxes to ensure Canadians represent "key creative personnel" had little to do with Canadian cultural sovereignty, much less ensuring access to Canadian stories. Yet while the panel emphasized "the importance of story", when confronted with the question of whether current Canadian content rules achieve that objective, it stated "it is time to review the model for supporting Canadian content, but not the definition of Canadian content."[41] The panel was prepared to overhaul the regulatory rules for creating and delivering Canadian content, but not consider the rules that determine what qualifies as Canadian content.

Chapter 6. The Government Responds to the Yale Report: Bill C-10

Taking up the recommendations from the BTLR panel report, in November 2020 Canadian Heritage Minister Steven Guilbeault tabled an Internet regulation bill with the express aim to "get money from web giants".[42] As expected, Bill C-10 handed massive new powers to the CRTC to regulate online streaming services and opens the door to mandated Canadian content payments, discoverability requirements, and confidential information disclosures all backed by new fining powers. Given that many of the details will be sorted out by the CRTC, the specifics would have taken years to unfold had the bill become law.

Chapter 7. Responding to a fictional content crisis

In part, Bill C-10 responded to a fictional Canadian content "crisis." Canadian cultural lobby groups regularly claim that the sector is at risk.[43] Yet

---

40 Canada, Broadcasting and Telecommunications Legislative Review, *Canada's Communications Future,* 117.

41 Ibid.

42 As of August 2021, the bill had been passed by the House of Commons, but was subject to review within the Canadian Senate. With a national election call, the bill died on the order paper and may only be re-introduced by a new government.

43 "Bill on Broadcasting Act: CDCE welcomes the decisive changes for our cultural sovereignty," *CDCE*, November 3, 2020, https://cdec-cdce.org/en/publications/r

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the reality is that spending on film and television production in Canada was at record highs. This included both certified Canadian content and so-called foreign location and service production in which the production takes place in Canada (thereby facilitating significant economic benefits) but does not meet the narrow criteria to qualify as "Canadian." The overall financing picture showed an industry that had record amounts of investment in film and television production with the total amount nearly doubling over the prior decade. Further, certified Cancon had also grown in recent years, with the top two years for certified Cancon television production occurring over the prior three years. In fact, 2019 was the biggest year for the production of French language Cancon over the prior decade.[44]

The data at the provincial level provided further confirmation of record-setting production. In February 2020, Ontario Creates, the Government of Ontario's agency for cultural creation, touted a "record breaking year" for Ontario's film and television production sector, citing more than $2 billion in production spending for 343 productions.[45] Of the $2.1 billion, there was a near-even split between domestic and foreign production: $1.1 billion in foreign production and $1 billion on domestic productions. In further support, Carleton professor Dwayne Winseck's 2020 review of the state of the network economy in Canada also found that film and television production investment in Canada had continuously increased for two decades, most recently "driven by massive investments from streaming services such as Netflix and Amazon Prime."[46] Politicians and regulators knew this to be the case. In fact, CRTC chair Ian Scott described Netflix as "probably the biggest single contributor to the [Canadian] production

---

elease-bill-on-broadcasting-act/; Maxime-Pierre Gazeau, "Bill C-10 to Extend the Broadcasting Act to Webcasters," *Artisti*, November 11, 2020, https://www.artisti.ca/en/bill-c-10-to-extend-the-broadcasting-act-to-webcasters/.

44 CMPA, *Profile 2019: Economic Report on the Screen-Based Media Production Industry in Canada*, (Ottawa: CMPA, 2019), Exhibit 1-1, Pg. 7, https://cmpa.ca/wp-content/uploads/2020/04/CMPA_2019_E_FINAL.pdf.

45 Michael Geist, "Ontario's Record Breaking, Multi-Billion Dollar Film Production Year: 'A Healthy Balance Between Domestic and Foreign Production'," *Michael Geist* (blog), March 4, 2020, https://www.michaelgeist.ca/2020/03/ontarios-record-breaking-multi-billion-dollar-film-production-year-a-healthy-balance-between-domestic-and-foreign-production/.

46 Winseck, Dwayne, *Growth and Upheaval in the Network Media Economy in Canada, 1984-2019*, (Ottawa, Canadian Media Concentration Research Project, Carleton University, 2020), 45, http://www.cmcrp.org/wp-content/uploads/2020/11/Growth-Report-2020-11162020v2.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

sector today."[47] Further, at the press conference introducing the bill, Guilbeault acknowledged that the Internet companies are already investing in Canada, but argued that the bill was needed to ensure those investments were not voluntary.[48]

Chapter 8. The myth of the level playing field

A central part of Guilbeault's argument for Bill C-10 was that it levels the playing field between traditional and online broadcasters. It is true that conventional broadcasters and broadcast distributors face mandated payments to support Canadian content as part of their licensing requirements. Leaving aside the fact that broadcasters were seeking reductions in payments at the CRTC,[49] the notion that the only regulatory burden or benefit is mandated Cancon contributions misreads the law. The reality is that broadcasters receive benefits worth hundreds of millions of dollars in return for those payments as part of what amounts to a regulatory quid pro quo. None of those benefits are available to Internet streaming services, yet the "level the playing field" discussion focused exclusively on equivalent payment requirements.

Some of the regulatory and policy benefits enjoyed by traditional broadcasters and broadcast distributors not available to Internet streaming services included:

1. **Simultaneous Substitution Policies**, which allow Canadian broadcasters to replace foreign signals with their own. The industry says this policy alone generates hundreds of millions of dollars in revenues for

47 Terry Pedwell, "Streaming companies like Netflix will have to fund Canadian content: CRTC chair," *National Post*, January 8, 2020, https://nationalpost.com/pmn/news-pmn/canada-news-pmn/streaming-companies-like-netflix-will-have-to-fund-canadian-content-crtc-chair.

48 "Heritage minister discusses bill to update Broadcasting Act – November 3, 2020," *CPAC*, 2020, YouTube Video, https://www.youtube.com/watch?v=qkV1Wp4JduU.

49 Canadian Radio-television and Telecommunications Commission, *Broadcasting Notice of Consultation CRTC 2020-336* (Ottawa: Canadian Radio-television and Telecommunications Commission, 2020), https://crtc.gc.ca/eng/archive/2020/2020-336.htm.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Canadian broadcasters.[50] There is no equivalent to the hundreds of millions generated by this policy for Internet streaming services.

2. **Must-Carry Regulations**, which require broadcast distributors to include many Canadian channels on basic cable and satellite packages.[51] These rules provide guaranteed access to millions of subscribers, thereby increasing the value of the signals and the fees that can be charged for their distribution. Internet streamers compete for subscribers with no guaranteed access.
3. **Copyright Retransmission Rules**, which create an exemption in the Copyright Act to allow broadcast distributors to retransmit signals without infringing copyright.[52] This retransmission occurred for many years without any compensation. There is no equivalent for Internet streamers.
4. **Bundling Benefits**, which allow broadcast distributors to bundle less popular Canadian channels with more popular U.S. signals, thereby guaranteeing more revenues to the Canadian broadcasters.[53] There is no equivalent for Internet streamers.
5. **Market Protection**, which shielded Canadian broadcasters from foreign competition such as HBO or ESPN for decades.[54] Internet streamers compete for subscribers with no market protections and the prospect of users unsubscribing at any time.
6. **Foreign Investment Restrictions**, which limits the percentage that foreign companies may own of Canadian broadcasters or broadcast distributors, which has the effect of creating a protected marketplace with reduced competition.

---

50 Christine Dobby, "Bell launches new appeal of CRTC's Super Bowl ad policy," *The Globe and Mail*, December 28, 2016, https://www.theglobeandmail.com/report-on-business/nfl-hopes-trudeau-government-will-overturn-crtc-super-bowl-ad-ruling/article33442315/.

51 Michael Geist, "The Broadcasting Act Blunder, Day 2: What the Government Doesn't Say About Creating a 'Level Playing Field'," *Michael Geist* (blog), November 20, 2020, https://www.michaelgeist.ca/2020/11/the-broadcasting-act-blunder-day-two-what-the-government-doesnt-say-about-creating-a-level-playing-field/#:~:text=Simultaneous%20Substitution%20policies%2C%20which%20allows%20Canadian%20broadcasters%20to,millions%20of%20dollars%20in%20revenues%20for%20Canadian%20broadcasters.

52 Ibid.

53 Ibid.

54 Ibid.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

7. **Eligibility for Canadian Funding Programs**, which are available to Canadian entities to support content creation but may be unavailable to foreign entities such as the Internet streamers.[55]
8. **Unlimited Distribution Without Caps or Usage Charges**, unlike Internet-based services, whose subscribers often face high data costs for accessing those services.
9. **Intellectual Property Preferences**, which requires that producers be Canadian in order to be certified as Cancon.[56] This leads to rules that preclude foreign companies from producing Cancon and requiring domestic IP ownership. As a result, Internet streamers are excluded from accessing the same funding available to Canadian producers.
10. **Trade Agreement Protections**, which exempt the Canadian government from treating foreign providers in the culture sector in the same manner as domestic firms.[57] While this provision is subject to potential tariff retaliation (as will be discussed later in the series), it means that standard practices regarding equal treatment do not apply to Internet streamers.

Chapter 9. Missing economic thresholds

Guilbeault also tried to assure the House of Commons that the bill featured several "guardrails" against over-broad regulation. In particular, he stated that online entities would need to reach an economic threshold before being subject to any regulation.[58] However, there was no specific economic threshold established by the bill. The starting point was that all Internet streaming services carried on in whole or in part within Canada are subject to Canadian regulation.

Guilbeault was presumably referring to the fact that section 6(4) of the bill gave the CRTC the power to exempt services from regulation.[59]

---

55 Ibid.
56 Ibid.
57 Ibid.
58 Guilbeault, Steven, *House of Commons Debates Canada* (Ottawa: House of Commons, 2020), https://www.ourcommons.ca/DocumentViewer/en/43-2/house/sitting-31/hansard, Para 1640.
59 House of Commons of Canada, *An Act to amend the Broadcasting Act and to make consequential amendments to other Acts*, Bill C-10, 2nd sess, introduced in House November 3, 2020, https://www.parl.ca/LegisInfo/BillDetails.aspx?Language=E&billId=10926636.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

While the CRTC could certainly establish some thresholds for regulation following the enactment of the bill, the approval of a policy directive, and a full hearing on the implementation issues, the possibility that the CRTC could create thresholds is not the same as claiming that the law contains significant economic thresholds. In fact, it is likely that the CRTC would not limit the regulatory model to "companies that generate large revenues in Canada", whatever that means. In order for the CRTC to determine who might be exempt, it was likely to require even smaller foreign services to register with the regulator and to provide it with confidential subscriber and revenue data.

The uncertainty of who is caught by the regulation was sure to have an impact on the market. Internet streaming services thinking about the Canadian market might put those plans on hold until they have some visibility over what they face from a regulatory perspective, leading to less competition and less choice for Canadians. Should the CRTC establish an economic threshold, that too could have an unexpected impact. If it set a high threshold that is limited to a handful of large, U.S.-based streaming services, it invited the possibility of a trade challenge. If a low threshold becomes the standard, foreign services may avoid the Canadian market altogether given the regulatory costs.

Chapter 10. Removing Canadian ownership requirements

One of the more controversial aspects of Bill C-10 proved to be the decision to remove the very first policy declaration in the Broadcasting Act as found in Section 3(1)(a): "the Canadian broadcasting system shall be effectively owned and controlled by Canadians." For years, Canada has prioritized a Canadian broadcast system with Canadian ownership requirements and Canadian content rules. With Bill C-10, the government signalled that it believed the benefits that come from mandatory contributions from foreign companies (bearing in mind that the companies voluntarily invest in the market) were worth sacrificing the longstanding policy of keeping the Canadian system Canadian.

---

(4) The Commission shall, by order, on the terms and conditions that it considers appropriate, exempt persons who carry on broadcasting undertakings of any class specified in the order from any or all of the requirements of this Part, of an order made under section 9.1 or of a regulation made under this Part if the Commission is satisfied that compliance with those requirements will not contribute in a material manner to the implementation of the broadcasting policy set out in subsection 3(1).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Guilbeault was asked about the ownership during the first day of House of Commons debate on the bill. He responded that the amendment to section 3(1) was necessary in order to allow the government to collect money from "web giants".[60] Guilbeault was right that Canada cannot have it both ways. It cannot argue that foreign companies must be part of – and contribute to – the Canadian system and then also argue that the system must be owned and controlled by Canadians. Either foreign companies are part of the system or they are not.

Guilbeault' proposed solution was to remove the policy of Canadian ownership and control, but use licensing to ensure that Canadian companies retain that same control. Indeed, many countries have removed foreign ownership requirements given the lack of a link between domestic content requirements and domestic ownership. Guilbeault therefore said the removal of Section 3(1)(a) was immaterial since licensing requirements would still apply to broadcasters and could be used to ensure that they remain in Canadian hands. Yet the obvious trajectory of the new Canadian system is to shift away from that licensing system. The government claimed it is creating a level playing field, but broadcasters in the licensed world would increasingly look at the unlicensed Internet world that is free from foreign investment restrictions and conclude that they prefer the unlicensed system.

The issue could become particularly acute if Canadian broadcasters are forced to compete with companies like Netflix and Disney for Canadian content as all participants race to meet their regulatory Cancon requirements. The disadvantages of remaining Canadian-owned would become increasingly apparent as more broadcasters surrender their licences in favour of switching to streaming-only services that remain unlicensed and have the advantage of no foreign ownership limitations. The Canadian market would feature an increasingly prominent foreign ownership presence, not only in the form of foreign streamers but also Canadian-originated streamers that become foreign-controlled through new investment.

Chapter 11. Discoverability requirements

Among the issues that Bill C-10 was intended to remedy, Guilbeault cited the need to improve the "discoverability" of Canadian content. Under section 9.1(1) the Bill permits the CRTC to make orders, including those

60 Canada, *House of Commons Debates,* Para 1645.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

with respect to program presentation and discoverability. The term "discoverability" does not appear elsewhere in the bill and is not defined. It would therefore fall to the CRTC to decide what it means and what conditions are imposed on Internet services as a result. Based on the Canadian cultural debate of the past few years, it would be expected that the CRTC would be urged to require services such as Netflix or Disney+ to override their algorithms that identify what subscribers are likely to want to see by actively promoting Canadian content regardless of their preferred content.

The BTLR panel, which recommended discoverability regulations, went looking for evidence of a discoverability problem and found very little. That report identified just two sources: a 2017 PriceWaterhouseCoopers report[61] and a 2016 report from Telefilm Canada.[62] The PriceWaterhouseCoopers report involved a survey of 1,000 U.S. residents, had nothing to do with Canada, and said absolutely nothing about the ability to find or recognize Canadian content. The Telefilm Canada report was focused on Canada but did not find that Canadians have trouble finding Canadian content. Rather, it found a range of experiences and emphasized that "word-of-mouth is Canadians' main discoverability method." Two reports – one from the U.S. and the other four years old – do not make the case for new regulations requiring the CRTC to regulate the way online services make their content available to subscribers in Canada.

Chapter 12. Downgrading the Role of Canadians in their Own Programming

One of the benefits of Bill C-10 touted by Guilbeault was that it was a big win for Canadian creators. Section 3(1)(f) of the current Broadcasting Act features the policy on use of Canadian creative talent, saying that "each broadcasting undertaking shall make *maximum use, and in no case less than predominant use*, of Canadian creative and other resources in the creation

61 Mark McCaffrey, Paige Hayes and Jason Wagner, *Can you find that show I didn't know I wanted to watch?: How tech will transform content discovery*, (London: PriceWaterhouseCoopers, 2017) https://gsma.force.com/mwcoem/servlet/servlet.FileDownload?file=00P1r00001kQ5tHEAS

62 The Telefilm Canada report is incorrectly cited as a 2018 report but actually dates to 2016:
Telefilm Canada, *Discoverability: Toward a Common Frame of Reference: Part 2: The Audience Journey*, (Montreal: Telefilm Canada, 2016) https://telefilm.ca/en/studies/discoverability-toward-common-frame-reference-part-2-audience-journey.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

and presentation of programming". Bill C-10 dropped the expectation of maximum or predominant use. The policy provision instead would state:

> *each broadcasting undertaking shall make use of Canadian creative and other resources in the creation and presentation of programming to the extent that is appropriate for the nature of the undertaking*

No one knew what that means since it would fall to the CRTC to determine what is "appropriate given the nature of the undertaking". Presumably the change was needed given the expansive regulatory approach taken by Bill C-10. Since the bill effectively captured foreign streaming sites both big and small, news sites, and podcasters, the government apparently felt that it could no longer require predominantly Canadian creative talent or even meet "the greatest practicable use of those resources" standard.

Chapter 13. The "Regulate Everything" Approach

The government was careful to note that it was not creating a new licensing system for Internet services with Bill C-10. For example, the Canadian Heritage FAQ stated "Canadians will still be able to watch all of their favourite programs and access their preferred services. This Bill in no way prevents online streaming services from operating in Canada, or *requires them to be licensed*."[63]

Bill C-10 was clear that in contrast to conventional broadcasters, online undertakings such as Internet streaming services would not require a licence to operate in Canada. While conventional broadcast undertakings (ie. programming undertakings) require either a licence or an exemption from the CRTC, online undertakings do not require either. Yet given the regulatory requirements, the absence of a licence would mean little for services operating in Canada, thinking about operating in Canada, or simply having Canadian users. For them, Bill C-10 provided a whole new regime that replaced licensing or exemption with "registration" subject to "conditions".

Bill C-10 created this new regime through amendments to sections 9, 10 and 11 of the Act. These new powers would allow the CRTC to:

---

63 Government of Canada, *Frequently asked questions – Modernizing the Broadcasting Act for the Digital Age* (Ottawa: Government of Canada, 2021), https://www.canada.ca/en/canadian-heritage/services/modernization-broadcasting-act/faq.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

- require registration of any broadcasting undertaking (section 10(1)(i))
- impose, by order, conditions that are virtually indistinguishable from licensing requirements (s.9.1(1))
- implement a wide range of additional regulations (sections 10 and 11).

Section 10(1)(i) gave the CRTC the power to establish regulations that could require all broadcasting undertakings – including online undertakings – to register with the Commission. Given how broadly the bill defined the jurisdictional scope, this included smaller streaming services, video news sites, podcasters, or even user generated content sites that include anything other than user generated content. Unless the CRTC decided to establish new thresholds or exemptions, all of these sites and services were caught by the bill and subject to Canada's new registration requirement.

The regulatory power extended beyond registration requirements, however. The CRTC could establish registration fees (the bill limited the fees to the costs incurred by the Commission) as well as regulations on Canadian programming, advertising rules, and audit rules that would have allowed the CRTC to examine records and books of any registered entity. These were all regulations that specifically could have been targeted at online undertakings such as Internet-based services. To be clear, failure to comply with these regulations carried the possibility of stiff penalties.[64]

Further, Section 34.4 established the possibility of administrative monetary penalties (AMPs) for contravening these regulations that ran into the millions of dollars. So while the government argued that it was not licensing Internet services, it created a regulation system that included registration, mandated audits, and Cancon conditions all backed by millions in potential penalties for failure to comply.

But Bill C-10 went beyond those regulatory requirements. Section 9.1 (1) featured numerous conditions that could have been imposed on any broadcast undertaking – including online undertakings. In addition the aforementioned discoverability conditions, the CRTC could have imposed conditions related to:

---

64 House of Commons of Canada, *An Act to amend the Broadcasting Act,* s.33. Section 33 provides: *Every person who contravenes any regulation or order made under this Part is guilty of an offence punishable on summary conviction and is liable (a) in the case of an individual, to a fine of not more than $25,000 for a first offence and of not more than $50,000 for each subsequent offence; or (b) in the case of a corporation, to a fine of not more than $250,000 for a first offence and of not more than $500,000 for each subsequent offence.*

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

- the proportion of programs to be broadcast that are Canadian
- access by persons with disabilities to programming, including the identification, prevention and removal of barriers to such access
- the carriage of emergency messages
- providing the CRTC with information on ownership, governance and control of the services as well as any affiliates
- providing the CRTC with any other information it requires, including financial or commercial information, programming information, expenditure information, and any information related to the provision of broadcasting services

While these provisions may fit within a licensed, Canadian-only environment, the conditions could have been applied by the CRTC to foreign online services with no presence in Canada. In fact, without any economic thresholds in the bill, the starting point was that all services around the world were potentially covered by these conditions so long as they have some Canadian subscribers. The CRTC may have ultimately limited the reach of the rules following extensive hearings, but for services thinking about the Canadian market, the regulatory environment might well have been reason to block Canadian subscribers. Guilbeault claimed that Bill C-10 would not result in less consumer choice, yet the more likely outcome was a Canadian regulatory firewall that had new entrant streaming services thinking twice before entering the market.

While the CRTC would have been tasked with establishing the specifics, the bill was also notable in that it granted the Commission the power to target individual services or companies with unique or individualized requirements. In other words, rather than establishing a "level playing field", Guilbeault opened the door to multiple fields with individual companies potentially each facing their own specific requirements and conditions to operate in Canada.

The source of this targeted approach was Section 9.1 (2), which provides:

> *(2) An order made under this section may be made applicable to all persons carrying on broadcasting undertakings, to all persons carrying on broadcasting undertakings of any class established by the Commission in the order or to a particular person carrying on a broadcasting undertaking.*[65]

65 House of Commons of Canada, *An Act to amend the Broadcasting Act,* s. 9.1 (2).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The regulatory ability to single out individual services for specific conditions (as opposed to common rules for all) created significant regulatory uncertainty, invited the possibility of a trade challenge, could have sparked allegations of unfair treatment, and raised further doubts for potential entrants into the Canadian market. The government claimed that consumer choice would not be affected by Bill C-10, but the likely repercussions of its legislative proposal strongly suggested otherwise.

Chapter 14. Risk to Canadian Ownership of Intellectual Property

At a time when the government emphasized the importance of intellectual property, the bill opened the door to less Canadian control and ownership over its IP. There was no reference to intellectual property in the bill nor any discussion of it within Canadian Heritage's FAQ or departmental materials.[66] Other than a background document reference to IP that suggested it could be included in a policy direction to the CRTC, intellectual property was not prioritized in the bill. In fact, by mandating that foreign services pay to support Canadian content and claiming they should be treated as equivalent to Canadian services for regulatory purposes, the government placed policy measures designed to safeguard intellectual property at risk.

IP policy has long been viewed as an important part of Canadian content policy. A production can be certified as Canadian content either through access to tax credits and/or Canadian Media Fund subsidy, or by the CRTC. All three require Canadian ownership of IP. For example, tax credits favour Canadian copyright ownership with larger credits available under the Canadian Film or Video Production Tax Credit (which requires Canadian copyright ownership) than with Film or Video Production Services Tax Credit (which does not).

These policies have prioritized domestic IP ownership and precluded foreign companies from producing and owning fully-financed Canadian content. As a result, revivals of Canadian programs such as Trailer Park Boys (Netflix) or Kids in the Hall (Amazon) would not meet the qualification requirements as Cancon where those companies are the sole funders and producers. The problem with Bill C-10 was that since no production fully-financed and owned by a foreign entity can be certified as Canadian content and the government sought to mandate such financing, the Canadian content rules would have had to change. If those changes meant

66 Government of Canada, *Frequently asked questions*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

removing the IP ownership link between tax credits and subsidies, and well-financed foreign streamers were allowed to fully-finance and own Canadian content, they could easily outbid Canadian producers for the best content. The end result could be that the best Canadian IP is owned by foreign streaming services, not Canadians.

Chapter 15. Mandated Confidential Data Disclosures May Keep Companies Out of Canada

Bill C-10 established significant confidential data disclosure requirements as a condition that could be imposed on Internet services both big and small around the world.

Section 9.1(1)(j) gave the CRTC the power to set a requirement on all broadcast undertakings, including online undertakings, to provide information the Commission considered necessary for the administration of the Act[67], including:

I. *financial or commercial information*,
II. *information related to programming,*
III. *information related to expenditures made under section 11.1*,
IV. *information related to audience measurement, other than information that could identify any individual audience member, and*
V. *other information related to the provision of broadcasting services*

In other words, the CRTC could demand everything: financial data, programming data, expenditure information, audience measurement data, and anything else it deemed relevant. In many cases, this information is commercially sensitive, not publicly available, and not required by other regulators.

While the CRTC needs good data to make effective decisions, the broad approach to mandated confidential information disclosure carried some significant risks. As noted previously, the condition on information disclosure could be limited to specific companies. For example, the CRTC could require companies such as Netflix or YouTube to disclose detailed audience and algorithmic data, which is data that those companies have been reluctant to make available anywhere in the world.

Moreover, the disclosure requirements were likely to extend to a very broad range of services, many of which may have limited or little connec-

---

67 House of Commons of Canada, *An Act to amend the Broadcasting Act*, 9.1(1)(j).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tion to Canada. While the bill did not contain economic thresholds the CRTC could establish such thresholds after extensive hearings. If it did so, companies could have been required to provide the Commission with confidential subscribers and financial data as evidence that they qualify for an exemption. In other words, services of all sizes and from all over the world would find themselves caught by CRTC regulation and requirements to disclose their confidential data. Their response may well have been to give Canada a pass by actively blocking Canadian users to reduce the risk of regulation, thereby leaving consumers with less choice and competition.

Chapter 16. Mandated Payments Likely to Bring in Less Than the Government Claims

Guilbeault made mandated payments the centrepiece of his Bill C-10 strategy, claiming that this would result in a billion dollars a year by 2023 in new funding.[68] The mandatory payment system was established in Section 11.1(1) of the bill and left it open to the CRTC to decide precisely who contributes, how they contribute, and how much they contribute.

Yet despite the fact that the CRTC would determine actual amounts, Guilbeault still clearly had a number in mind given the claims of $1 billion in new revenues. In fact, the number was $830 million when the bill was launched,[69] but the Minister was soon claiming nearly a billion instead.[70] In fact, Guilbeault went even further in the House, suggesting "it is actually more than $1 billion, because if nothing is done by 2023, Canadian productions and Canadian artists will miss out on $1 billion."[71]

The claim appears to simply represent a rough estimate on Canadians revenues from services such as Netflix with mandated payments of about 30 percent of those revenues. In the case of Netflix, its publicly stated revenues for Canada in 2019 were $780 million in revenue during the first 9 months,[72] so about $975 million for the year. At 30 percent, Netflix contribution would be around $293 million or about 30 percent of Guil-

68 Canada, *House of Commons Debates*, para 1650.
69 Government of Canada, *Frequently asked questions*.
70 Canada, *House of Commons Debates,* para 1625.
71 Ibid., para 1650.
72 Kelly Townsend, "Netflix has earned $780M in Canadian revenue in 2019," *Playback,* December 17, 2019, https://playbackonline.ca/2019/12/17/netflix-has-earned-780m-in-canadian-revenue-in-2019/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

beault's projected billion dollars in 2023, a number that could grow as revenues climb.

That will sound tempting to many, but it isn't the entire story. In the case of Netflix, it committed in 2017 to spend $500 million on productions in Canada over the following five years.[73] One year later, the company said it was on track to exceed that commitment.[74] In other words, Netflix was already spending hundreds of millions of dollars on production in Canada. While it is uncertain how the CRTC would mandate spending, it seems likely that the lion share of spending would be re-allocated money, not new funding. The same would apply to many other services that are already producing in Canada with money being reallocated to meet the regulatory requirements. To suggest that this will mean one billion dollars per year in new funding is at best a stretch.

Chapter 17. Misleading Comparison to the European Union

Guilbeault regularly cited the situation in Europe as evidence that the concerns about how Bill C-10 was likely to increase costs for consumers and decrease choice were unfounded. For example, he told the House of Commons that "European Union has adopted new rules on streamers resulting in increased investment, jobs, choice of content and ability to assert one's own cultural sovereignty"[75] and told the media that the European Union has had a requirement since 2018 that 30% of Internet streaming services content must be European content without resulting in higher fees.[76]

Guilbeault's comparison of Bill C-10 to the situation in Europe was misleading at best. A closer look reveals that after 10 years of regulatory work, less than a handful of EU member states have actually implemented

---

73 Catherine Cullen, "Netflix to commit $500M over 5 years on new Canadian productions: sources," *CBC News,* September 27, 2017, https://www.cbc.ca/news/politics/netflix-canadian-content-broadcaster-1.4309381#:~:text=Politics-,Netflix%20to%20commit%20%24500M%20over%205%20years%20on%20new,productions%2C%20CBC%20News%20has%20learned.

74 Corie Wright, "A Busy First Year for Netflix Canada," *Netflix,* September 28, 2018, https://about.netflix.com/en/news/a-busy-first-year-for-netflix-canada.

75 Canada, *House of Commons Debates,* para 1635.

76 Alex Boutilier, "Liberals propose law forcing Netflix, Spotify and others to support Canadian content," *The Star,* November 3, 2020, https://www.thestar.com/politics/federal/2020/11/03/liberals-propose-law-forcing-netflix-spotify-and-others-to-support-canadian-content.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the rules. Those that have done so have opted for much lower obligations with payment requirements that are a fraction of what Guilbeault had in mind. Moreover, scale matters and attempts to compare quotas intended for a market of 450 million people and 28 countries to a single country of 38 million is apples and oranges.

The European Audiovisual Media Services Directive was passed by the EU Parliament and Council in November 2018[77] and features at least four elements that bear some similarity at first glance to Bill C-10:

1. The designation of social media platforms as video-sharing platforms. This bring companies like Facebook, Instagram, and YouTube into the same regulatory sphere as Netflix, Apple TV, Amazon Prime and other streaming sites.
2. The imposition of the obligation for all Video on Demand (VOD) services (i.e.: streaming services) to have at least 30% of their catalogue be European works. This means all streaming services operating across European countries must have at least 30% of their country-specific catalogue be European.
3. The 30% obligation is accompanied by a prominence requirement which mandates all VOD services to have an EU works section on their platform so European films and movies are easily discoverable by users.
4. The directive provides each member state the ability to require VOD service providers to invest in EU works. These funding requirements can be applied to service providers targeting audiences in a member state even when they are under the jurisdiction of another member state.

In other words, the directive includes content and discoverability requirements, but does not mandate a funding requirement. In that regard, it is different from Bill C-10, which emphasized funding over content requirements.

While Guilbeault suggested that the European directive has not had a negative effect on consumers, the reality is that few member states have actually implemented it despite an obligation to do so by September 2020.[78] In fact, even those that have implemented the directive have adopt-

77 Audiovisual and Media Services Directive, European Commission (2021), https://ec.europa.eu/digital-single-market/en/audiovisual-media-services-directive-avmsd.

78 Glenn Carstens Peters, "Member States fail to meet the Audiovisual Media Services Directive deadline," *Society of Audiovisual Authors* (blog), September 17, 2020, https://www.saa-authors.eu/en/blog/667-member-states-fail-to-meet-the-audiovisual-media-services-directive-deadline#.YGe2SmRKg3R.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ed differing approaches. For example, countries such as the Netherlands, Croatia, Poland, and Denmark have investment quotas of under 6%, a far cry from the 30% envisioned by Guilbeault.[79] Meanwhile, Germany is still debating the levy and Spain is thinking about a 5% requirement.[80] The overall approach to date suggests that the 30% payment requirement is dramatically out-of-step with what is found in Europe. Given the far higher payment requirements in Canada, the consumer implications would undoubtedly be far greater.

The content requirements are also an inapt comparison to Canada. The 30% requirement covers European content, not content from a single country. Given the size of the European market and the number of member states, the actual per country requirement is effectively just over 1% if divided evenly among the member states. The reality is that services will surely exceed that number locally since it is in their interests to do so in order to attract local customers. However, any attempt to compare a 30% requirement that draws on a population of approximately 450 million people and 28 member states with Canada just doesn't work.

Finally, consider how long the process in Europe has taken (and continues to take). While Guilbeault talked about a regulatory process concluding by the end of 2021, Europe has taken more than ten years to develop its rules and the majority of member states still have not implemented them at a domestic level. The European experience highlights that these are complex issues that require careful study, not a "trust us" approach that leaves most of the key issues to a policy directive or the regulator.

Chapter 18. Bill C-10 and the Regulation of User Generated Content

The public paid little attention to Bill C-10 for months after it was introduced. Indeed, by late April 2021, the bill had steadily and stealthily worked its way through the Parliamentary process with only a few hurdles left to clear before passing the House of Commons. However, the bill was suddenly thrust onto the front page of newspapers across the country toward the end of its review journey with the public seizing on an

79 Nick Vivarelli, "Europe's New Rules of Engagement With Streaming Making Slow But Steady Progress", Variety, March 5, 2021, https://variety.com/2021/digital/news/europe-avms-streamers-1234915013/.

80 Ibid.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

unexpected change that opened the door to government regulation of the Internet content posted by millions of Canadians.

The change involved the removal of a clause that exempted from regulation user generated content on social media services such as TikTok, Youtube, and Facebook. The government had maintained that it had no interest in regulating user generated content, but the policy reversal meant that millions of video, podcasts, and the other audiovisual content on those popular services would be treated as "programs" under Canadian law and subject to some of the same rules as those previously reserved for programming on conventional broadcast services. Indeed, when Guilbeault appeared before the Standing Committee on Canadian Heritage he was asked by Liberal MP Tim Louis about "misinformation that somehow this [Bill C-10] would control, or regulate, or censor social media." He responded:

> *In the case of YouTube, for example, we're not particularly interested in what people…you know, when my great-uncle posts pictures of his cats, that's not what we're interested in as a legislator. When YouTube or Facebook act as a broadcaster, then the legislation would apply to them and the CRTC would define how that would happen. But really, we're not interested in user-generated content. We are interested in what broadcasters are doing.*[81]

Guilbeault was referring to a specific exception in Bill C-10 that excluded user generated content from the scope of broadcast regulation. The provision stated:

> *This Act does not apply in respect of*
> *(a) programs that are uploaded to an online undertaking that provides a social media service by a user of the service – who is not the provider of the service or the provider's affiliate, or the agent or mandatary of either of them – for transmission over the Internet and reception by other users of the service; and*
> *(b) online undertakings whose broadcasting consists only of such programs.*[82]

---

81 Parliament, House of Commons, Canadian Heritage Committee, *Minutes of Proceedings 43rd Parliament, Meeting No 18* (Ottawa: Parliament, House of Commons, Canadian Heritage Committee, 2021), https://openparliament.ca/committees/canadian-heritage/43-2/18/steven-guilbeault-10/.

82 House of Commons of Canada, *An Act to amend the Broadcasting Act,* s.4.1.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Without this provision, anything uploaded by users – whether cat videos or kids dancing in the kitchen – would be treated by Canadian law as a "program" and subject to CRTC regulation. In fact, government officials confirmed that interpretation:

> *Ms. Dabrusin has signalled the government intends to repeal, or suggest a repeal, of Section 4.1 altogether, meaning that there would no longer be any exclusion for social media services at all. For the benefit of the committee, in our previous sessions, the committee upheld the exclusion for users of social media companies. In other words, when you or I upload something to YouTube or some other sharing service, we will not be considered broadcasters for the purposes of the Act. The CRTC couldn't call us before them and we couldn't be subject to CRTC hearings.*
> *But if the exclusion is removed – if 4.1 is struck down – the programming we upload to Youtube, that programming that we place on that service would be subject to regulation moving forward, but would be the responsibility of Youtube or whatever the sharing service is. The programming that is uploaded could be subject to discoverability requirements or certain obligations like that. If the way forward is to maintain the exclusion for individual users but to strike down the exclusion for social media companies, that means that all the programming that is on those services would be subject to the Act regardless of whether it was put there by an affiliate or a mandatary of the company.*[83]

The change in approach sparked widespread public concern with the main opposition party vowing to repeal the legislation if enacted. The government's initial response to the controversy focused on two issues: the constitutionality of the change and attempts to limit regulatory power over user generated content to data disclosures by Internet services and the previously discussed discoverability requirements.

With respect to compliance with the Canadian Charter of Rights and Freedoms, the government provided an updated Charter statement which shed little light on the concerns involving the regulation of user generated content as a "program" under the law, however. Instead, it simply emphasized that users are not regulated as broadcasters and the CRTC is required to rule in a manner consistent with the Charter.

---

83 ParlVu, "CHPC Meeting No. 26 – Standing Committee on Canadian Heritage," *ParlVu*, April 23, 2021, https://parlvu.parl.gc.ca/Harmony/en/PowerBrowser/PowerBrowserV2/20210423/-1/35243.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The government also sought to justify the broader regulatory scope by pointing to the need for discoverability requirements for user generated content. Yet that too faced public criticism. First, as noted above, there was little evidence supporting claims of a problem in discovering Canadian content. Second, beyond the ease with which Canadian content can be found on audio and video-on-demand services, critics noted no other country mandates domestic content requirements on a user generated content platforms. That includes the European Union approach, which explicitly treats audiovisual media services (such as Netflix) and video sharing platform services (such as Youtube) differently. Audiovisual media services that engage in curating content face content requirements similar to those found for conventional broadcasters. Video sharing platform services face rules with respect to removing certain illegal or harmful content, but there are no quotas or no positive obligations to prioritize some content over others.

Chapter 19. The Bill C-10 Endgame

Faced with ongoing opposition in the House of Commons and lengthy debates and delays during review of Bill C-10, the government ultimately joined forces with two smaller opposition parties to impose a process known as "time allocation", which limited the time available for further study to five hours. The process, which had not been used in more than twenty years for a committee study, was highly controversial with many noting that it tainted the legitimacy of the review process.

The committee was forced to comply with the time allocation order, however, leading to a rapid conclusion to the study of Bill C-10 with Members of Parliament voting on dozens of amendments that were not made public at the time nor subject to any debate. Yet days later, the Speaker of the House of Commons declared many amendments "null and void", forcing the government to re-introduce the amendments within the House of Commons. In order to pass the amendments and bring the debate on Bill C-10 to a close, the government passed multiple motions to cut short debate and ultimately passed the bill in the middle of the night when few Canadians were still awake.[84]

84 House of Commons Canada, *Vote No.* 174 (Ottawa: House of Commons, 2021), https://www.ourcommons.ca/members/en/votes/43/2/174.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

With only days remaining before the summer recess, the bill was sent to the Senate for review. The Senate Bill C-10 debate wrapped up with several speeches and a vote to send the bill to committee for further study. Given that the Senate declined to approve summer hearings for the bill, with the earliest possible time for the study to begin in late September 2021. Yet with the late summer election call, Bill C-10 died on the order paper before the Senate study could begin in earnest.

While the debate in the Senate was marked by consistent calls for more study, the final debate was punctuated by a powerful speech from Senator David Adams Richards. One of Canada's leading authors, Senator Richards has won the Governor General's Award for both fiction and non-fiction, the Giller Prize, and is a member of the Order of Canada. Senator Richards, appointed by Prime Minister Trudeau to the Senate in 2017, warns against government or cultural decision makers and the parallels to Bill C-10:

> *Some years ago, I was at a dinner with some very important, famous people. One academic mentioned that he had given his entire life for Canadian literature. Others there applauded him for doing so. When I was writing my fourth novel, we sold our 20-year-old car to pay the rent; and my wife, to keep us alive, was selling Amway door-to-door in the middle of winter. I believe she gave her life for Canadian literature as well, but she didn't get to that dinner. For that reason, in her honour, I will always and forever stand against any bill that subjects freedom of expression to the doldrums of governmental oversight, and I implore others to do the same. I don't think this bill needs amendments; I think, however, it needs a stake through the heart.*[85]

After months of Canadian Heritage Minister Steven Guilbeault invoking the names of cultural lobby groups as evidence of support for Bill C-10, it took one of Canada's most celebrated authors to set the record straight and bring the debate to a close. In doing so, Senator Richards placed the spotlight on the challenges of reshaping Canada's broadcasting laws and difficulty in striking a balance between modernized Internet rules and freedom of expression safeguards.

---

85 Senator David Adams Richards, Senate Hansard, *Senators' Statements* (Ottawa: Senate of Canada, 2021), https://sencanada.ca/en/content/sen/chamber/432/debates/056db_2021-06-29-e#35.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Bibliography

Audiovisual and Media Services Directive, European Commission (2021). https://ec.europa.eu/digital-single-market/en/audiovisual-media-services-directive-avmsd.

Bibic, Mirko. *Transcript of Proceedings.* Quebec: Canadian Radio-Television and Telecommunications Commission, 2009. https://crtc.gc.ca/eng/transcripts/2009/tb0311.html.

Bill C-10. *An Act to amend the Broadcasting Act and to make consequential amendments to other Acts*. 2nd sess. 43rd Parliament. 2020. https://www.parl.ca/LegisInfo/BillDetails.aspx?Language=E&billId=10926636.

Boutilier, Alex. "Liberals propose law forcing Netflix, Spotify and others to support Canadian content." *The Star*. November 3, 2020. https://www.thestar.com/politics/federal/2020/11/03/liberals-propose-law-forcing-netflix-spotify-and-others-to-support-canadian-content.html.

Canada. Broadcasting and Telecommunications Legislative Review. *Canada's Communications Future: Time to Act.* Ottawa: Innovation, Science and Economic Development Canada, 2020. https://www.ic.gc.ca/eic/site/110.nsf/vwapj/BTLR_Eng-V3.pdf/$file/BTLR_Eng-V3.pdf.

Canada. Canadian Radio-television and Telecommunications Commission. *Harnessing Change: The Future of Programming Distribution in Canada*. Ottawa: Canadian Radio-television and Telecommunications Commission, 2018. https://crtc.gc.ca/eng/publications/s15/.

Guilbeault, Steven. *House of Commons Debates.* Ottawa: House of Common Debates, 2020. https://www.ourcommons.ca/DocumentViewer/en/43-2/house/sitting-31/hansard.

Canadian Heritage, *Minister Joly Announces Creative Canada: A Vision for Canada's Creative Industries in the Digital Age*. Ottawa: Canadian Heritage, 2017. https://www.canada.ca/en/canadian-heritage/news/2017/09/minister_joly_announcescreativecanadaavisionforcanadascreativein.html.

Canada. Parliament. House of Commons. Canadian Heritage Committee. *Minutes of Proceedings.* 2nd sess. 43rd Parliament. Meeting No. 18, 2021. https://openparliament.ca/committees/canadian-heritage/43-2/18/steven-guilbeault-10/.

Canadian Radio-television and Telecommunications Commission. *Broadcasting Notice of Consultation CRTC 2020-336.* Ottawa: Canadian Radio-television and Telecommunications Commission, 2020. https://crtc.gc.ca/eng/archive/2020/2020-336.htm.

Canadian Radio-television and Telecommunications Commission. *Broadcasting and Telecom Notice of Consultation CRTC 2011-344.* Ottawa: Canadian Radio television and Telecommunications Commission, 2011. https://crtc.gc.ca/eng/archive/2011/2011-344.pdf.

Canadian Radio-television and Telecommunications Commission. *Broadcasting Regulatory Policy CRTC 2009-329.* Ottawa: Canadian Radio-television and Telecommunications Commission, 2009. https://crtc.gc.ca/eng/archive/2009/2009-329.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Canadian Radio-television and Telecommunications Commission. *Broadcasting Notice of Consultation CRTC 2017-359.* Ottawa: Canadian Radio-television and Telecommunications Commission, 2017. https://crtc.gc.ca/eng/archive/2017/2017-359.htm.

Canadian Radio-television and Telecommunications Commission. *Public Notice CRTC 1998-82 – New Media – Call for Comments.* Ottawa: Canadian Radio-television and Telecommunications Commission, 1998. https://crtc.gc.ca/eng/archive/1998/PB98-82.htm.

Canadian Radio-television and Telecommunications Commission. *Public Notice CRTC 1999-84 – Report on New Media.* Ottawa: Canadian Radio-television and Telecommunications Commission, 1998. https://crtc.gc.ca/eng/archive/1999/PB99-84.htm

Canadian Radio-television and Telecommunications Commission. *Transcript of Proceedings.* Ottawa: Canadian Radio-Television and Telecommunications Commission, 2009. https://crtc.gc.ca/eng/transcripts/2009/tb0226.htm.

Canadian Radio-television and Telecommunications Commission. *Transcript of Proceedings.* Ottawa: Canadian Radio-Television and Telecommunications Commission, 2009. https://crtc.gc.ca/eng/transcripts/2009/tb0310.html. https://crtc.gc.ca/eng/archive/1999/PB99-84.htm.

CMPA. *Profile 2019: Economic Report on the Screen-Based Media Production Industry in Canada*. Ottawa: CMPA, 2019. https://cmpa.ca/wp-content/uploads/2020/04/CMPA_2019_E_FINAL.pdf.

Crull, Kevin. *Transcript of Proceedings.* Quebec: Canadian Radio-Television and Telecommunications Commission, 2011. https://crtc.gc.ca/eng/transcripts/2011/tb0404.html.

Cullen, Catherine. "Netflix to commit $500m over 5 years on new Canadian productions: sources." *CBC News*, September 27, 2017. https://www.cbc.ca/news/politics/netflix-canadian-content-broadcaster-1.4309381#:~:text=Politics-,Netflix%20to%20commit%20%24500M%20over%205%20years%20on%20new,productions%2C%20CBC%20News%20has%20learned.

Department of Canadian Heritage. *Canadian Content in a Digital World: Focusing the Conversation.* Ottawa: Department of Canadian Heritage, 2016. https://www.canadiancontentconsultations.ca/system/documents/attachments/e328d01aaa5d8b25b5b2e769f0f3ccb59f63893e/000/004/022/original/PCH-DigiCanCon-Consultation_Paper.pdf.

Dobby, Christine. "Bell launches new appeal of CRTC's Super Bowl ad Policy." *The Globe and Mail*, December 28, 2016. https://www.theglobeandmail.com/report-on-business/nfl-hopes-trudeau-government-will-overturn-crtc-super-bowl-ad-ruling/article33442315/.

Gazeau, Maxime-Pierre. "Bill C-10 to Extend the Broadcasting Act to Webcasters." *Artisti*, November 11, 2020. https://www.artisti.ca/en/bill-c-10-to-extend-the-broadcasting-act-to-webcasters/.

Geist, Michael. "Budget 2017: Why Canada's Digital Policy Future is Up For Grabs." *Michael Geist* (blog). March 22, 2017. https://www.michaelgeist.ca/2017/03/budget-2017-canadas-digital-policy-future-grabs/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Geist, Michael. "Ontario's Record Breaking, Multi-Billion Dollar Film Production Year: 'A Healthy Balance Between Domestic and Foreign Production'." *Michael Geist* (blog). March 4, 2020. https://www.michaelgeist.ca/2020/03/ontarios-record-breaking-multi-billion-dollar-film-production-year-a-healthy-balance-between-domestic-and-foreign-production/.

Geist, Michael. "Sunlight on the Submissions: Why the Broadcasting and Telecommunications Legislative Review Panel Should Reverse Its Secretive Approach." *Michael Geist* (blog). January 18, 2018. https://www.michaelgeist.ca/2019/01/sunlight-on-the-submissions-why-the-btlr-should-reverse-its-secretive-approach/.

Geist, Michael. "The Broadcasting Act Blunder, Day 2: What the Government Doesn't Say About Creating a 'Level Playing Field'." *Michael Geist* (blog). November 20, 2020. https://www.michaelgeist.ca/2020/11/the-broadcasting-act-blunder-day-two-what-the-government-doesnt-say-about-creating-a-level-playing-field/#:~:text=Simultaneous%20Substitution%20policies%2C%20which%20allows%20Canadian%20broadcasters%20to,millions%20of%20dollars%20in%20revenues%20for%20Canadian%20broadcasters.

"GST on Netflix still a possibility as Liberals review cultural production." *CTV News,* October 16, 2016. https://www.ctvnews.ca/politics/gst-on-netflix-still-a-possibility-as-liberals-review-cultural-production-1.3115996.

Government of Canada. *Frequently asked questions – Modernizing the Broadcasting Act for the Digital Age*. Ottawa: Government of Canada, 2021. https://www.canada.ca/en/canadian-heritage/services/modernization-broadcasting-act/faq.html.

Headline Politics. "Broadcasting & Telecommunications Legislative Review Panel Releases Final Report." *CPAC*. January 29, 2020. https://www.cpac.ca/en/programs/headline-politics/episodes/66143249/#.

"Heritage minister discusses bill to update Broadcasting Act – November 3, 2020." *CPAC*. 2020. YouTube Video. https://www.youtube.com/watch?v=qkV1Wp4JduU.

House of Commons of Canada. *An Act to amend the Broadcasting Act and to make consequential amendments to other Acts, Bill C-10*. Ottawa: House of Commons of Canada, 2020. https://www.parl.ca/LegisInfo/BillDetails.aspx?Language=E&billId=10926636.

Innovation, Science and Economic Development Canada and Canadian Heritage. *Government of Canada launches review of Telecommunications and Broadcasting Acts*. Ottawa: Canadian Heritage, 2018. https://www.canada.ca/en/canadian-heritage/news/2018/06/government-of-canada-launches-review-of-telecommunications-and-broadcasting-acts.html.

Ipsos. *What we heard across Canada: Canadian Culture in a Digital World.* Ottawa: Ipsos Public Affairs, 2017. https://www.canadiancontentconsultations.ca/system/documents/attachments/7fbdb8859168fdacec048735532bfdf6c45789a0/000/005/630/original/PCH-DigiCanCon-Consultation_Report-EN_low.pdf

McCaffrey, Mark, Hayes, Paige and Jason Wagner. *Can you find that show I didn't know I wanted to watch? How tech will transform content discovery.* London: PriceWaterhouseCoopers, 2017. https://gsma.force.com/mwcoem/servlet/servlet.FileDownload?file=00P1r00001kQ5tHEAS.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ParlVu. "CHPC Meeting No. 26 – Standing Committee on Canadian Heritage." ParlVu. April 23, 2021. https://parlvu.parl.gc.ca/Harmony/en/PowerBrowser/PowerBrowserV2/20210423/-1/35243.

Pedwell, Terry. "Streaming companies like Netflix will have to fund Canadian content: CRTC chair." *National Post*. January 8, 2020. https://nationalpost.com/pmn/news-pmn/canada-news-pmn/streaming-companies-like-netflix-will-have-to-fund-canadian-content-crtc-chair.

Peters, Glenn Carstens. "Member States fail to meet the Audiovisual Media Services Directive deadline." *Society of Audiovisual Authors* (blog). September 17, 2020. https://www.saa-authors.eu/en/blog/667-member-states-fail-to-meet-the-audiovisual-media-services-directive-deadline#.YGe2SmRKg3R.

Robertsen, Paul. *Transcript of Proceedings.* Quebec: Canadian Radio-Television and Telecommunications Commission, 2011. https://crtc.gc.ca/eng/transcripts/2011/tb0604.html.

Rodriguez, Pablo (@pablorodriguez). "Thanks to @JanetYale1 & panel for their work. We will be ready to legislate once we receive their recommendations." Twitter. June 26, 2019. 11:38 a.m. https://twitter.com/pablorodriguez/status/1143906301002620928.

Senator David Adams Richards, Senate Hansard. *Senators' Statements*. Ottawa: Senate of Canada, 2021, https://sencanada.ca/en/content/sen/chamber/432/debates/056db_2021-06-29-e#35Telefilm Canada. *Discoverability: Toward a Common Frame of Reference: Part 2: The Audience Journey.* Montreal: Telefilm Canada, 2016. https://telefilm.ca/en/studies/discoverability-toward-common-frame-reference-part-2-audience-journey.

Townsend, Kelly. "Netflix has earned $780M in Canadian revenue in 2019." *Playback.* December 17, 2019. https://playbackonline.ca/2019/12/17/netflix-has-earned-780m-in-canadian-revenue-in-2019/.

Vivarelli, Nick. "Europe's New Rules of Engagement With Streaming Making Slow But Steady Progress", *Variety*, March 5, 2021, https://variety.com/2021/digital/news/europe-avms-streamers-1234915013/.

Winseck, Dwayne. *Growth and Upheaval in the Network Media Economy in Canada, 1984-2019.* Ottawa: Canadian Media Concentration Research Project, Carleton University, 2020. http://www.cmcrp.org/wp-content/uploads/2020/11/Growth-Report-2020-11162020v2.pdf.

Write, Corie. "A Busy First Year for Netflix Canada." *Netflix*. September 28, 2018. https://about.netflix.com/en/news/a-busy-first-year-for-netflix-canada.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The UK's Approach to Regulation of Digital Platforms

Lorna Woods

Abstract: This chapter provides an overview of three main strands of regulation in the UK that would affect the regulation of digital platforms in general and social media in particular: data protection; competition; and the online harms agenda. In doing so, it considers the extent to which existing powers have been used and the extent to which new regimes have been proposed or are required. All the regimes have a regulator and, despite potential overlap and tensions between regimes, a number of commonalities exist between them, notably the focus on the impact of design choices, and risk-based approaches to the applicability of the regimes. A further similarity is the question of whether the regulators have adequate powers and resources. A final theme is the response, particularly of large companies, to enforcement of regulation.

Keywords: data protection – competition – consumer protection – online harms – online safety – targeted advertising – age appropriate design code – digital markets unit

Chapter 1. Introduction

The last decade has seen the beginning of attempts to regulate online platforms, a trend which has picked up pace since the Cambridge Analytica scandal and other *causes célèbres*. This contribution outlines policy developments in the UK across three relevant policy fields: data protection; competition and consumer protection; and online harms. In so doing, it considers both the re-purposing of existing powers and the proposal of entirely new regimes. This paper will identify how existing regimes have been used; what new measures are proposed and where the legislative process currently sits in a policy environment dominated by Brexit and COVID-19.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 2. Data Protection

Although data protection has received a much higher profile[1] with the introduction of the General Data Protection Regulation (GDPR)[2] (implemented in the UK by the Data Protection Act 2018 (DPA18)), the fundamental principles of data protection have not changed radically from the previous regime (Data Protection Act 1998 (DPA98), implementing the Data Protection Directive[3]). It is enforced by the Information Commissioners Office (ICO), an independent regulatory authority.[4] While much of the ICO's enforcement activity has focused on inadequate security,[5] but beyond this there are three interconnected areas affecting online platforms: the Cambridge Analytica investigation (including the investigation regarding political campaigns); the investigation into the online advertising sector; and the Age Appropriate Design Code.

Chapter 3. Cambridge Analytica and the Use of Data for Political Purposes

The ICO commenced an investigation into the use of data analytics in political campaigning in the light of concerns about "invisible processing" and micro-targeting of political adverts[6] triggered by the Cambridge Analytica scandal. Cambridge Analytica, a political consultancy firm, combined data obtained from a quiz app with data obtained through Facebook's Graph API and other data sources to profile users in furtherance of its clients' objectives. Users of the quiz app were unaware of the data collection and use. The ICO investigation covered social media platforms, but

1 C. Sellars, "GDPR: one year on - ICO pulls back the curtain on the impact of the new regime", (2019) 25 *CTLR* 172, pp. 172 and 173.
2 Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ L119/1.
3 Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31.
4 The relationship with the Department of Digital Culture Media and Sport is set out in a management agreement between the ICO and DCMS, https://ico.org.uk/media/about-the-ico/documents/2259800/management-agreement-2018-2021.pdf.
5 A. Bevitt and A. Collins, "UK Enforcement: Five focus areas", (2020) 20 *Privacy and Data Protection* 10.
6 Select Committee on Digital Culture Media and Sport, *Disinformation and 'fake news': Final Report*, 18 February 2019, https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/179102.htm.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

also data brokers, analytics firms, political parties and campaign groups. Its report concluded that there were risks in relation to the processing of personal data by many political parties. Particular concerns included the purchasing of marketing lists and lifestyle information from data brokers without sufficient due diligence; a lack of fair processing; and the use of third party data analytics companies, with insufficient checks around consent.[7] Formal warnings were issued to 11 political parties, and a number of fines imposed (including one of £500,000 on Facebook[8] – the maximum allowable under the DPA98 which applied at the time the incidents occurred). The interim report[9] also contained assessment notices to the three main credit reference agencies – Experian, Equifax and Call Credit.[10] Experian's response was subsequently found to be insufficient, and the ICO issued an enforcement notice (but not yet a fine).[11] The investigation concluded that there had not been significant interference in

7 For the interplay between data protection rules and rules pertaining to electoral advertising, see B. Shiner, "Big data, small law: how gaps in regulation are affecting political campaigning methods and the need for fundamental reform", (2019) *Public Law* 362; concerns about micro-targeting have also been expressed at EU level: C. Wenn, "Can data protection solve the problem of microtargeting, manipulation of internet users and fake news?", (2018) 29 *Ent. LR* 216.

8 ICO, Press Release, ICO issues maximum £500,000 fine to Facebook for failing to protect users' personal information, https://ico.org.uk/facebook-fine-20181025#.

9 ICO, *Democracy disrupted? Personal information and political influence*, 11 July 2018, https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf; ICO, Investigation into the use of data analytics in political campaigns: Investigation update, 11 July 2018, https://ico.org.uk/media/action-weve-taken/2259371/investigation-into-data-analytics-for-political-purposes-update.pdf; and ICO *Investigation into the use of data analytics in political campaigns: A Report to Parliament*, 6 November 2018, https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf.

10 See report of the investigation: ICO, Investigation into data protection compliance in the direct marketing data broking sector, October 2020, https://ico.org.uk/media/action-weve-taken/2618470/investigation-into-data-protection-compliance-in-the-direct-marketing-data-broking-sector.pdf. Regulatory action (but not an audit) was taken in relation to a data broker, Emma's Diary: ICO, Emma's Diary fined £140,000 for selling personal information for political campaigning, 9 August, 2018, Press Release: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/08/emma-s-diary-fined-140-000-for-selling-personal-information-for-political-campaigning/.

11 ICO, ICO takes enforcement action against Experian after data broking investigation, 27 October 2020, https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/10/ico-takes-enforcement-action-against-experian-after-data-broking-investigation/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

elections[12] (seemingly some claims of influence by Cambridge Analytica may have been unfounded). Given the general poor compliance with basic principles of data protection law in this area, however, the ICO has published guidance to political parties on the use of personal data in political campaigns.[13]

Despite what might seem unremarkable conclusions, the investigation is significant not just for the interpretation of the substantive obligations but also for the ICO's use of its enforcement powers which had been extended by the DPA18 (e.g., compulsory audit under Article 58(1)(b) GDPR). The wide territorial reach of the GDPR is illustrated by the investigation of Aggregate IQ (AIQ), a Canadian analytics firm linked to Cambridge Analytica. The ICO served an enforcement notice under section 149 DPA18,[14] its first notice under the GDPR/DPA18 regime, requiring AIQ to cease processing the personal data of UK and EU citizens, processing that was in violation of Articles 5, 6 and 14 GDPR. It was also the first time the ICO, relying on Article 3(2)(b) GDPR, had attempted to enforce against an entity outside the jurisdiction. It determined that the GDPR rather than just the Directive was relevant because, although the data were collected before the entry into force of the GDPR, AIQ continued to hold (and therefore process) the data afterwards.

The possibility of such extraterritoriality had been recognised as the GDPR came into force[15]. The election investigation showed that extraterritoriality might also operate when the relevant parties' locations were reversed. The ICO served an enforcement notice on SCL Elections Ltd (a UK company) to compel it to deal properly with a data subject access request from an American, Professor Carroll. SCL responded that non-UK citizens had no rights under the GDPR, a view the ICO did not share. It took the position that SCL was based in the UK and therefore subject to the law of that jurisdiction. The question has not yet been judicially considered.

---

12 ICO, Letter to Digital, Culture and Media and Sport Select Committee, 2 October 2020, https://ico.org.uk/media/action-weve-taken/2618383/20201002_ico-o-ed-l-rtl-0181_to-julian-knight-mp.pdf.

13 ICO, Guidance for the use of personal data in political campaigning, https://ico.org.uk/for-organisations/guidance-for-the-use-of-personal-data-in-political-campaigning/.

14 https://ico.org.uk/media/2259362/r-letter-ico-to-aiq-060718.pdf.

15 See e.g. K Hon, "GDPR's extra-territoriality means trouble for cloud computing", (2016) 140(Apr) *Privacy Laws and Business International Newsletter* 25.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

In addition to the ICO enforcement notice, AIQ (along with other companies – for example Facebook) were subject to investigation in other jurisdictions. So, this case also illustrates the importance of international regulatory cooperation.

In this enforcement action, the ICO also used its criminal enforcement powers under s47(1) DPA98 against SCL, which had chosen to ignore the enforcement notice the ICO had issued in relation to Professor Carroll.[16] This tough approach to enforcement was reinforced by the ICO referring SCL, as it had become insolvent, to the Insolvency Service, which in turn disqualified the directors of the company from acting as such for a period of seven years.[17] Suggesting that it would not be easy for those behind a company to avoid regulation by establishing new companies, the ICO stated that it would "monitor closely any successor companies using our powers to audit and inspect".[18]

Another theme relating to this investigation concerns the challenges to the ICO's decisions. Facebook challenged its fine, alleging procedural unfairness, showing that decision-making processes are important and may be a point of dispute especially where penalties are significant. The parties settled the action, with Facebook agreeing to pay the fine but, significantly, making no admission of liability as to the basis on which the fine was levied (though it had carried out an app audit[19] and changed its

16 ICO, SCL Elections prosecuted for failing to comply with enforcement notice, (January 2019), https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/01/scl-elections-prosecuted-for-failing-to-comply-with-enforcement-notice/.

17 The Insolvency Service investigation determined that "he caused or permitted SCL Elections Ltd or associated companies to market themselves as offering potentially unethical services to prospective clients; demonstrating a lack of commercial probity": The Insolvency Service, Press Release: 7-year disqualification for Cambridge Analytica boss, 24 September 2020, https://www.gov.uk/government/news/7-year-disqualification-for-cambridge-analytica-boss.

18 ICO, ICO statement: investigation into data analytics for political purposes, 2 May 2018, https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2018/05/ico-statement-investigation-into-data-analytics-for-political-purposes/.

19 Facebook, An Update on Our App Investigation and Audit, 14 May 2018, https://about.fb.com/news/2018/05/update-on-app-audit/; Facebook, An Update on Our App Developer Investigation, 20 September 2019, https://about.fb.com/news/2019/09/an-update-on-our-app-developer-investigation/; Facebook Taking Legal Action Against Those Who Abuse Our Platform, 27 August 2020, https://about.fb.com/news/2020/08/taking-legal-action-against-those-who-abuse-our-platform/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

process regarding access to the API).[20] Facebook is not alone in turning to litigation, and the substance of the ICO's reasoning has been challenged as well as its processes. Leave.EU challenged[21] a fine imposed (under the Privacy and Electronic Communications Regulations[22] (PECR)) for including marketing materials in communications with Leave.EU's subscribers. It lost at first instance and was unsuccessful on appeal.[23] It has announced its intention to appeal again. Appeals by the Liberal Democrats and UKIP were dismissed at first instance.[24] Experian also plans to challenge the ICO's interpretation of the GDPR.[25]

Chapter 4. Online Advertising

The ICO cited web and cross-device tracking as one of its three regulatory priority areas in its Technology Strategy 2018-2021. Advertising and the use of data is a broad topic, but 'adtech' and real-time bidding (RTB) systems are central. Adtech is the umbrella term for the range of software and tools used to target, deliver, and analyse their digital advertising. RTB is a real-time automated digital auction process that allows advertisers to bid for online ad space from publishers, with the highest bid usually winning. Significantly, the ads are personalised; to make the assessment as to whether or how to bid, data about the person viewing the page/app must be shared through the RTB system. This brings data protection rules into play.

---

20 ICO, Statement on an agreement reached between Facebook and the ICO, 30 October 2019, https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/10/statement-on-an-agreement-reached-between-facebook-and-the-ico/#.

21 A case challenging another fine for using an insurance company's mailing list to send out political material was withdrawn.

22 The Privacy and Electronic Communications (EC Directive) Regulations 2003 (SI 2003/2426), implementing the e-Privacy Directive (Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector [2002] OJ L201/37).

23 *Leave.EU and Eldon v Information Commissioner* [2021] UKUT 26 (AAC).

24 ICO, Letter to the Chair of DCMS Select Committee, 2 October, 2020, https://ico.org.uk/media/action-weve-taken/2618383/20201002_ico-o-ed-l-rtl-0181_to-julian-knight-mp.pdf.

25 Experian, Response to ICO Enforcement Notice in relation to UK marketing services, 27 October 2020, https://www.experianplc.com/media/news/2020/response-to-ico-enforcement-notice-in-relation-to-uk-marketing-services/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The ICO announced an investigation into adtech because of its complexity and scale, the risks posed to the rights and freedoms of individuals, as well as concerns expressed by some actors about the use of the technology. The ICO released an interim report[26] identifying a number of issues of concern. Risks were found to arise from profiling within the meaning of Article 4(4) GDPR and automated decision-making; large-scale processing including of special categories of data; the use of new/innovative technologies; combining and matching data from multiple sources; geolocation tracking; the tracking of behaviour; and the fact the processing was effectively invisible (as also found in the political advertising investigation). In particular, the ICO highlighted transparency and consent to processing; while some actors sought to rely on 'legitimate interests', the ICO noted that the circumstances in which this basis for processing would be available would be limited. In general, the scale of both the creation of data and the sharing of those data was assessed as disproportionate, intrusive and unfair – especially given that data subjects were unaware that this is happening. Further, the sharing of data through the supply chain, which relied on contractual arrangements (especially standard terms and conditions), was viewed as problematic, particularly given the type of personal data shared and the number of intermediaries involved.

The interim report gave the industry six months to respond to the issues raised.[27] Despite some changes[28], the ICO subsequently commented:

"while many organisations are on board with the changes that need making, some appear to have their heads firmly in the sand".[29]

Its activities on this project were suspended as a result of the pandemic; it was only in January 2021 that the ICO announced that its investigation was to re-start.[30] Nonetheless, there has been some concern amongst civil society actors as to the rate of progress; against this background there is

---

26 ICO Update report into adtech and real time bidding, 29 June 2019, https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906-dl191220.pdf.

27 S. McDougall, Blog: Adtech - the reform of real time bidding has started and will continue, 17 January 2020, https://ico.org.uk/about-the-ico/news-and-events/blog-adtech-the-reform-of-real-time-bidding-has-started/.

28 See e.g. M. Dunphy-Moriel and S. Dittel, "A real-time bid to restore trust in online advertising: DMA's seven-step ad tech and other industry initiatives" (2020) 31 *Ent. LR* 233.

29 McDougall (n. 27).

30 ICO, Press Release: Adtech investigation resumes, 22nd January 2021, https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2021/01/adtech-investigation-resumes/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

increasing strategic litigation (including collective actions).[31] The role of representative bodies here could be significant, given the asymmetries of information and resources between individual data subjects and global businesses. The Government has, however, decided not to implement Article 80(2) GDPR.

Chapter 5. Age Appropriate Design Code

The requirement for an Age Appropriate Design Code (AADC) derives from section 123 DPA18. There were concerns that a focus on the age at which children could consent, as found in the GDPR, was insufficient to tackle problems arising from the ways platforms are designed and which do not take the various levels of children's understanding into account. The AADC aimed at ensuring that those companies which provided services to children provided services that were appropriate to the children's respective developmental stages. The inclusion of this obligation was a significant step in recognising the impact of design choices in this context.

Section 123 specifies that the Information Commissioner must prepare a code of practice on standards of age appropriate design of "relevant information society services"[32] which are likely to be accessed by children - not just those which actively target children. The draft AADC was subject to Parliamentary approval.[33] 'Age appropriate' means that the services should be designed to be appropriate for children bearing in mind their developmental stage – so that design issues will be different depending on the age group served. Note that there are no requirements as to specific technology; by contrast, the Digital Economy Act 2017, Part 3 dealing with children's access to online pornography specified age verification technology for all age groups (though these provisions were not brought into force). The DPA18 nonetheless specified a minimum range of issues

31 K. Brimstead, "All I want for Christmas is not to be sued (by you and you and you...)!" (2020) 21 *Privacy and Data Protection* 6 provides an overview of procedure; *Lloyd v Google* [2019] EWCA Civ 1599 deals with the conditions for determining whether there is a class; the Supreme Court has heard an appeal but at the time of writing the judgment had not been handed down; *CMO v TikTok* is at an early stage, progress depending on the outcome of *Lloyd*; *Rumbul* v *Oracle and Sales Force* is also at a preliminary stage.

32 Defined s 123(7) DPA 18.

33 Section 125(3) and (4) of the DPA18.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to be considered. While many of these relate to rights and requirements found in the GDPR, some might be seen as going beyond that.

The AADC incorporates the principle that the best interests of the child should be a primary consideration in all actions concerning children; in this it borrows from the approach found in the UN Convention on the Rights of the Child (UNCRC). Following the UNCRC, a child is anyone under the age of 18. The AADC sets out 15 principles of age appropriate design, reflecting the concerns identified in section 123. In addition to the focus on the best interests of the child, they are: the need to carry out data impact assessments; that approaches adopted are age appropriate; transparency requirements; children's personal data should not be used in ways detrimental to their well-being; up-hold the services policies and standards; high privacy settings should be the default position; data minimisation; children's data should not be disclosed without a compelling reason; geolocation should be switched off by default; the child should be given age appropriate information about the existence of any parental controls; consent to profiling should be opt-in rather than opt-out and only allowed when appropriate measures are in place to protect children from any harmful effects; nudge techniques should not be used to get children to turn off protections; effective tools to be provided in connected toys; and prominent and accessible tools should be provided for children to exercise their rights. These are not technical design requirements but are a set of technology-neutral design principles; as with the DPA18, the AADC does not mandate any particular solutions. Assessment and mitigation of risk falls to service providers.[34] It remains to be seen how these requirements will be implemented by the ICO. While the AADC is now in force, the ICO allowed a 12-month transitional period, starting on 2 September 2020, to allow business time to prepare to comply with these obligations.

The AADC was not required by the GDPR and could be seen as a domestic experiment; other countries are, however, considering design codes. The Irish Data Protection Commissioner (DPC), for example, published draft "Fundamentals for a child oriented approach to data processing" in 2020. While these 'fundamentals' are not the same as the AADC, there is some consistency between the two, for example as regards the emphasis on data protection impact assessments, approach to profiling, data minimisation, geolocation and sharing of data.

---

34 For discussion of some initial concerns see e.g. R. Jay, "The Age Appropriate Design Code", (2020) 21 *Privacy and Data Protection* 3.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 6. Competition and Markets Authority

The Competition and Markets Authority (CMA) is an independent non-Ministerial government department dealing with competition enforcement and consumer protection. It will be granted further powers in relation to digital markets as envisaged in a number of reviews and reports on digital markets.[35] The Furman Report recommended the creation of a Digital Markets Unit (DMU) and the Government established the Digital Markets Taskforce (the "Taskforce"), led by the CMA, to make recommendations on the establishment of a regulatory framework for digital markets.[36] The CMA will also be expected to collaborate with the other regulators with competence in the digital sectors: Ofcom, the ICO and the Financial Conduct Authority (FCA). Together they established the Digital Regulators Cooperation Forum (DRCF).[37] It should be noted that while these regulators may have the most involvement with digital markets, they are not the only regulators those markets may touch. DRCF acknowledges this, as well as the likely need for engagement internationally.

Chapter 7. Competition Policy

The CMA has responsibility under the Competition Act 1998 for enforcing the prohibition on agreements and conduct which prevent, restrict or distort competition (Chapter 1 prohibition), and conduct which constitutes an abuse of a dominant position (Chapter 2 prohibition). The CMA has the power to impose fines and, in relation to cartels, criminal sanctions may be available. The Enterprise Act 2002 (EA02) introduced

35 Report of the Digital Competition Expert Panel, Unlocking Digital Competition, March 2019 (Furman Report); Lear, Ex-post Assessment of Merger Control Decisions in Digital Markets: Prepared for the Competition and Markets Authority, 9 May 2019 (Lear Report).

36 Digital Markets Taskforce Terms of Reference: https://www.gov.uk/government/publications/digital-markets-taskforce-terms-of-reference/digital-markets-taskforce-terms-of-reference--3.

37 CMA, Ofcom, ICO, Digital Regulation Cooperation Forum Launch Document, 1 July 2020, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/896827/Digital_Regulation_Cooperation_Forum.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

market studies[38] and market investigations to the CMA toolbox. The CMA also has responsibility for reviewing mergers under the EA02.

Following the Furman Report's recommendation that there be a market study into the digital advertising market,[39] the CMA investigated three broad heads of harm: the impact on consumers of online platforms' market power; the ability of consumers to control how data about them is used and collected by online platforms; and distortion in the digital advertising market caused by any market power held by platforms. The CMA concluded that "concerns we have identified in these markets are so wide ranging and self-reinforcing that our existing powers are not sufficient to address them".[40] The Government envisaged the recommendations from the market study would be taken forward through the establishment of the DMU.[41]

The Taskforce's advice[42] envisaged that the DMU be established within the CMA. It will operate a new regime applying to certain digital businesses designated as having "strategic market status" (SMS). The test for SMS is where a company has a "substantial, entrenched market power in at least one digital activity, providing the firm with a strategic position". SMS status would apply to the entire group of which the relevant company formed part. These businesses would be subject to an *ex ante* regime with three main elements. The first is a binding statutory code of conduct (with financial penalties of up to 10% of worldwide turnover for breach of the code). Secondly, the DMU would initiate proactive interventions

38 Market studies examine why a particular market may not be working well. The range of possible outcomes includes recommendations to government or initiation of a market investigation.

39 An investigation was also recommended by the Cairncross Review into Sustainable Journalism, 12 February 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/779882/021919_DCMS_Cairncross_Review_.pdf as well as the House of Lords Report, Regulating in a Digital World (HL Paper 299), 9 March 2019, available: https://publications.parliament.uk/pa/ld201719/ldselect/ldcomuni/299/299.pdf.

40 CMA, *Online platforms and digital advertising Market study final report,* 1 July 2020, p. 5, https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ALT_TEXT.pdf.

41 BEIS and DCMS, Government Response to the CMA's market study into online platforms and digital advertising, November 2020, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/939008/government-response-to-cma-study.pdf.

42 CMA, *A new pro-competition regime for digital markets: Advice of the Digital Markets Taskforce*, December 2020 (CMA 135), https://www.gov.uk/cma-cases/digital-markets-taskforce.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

targeted at SMS firms, including interventions relating to personal data mobility, interoperability and access to data so as to promote competition and innovation. Finally, special merger rules will require SMS firms to report all transactions to the CMA; normally the UK merger regime does not require parties to notify transactions. The new regime will also impose mandatory and suspensory notification requirements for transactions that meet certain thresholds. Although the Government has committed to introducing legislation to introduce the new regime, it is unclear when there will be Parliamentary time for the bill. Nonetheless, the DMU itself was launched on a non-statutory basis to focus on operationalising and preparing for the new regime on 7 April 2021.[43]

The CMA has also reviewed its Merger Assessment Guidelines[44] to reflect its recent decisional practice under the Competition Act which takes account of a broader context and the risk of consumer harm in assessing whether the threshold for intervention is met. Its approach to the 'share of supply' test,[45] allowed it to intervene in deals involving targets with very low (or even no) turnover, for example when technology rights are involved, and has been approved by the Competition Appeal Tribunal (CAT).[46] In all this, the CMA seems to take a comparatively interventionist stance,[47] and has challenged some deals that have been permitted by other competition authorities around the world.

The CMA's expansive use of its powers has, however, been subject to legal challenge.[48] Facebook appealed against the CMA's intervention[49] in Facebook's acquisition of Giphy, arguing the intervention was irrational, disproportionate and infringed the principle of legal certainty. The CAT unanimously dismissed the application,[50] and the CMA is now carrying

43 https://www.gov.uk/government/collections/digital-markets-unit.
44 CMA, *Merger Assessment Guidelines* (CMA129), 18 March 2021, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/986475/MAGs_for_publication_2021_-.pdf.
45 S. 23(2)(b) EA02.
46 *Sabre Corporation v Competition and Markets Authority* [2021] CAT 11.
47 M. Jephcott and V. Karadakova, 'The CMA's increasingly expansionist approach to the share of supply test in UK merger control: a threshold issue", (2020) 41(9) *European Competition Law Review* 466.
48 Section 120 EA02; *Sabre* (n47) is another example.
49 It imposed an initial enforcement order (IEO) aimed at preventing pre-emptive action by the companies involved which might otherwise restrict the CMA's ability to secure remedies at the conclusion of its merger review.
50 *Facebook v CMA* [2020] CAT 23.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

out a full merger inquiry.[51] Facebook, however, is pursuing the action against the CMA before the appellate courts.[52]

It should be noted that the CMA has not just applied its powers in relation to mergers. For example, it has opened an investigation into Apple's app store, in particular, the terms and conditions governing app developers' access to Apple's App Store under the Chapter II prohibition. It has similarly launched an investigation into Google's 'privacy sandbox' changes to its Chrome browser. Note that those changes introduced and potentially problematic in a competition context might be seen as a good thing from the data protection perspective.[53] The ICO and CMA have issued a joint statement,[54] but this tension highlights the importance of the DRCF as a venue for regulatory coordination and cooperation. Finally, the CMA has opened an investigation into whether Facebook has unfairly used the data gained from its advertising and single sign-on to benefit its own services, notably Facebook marketplace. It is noteworthy that the EU Commission has also launched an investigation. While the two investigations are separate, the CMA envisages working closely with the European Commission on this issue.[55]

Chapter 8. Consumer Protection

The CMA also has competence in the consumer protection field under the Enterprise Act 2002 (as amended) (EA02). It is not the only body with consumer protection powers: The Trading Standards Authority, for example, deals with misleading statements and acts as backstop regulator to the Advertising Standards Authority (ASA) in relation to advertisements not caught by the audiovisual regime. The CMA's enforcement powers

51 https://www.gov.uk/cma-cases/facebook-inc-giphy-inc-merger-inquiry.

52 *Facebook v CMA* (nyd); hearing available here: https://www.judiciary.uk/publications/facebook-inc-another-v-the-competition-and-markets-authority/.

53 This tension is discussed by D. Geradin et al, "GDPR Myopia: how a well-intended regulation ended up favouring large online platforms - the case of ad tech", (2021) 17 *European Competition Journal* 47.

54 CMA and ICO, *Competition and data protection in digital markets: a joint statement between the CMA and the ICO*, 19 May 2021, https://ico.org.uk/media/about-the-ico/documents/2619797/cma-ico-public-statement-20210518.pdf.

55 CMA, Press Release: CMA investigates Facebook's use of ad data, 4 June 2021, https://www.gov.uk/government/news/cma-investigates-facebook-s-use-of-ad-data.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

include both civil and criminal mechanisms.[56] Part 8 EA02 constitutes the main civil enforcement regime, giving the CMA the power to apply to the court for an enforcement order in relation to any rules identified by the EA02. These orders may include 'enhanced consumer measures' which require business to take additional steps for the protection of consumers. Alternatively, the CMA may accept an undertaking from the relevant business. The CMA also has powers under the Consumer Rights Act 2015 in relation to unfair terms.[57] The UK retained after Brexit rules[58] derived from the EU Consumer Protection Co-Operation Regulation.[59]

Using its current powers, the CMA has launched a number of consumer protection investigations in the online context: fake online reviews (leading to commitments from Facebook to do more to tackle the problem in 2020 and in 2021[60]); unfair roll-over contracts in subscriptions for online gaming[61] and anti-virus software;[62] problems with nudging techniques on hotel booking sites;[63] unclear policies especially as regards data sharing on data-sites;[64] and lack of disclosure of incentivised endorsements on social media platforms. The CMA has tackled a wide range of issues: these investigations have identified issues with content, business models as well as with platform design.

As with the ICO, international collaboration is important in this sector. The CMA's work on dating platforms was part of an international project on the fairness of platforms' terms and conditions.[65] The project overall aimed at securing disclosure around the data collection and privacy terms

56 Criminal enforcement powers are found in The Consumer Protection from Unfair Trading Regulations 2008 (SI 2008/1277), https://www.legislation.gov.uk/uksi/2008/1277/contents/made.

57 The CMA's approach to these powers is described in its guidance on unfair terms: CMA, Unfair Contract Terms Guidance, 31 July 2015 (CMA37) para 6.4, https://www.gov.uk/government/publications/unfair-contract-terms-cma37.

58 The Consumer Protection (Enforcement) (Amendment etc.) Regulations 2020 (SI 2020/484), https://www.legislation.gov.uk/uksi/2020/484/made.

59 Regulation 2017/2394 Consumer Protection Co-Operation Regulation [2017] OJ L345/1.

60 https://www.gov.uk/government/news/cma-intervention-leads-to-further-facebook-action-on-fake-reviews.

61 https://www.gov.uk/cma-cases/online-console-video-gaming.

62 https://www.gov.uk/cma-cases/anti-virus-software.

63 https://www.gov.uk/cma-cases/online-hotel-booking.

64 https://www.gov.uk/cma-cases/online-dating-services.

65 CMA, Blog: Why we're banging the drum for international fairness in the digital economy, 29 June 2018, https://competitionandmarkets.blog.gov.uk/2018/06/29/why-were-banging-the-drum-for-international-fairness-in-the-digital-economy/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of apps as well as to prevent nudging techniques being used in breach of consumer protection rules (e.g. pressure selling, scarcity claims and subscription traps).

The CMA has also worked with other UK regulators, for example the Gambling Commission, which was concerned about potentially unfair terms and practices in the online gambling sector.[66] The work on non-disclosed adverts by influencers has also fallen within the remit of the co-regulator, the ASA, which has targeted advertisers and influencers;[67] the CMA's work, by contrast, resulted in undertakings from the platforms themselves (as well as guidance to influencers[68]). The ASA's work is not limited to non-disclosure issues but extends to ensuring compliance with all advertising rules.

Nonetheless, the CMA has expressed concerns about the effectiveness of these powers, especially in the digital context, and has suggested that there be legislative reform of its consumer protection powers.[69] The CMA characterised its enforcement powers as weak; it highlighted the fact that it cannot order the cessation of practices it considers to be illegal, but must pursue businesses through the courts and even then no fines are available. It proposed bringing its consumer protection powers in line with its competition powers, so that the CMA would be able to decide whether consumer protection law has been broken; declare the fact publicly; direct businesses to bring infringements to an end; and impose fines. It also

---

66 https://www.gov.uk/government/news/gambling-sector-told-to-raise-its-game-after-cma-action; discussed J. Althoff, "Crackdown in the online gambling sector", (2018) 29(1) *Ent LR* 7.

67 ASA, *Influencer Ad Disclosure on Social Media - A report into Influencers' rate of compliance of ad disclosure on Instagram*, https://www.asa.org.uk/uploads/assets/dd740667-6fe0-4fa7-80de3e4598417912/Influencer-Monitoring-Report-March2021.pdf; see also ASA guidance for influencers: https://www.asa.org.uk/resource/influencers-guide.html; discussed O. Bray and V. Noto, "#Ad-vice for influencers and brands: how to comply with CAP's new influencer's guide", (2019) 30(1) *Ent. LR* 11; J. Agate et al., "Influencer advertising: the latest ASA findings" (2020) 31(1) *Ent LR* 14.

68 CMA, Guidance: Social media endorsements: being transparent with your followers, 23 January 2019, https://www.gov.uk/government/publications/social-media-endorsements-guide-for-influencers/social-media-endorsements-being-transparent-with-your-followers.

69 CMA, Letter to the Secretary of State for Business, Energy and Industrial Strategy, 21 February 2019, https://www.gov.uk/government/publications/letter-from-andrew-tyrie-to-the-secretary-of-state-for-business-energy-and-industrial-strategy/summary-of-proposals-from-andrew-tyrie-cma-chair-to-the-secretary-of-state-for-business-energy-and-industrial-strategy.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

suggested that it should be able to order the cessation of practices on an interim basis. In addition to considering fines, it suggested that the CMA should be able to seek the disqualification of company directors. This is the case for competition law but few cases have resulted in disqualification orders.[70]

The Taskforce highlighted specific issues based on the CMA's experience in digital markets. It noted the problematic use of dark patterns and nudging techniques, suggesting that a more explicit duty on firms "to take reasonable and proportionate steps to reflect consumers' interests in the design of their products and services" could be a means of tackling this issue.[71] In proposing this solution, the Taskforce noted that such an approach would complement the 'fairness by design' duty suggested in the CMA's market study final report,[72] as well as the statutory duty of care proposed in the Online Harms White Paper (OHWP).[73] The Taskforce also noted the importance of stronger enforcement of the Platform to Business Regulation[74] (as retained). Again, it is unclear what the legislative timetable would be for bringing in any changes.

---

70 Sections 9A- 9E Company Directors Disqualification Act 1986; CMA, Guidance on Competition Disqualification Orders, 8 February 2019 (CMA 102), available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/910485/CMA102_Guidance_on_Competition_Disqualification_Orders__FINAL__PDF_A-.pdf; C. Chijioke-Oforji, "Director accountability for breach of competition law: important practical lessons from the CMA's increased use of disqualification powers", (2021) 42 *European Competition Law Review* 24 and S. Caliskan 'Directors' disqualification in UK competition law: has the dog started barking?" (2020) 41 *European Competition Law Review* 509 discusses the recent use of these powers in the competition arena.

71 CMA, A new pro-competition regime for digital markets Advice of the Digital Markets Taskforce, December 2020 (CMA135), https://assets.publishing.service.gov.uk/media/5fce7567e90e07562f98286c/Digital_Taskforce_-_Advice.pdf, para 5.26.

72 CMA, Market Study into Online Platforms and Digital Advertising, 1 July 2020, available: https://assets.publishing.service.gov.uk/media/5fa557668fa8f5788db46efc/Final_report_Digital_ALT_TEXT.pdf, paras 8.123-8 and Appendix Y.

73 DCMS Online Harms White Paper (CP57), 8 April 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973939/Online_Harms_White_Paper_V2.pdf.

74 Regulation (EU) 2019/1150 on promoting fairness and transparency for business users of online intermediation services OJ [2019] L186/57.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 9. Internet Safety and Online Harms

In the Autumn of 2017, the Government published a Green Paper on Internet Safety.[75] It identified a wide range of concerns[76] but mainly envisaged self-regulation and media literacy as the tools to deal with them. Government policy underwent a rapid change. In Spring 2018, the Secretary of State announced that as part of its Digital Charter, the government would introduce laws to ensure that "the UK is the safest place in the world to be online", reflecting the words of then Prime Minister, Theresa May, at Davos in January 2018. The proposed approach was at that stage unclear; the Online Harms White Paper (OHWP) did not emerge until April 2019. Unusually for a white paper, a number of details were undecided so the OHWP also constituted a consultation on those elements. Further progress was slow. There was an interim response to the OHWP[77] before the Full Government Response (FGR)[78] was published on 15 December 2020. In the meantime, however, the UK implemented the changes to the Audiovisual Media Services Directive[79] meaning the provisions on video sharing platforms have been in force in the UK since 1 November 2020, with Ofcom, the independent communications and media regulator, as the competent body. The Government also decided not to bring into force Part 3 of the DEA, a decision which has been contentious. Following the Queen's Speech for the 2021-22 Parliamentary session, the Government published the draft Online Safety Bill (OSB) for pre-legislative scrutiny.[80]

The OHWP proposed imposing a statutory duty of care on operators within remit. While the OHWP did not specify the extent of this duty,

---

75 DCMS, *Internet Safety Strategy Green Paper*, 11 October 2017, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650949/Internet_Safety_Strategy_green_paper.pdf.

76 See Annex A to the *Internet Safety Green Paper* (n. 76).

77 DCMS, Online Harms White Paper - Initial consultation response, 12 February 2020, https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response.

78 DCMS, Online Harms White Paper: Full Government Response to the consultation, December 2020 (CP354), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944310/Online_Harms_White_Paper_Full_Government_Response_to_the_consultation_CP_354_CCS001_CCS1220695430-001__V2.pdf.

79 The Audiovisual Media Services Regulations 2020 (SI 2020/1062), https://www.legislation.gov.uk/uksi/2020/1062/made.

80 Pre-legislative scrutiny is not a part of the legislative process but allows members of parliament to see proposals and consider general issues arising before the bill is finalised and formally presented to Parliament.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the proposal bears a marked resemblance to the proposal put forward by the Carnegie UK Trust (discussed in chapter 1.3. above). Significantly, the OHWP constituted a change from a focus on regulating the content of speech, or focusing on the host platform's immunity (or the conditions for loss of immunity). It was more orientated towards how the platforms operated and the impact of their features, as can been seen by issues that the OHWP raised for consideration: for example, down-ranking or reducing the visibility of content that has been disputed by reputable fact-checkers; improving the transparency of political advertising; promoting diverse news content; providing tools to users to help them protect themselves against harassment; steps to stop banned users from opening new accounts; tools to detect fake and spam accounts; processes to stop algorithms promoting self-harm or suicide content to users. The OHWP also envisaged improved process around take-down of content. The OHWP stated that there should be a regulator; Ofcom was confirmed as the regulator in the interim response. The powers of the regulator, as the experience of the ICO and the CMA have already demonstrated, are an important aspect of the regime especially given the asymmetry in information and resources between service providers and users.

The draft OSB imposes a number of duties on operators within scope, rather than a single over-arching duty of care, and in this there seems to be a difference even as regards the position in the FGR. The OSB imposes different obligations on "user-to-user services" and "search services". For both types of service, there is a difference between adult services and those likely to be accessible by children[81] which are subject to more stringent obligations.[82] It also seems that the OSB envisages reliance on age verification, though this is described in a technology neutral manner.[83] As regards adult services, all must take action in relation to "illegal content". Illegal content[84] comprises all crimes where the intended victim is an individual; the Law Commission was instructed to review the criminal law in relation to communications offences, with the expectation that the law may be revised to deal with issues such as abuse of intimate images. Terrorism and child sexual abuse and exploitation material are specifically mentioned (and have specific enforcement features[85]). Additionally, the Secretary of

81 That is, those under 18.
82 Content that is harmful to children is defined in cl 45 OSB.
83 Clause 26(3) OSB.
84 Defined cl 41 OSB,
85 Ofcom must produce separate codes in relation to terrorism and CSAEM and may use a "technology warning notice" Cl 63-68.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

State may identify priority areas. Only 'Category 1' services – determined according to the FGR by reference to their level of risk[86] – need to take action in relation to content that is harmful to adults – that is, harm understood as a significant adverse physical or psychological impact on adults of ordinary sensibilities.[87] While not expressly mentioned in the duties, following the reasoning of the FGR[88] disinformation and misinformation that could cause significant harm to an individual will be within scope of content harmful to adults, a point reaffirmed by the fact that a committee is to be established to advise Ofcom on this issue.[89] It seems that some issues envisaged as within scope by the OHWP (eg transparency of political advertising) are not within scope of the OSB. Conversely, some concerns (online scams) that have been outside the proposed scope of the regime since the OHWP might not be included.[90]

Central to the regime is the idea that companies have effective systems and processes in place to understand the risk their services (including the design of those services) pose and to improve user safety; service providers are required to carry out and keep up-to-date risk assessments relevant to the types of content found on their service (an "illegal content risk assessment", a "children's risk assessment" and an "adults' risk assessment").[91] The service providers then have different duties to mitigate those risks. In relation to illegal content and to services likely to be accessed by children, a service operator must take proportionate steps to mitigate risks identified.[92] The requirements as regards 'harmful but legal' content ("adults' safety duty") seems limited to enforcing the Terms of Service, which need have no specific minimum content (save where "priority content"[93], "content of democratic importance"[94] and "journalistic content"[95] are concerned). Priority content must be specifically addressed, though the nature of this obligation depends of that category of content and, again,

86 The relevant provisions in OSB are cl 59 and Schedule 4.
87 Clause 46 OSB.
88 FGR (n. 79), para 2.82, 2.84.
89 CL 98 OSB
90 DCMS, Press Release: Landmark laws to keep children safe, stop racial hate and protect democracy online published, 12 May 2021, https://www.gov.uk/government/news/landmark-laws-to-keep-children-safe-stop-racial-hate-and-protect-democracy-online-published.
91 Cl 5(2), (4) and (5) OSB.
92 Cls 9(2) and 10(2) OSB respectively.
93 Cl 11(2) OSB.
94 Cl 13(4) OSB.
95 Cl 14(6) OSB.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the obligation in relation to the adults' safety duty is weak. Additionally, there are specific duties to have regard to freedom of expression and privacy. In addition, all companies in scope will have a specific legal duty to have effective and accessible reporting and redress mechanisms. They must also produce transparency reports (Ofcom providing guidance on form, content and process[96]).

Ofcom has a key role in understanding the nature of risk and the approach to mitigation and adding detail to the regime set out in outline in the draft OSB. It is obliged to carry out a risk assessment to identify, assess and understand the risks and develop risk profiles for different kinds of regulated services[97] and use that to provide guidance to service providers to assist them in their risk assessments. It is only after the guidance has been published that the regulated services will be required to carry out their risk assessments.[98] Ofcom will issue codes of practice in relation to the safety duties, as well as duties in regard content of democratic importance, journalistic content and reporting and redress. These codes will be subject to Parliamentary approval. Significantly, the Secretary of State has a power of direction to ensure that the code of practice reflects government policy or to ensure the protection of national security or public health.[99]

The OSB permits high fines (up to 10% global annual turnover[100]) and business disruption measures. These measures include the power to require providers to withdraw access to key services (and in this seem to be a development of the mechanisms in section 21 DEA) or, in the case of serious failures of the duty of care, to block the non-compliant service.[101] As in other areas of its remit, Ofcom will take a proportionate approach to enforcement. The OSB contains provisions in relation to criminal liability for company directors; these will only be brought into force if certain conditions regarding non-compliance with the regime are met. The regulator's decision may be challenged using judicial review principles.[102]

The regime also envisages a 'super-complaint' mechanism, whereby an 'eligible entity' may lodge a complaint with Ofcom about the existence of feature giving rise to significant harm to a large number of users.[103]

---

96 Cl 50 OSB.
97 Cl 61 OSB.
98 Cl 62 OSB.
99 Cl 33 OSB.
100 Cl 85-86 OSB.
101 Cl 91-94 OSB.
102 Cl 104 and 105 OSB.
103 CL 106-108 OSB.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

There is no individual right of complaint in regard to a particular instance of harm to the regulator; a user in such a case should use the existing causes of action (against the person posting the content). In this there is a difference from the DPA18; whether follow-on actions (as seen in competition law as well as data protection) will be viable is unknown.

Chapter 10. Conclusions

This review demonstrates that the trend towards regulation of platforms exists, but that it is not one initiative but a multiplicity of actions in a range of policy spheres. This is not surprising given that there are elements of virtually all aspects of life online. Two questions arise: which are the lead areas; and how do the different sectors interact? This currently is uncertain given the present state of policy development, especially as regards the online safety agenda. The main regulators – the ICO, the CMA and Ofcom – are working together already which is essential to eliminate the risk of conflicting regulatory requirements and, as existing practice demonstrates, international cooperation will also be required. Despite potential overlap and tensions between regimes, a number of commonalities exist between them, notably the focus on the impact of design choices, and risk-based approaches to the applicability of the regimes.

A second similarity is the significance of the role of the regulators, and consequently the need for resources and appropriate powers. This is particularly noticeable with regard to the CMA and Ofcom, where new powers are envisaged to deal with digital markets; the ICO's powers have recently been extended, but that was driven by the GDPR. All regulators had general powers that were applicable to this field and which have been deployed to a greater or lesser extent. The existence of these powers is important given the need for legislation (at least as regards the CMA and Ofcom), and that progress particularly on online harms/online safety has been slow. In this, there is the difficulty of dealing with very rich companies and companies based outside the jurisdiction. Large fines are nothing new but there are indications that experimentation with enforcement tools, for example director's liability or director's disqualification as well as business disruption, is being considered but which may need legislative underpinning.

The final theme is the response, particularly of large companies, to enforcement of regulation. There has been a significant amount of litigation, draining the resources of the regulators and putting off the day on which the company must comply. This response, whether or not it is seen

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

as desirable, is hardly surprising – especially from 'long-pocket' litigants. The role and impact of collective litigation by users has yet to be fully understood.

Bibliography

Althoff, Juliane. "Crackdown in the online gambling sector." *Entertainment Law Review* 29, no. 1 (2018): 7-10.

Bevitt, Ann and Collins, Amy. "UK Enforcement: Top five focus areas." *Privacy & Data Protection* 20, no. 4 (2020): 10-12.

Brimsted, Kate. "All I want for Christmas is not to be sued (by you and you and you...)!" *Privacy & Data Protection* 21, no. 2 (2020): 6-10.

Caliskan, Samet. "Directors' disqualification in UK competition law: has the dog started barking?" *European Competition Law Review* 41, no. 10 (2020): 509-513.

Chijioke-Oforji, Chijioke. "Director accountability for breach of competition law: important practical lessons from the CMA's increased use of disqualification powers." *European Competition Law Review* 42, no. 1 (2021): 24-29.

Dunphy-Moriel, Marta and Dittel, Alexander. "A real-time bid to restore trust in online advertising: DMA's seven-step ad tech and other industry initiatives." *Entertainment Law Review* 31, no. 7 (2020): 233-236.

Geradin, Damien et al. "GDPR Myopia: how a well-intended regulation ended up favouring large online platforms - the case of ad tech." *European Competition Journal* 17, no. 1 (2021): 47-92.

Hon, Kuan. "GDPR's extra-territoriality means trouble for cloud computing." *Privacy Laws & Business International Report* no. 140 (April 2016): 25-28.

Jay, Rosemary. "The Age Appropriate Design Code." *Privacy & Data Protection* 21, no. 1 (2020): 3-7.

Jephcott, Mark and Karadakova, Vassilena. 'The CMA's increasingly expansionist approach to the share of supply test in UK merger control: a threshold issue." *European Competition Law Review* 41, no. 9 (2020): 466-475.

McDougall, Simon. Blog: "Adtech - the reform of real time bidding has started and will continue." January 17, 2020. https://ico.org.uk/about-the-ico/news-and-events/blog-adtech-the-reform-of-real-time-bidding-has-started/.

Sellars, Clare. "GDPR: one year on - ICO pulls back the curtain on the impact of the new regime." *Corporate and Trade Law Review* 25, no. 7 (2019): 172-174.

Shiner, Bethany. "Big data, small law: how gaps in regulation are affecting political campaigning methods and the need for fundamental reform." *Public Law 2019 (2):* 362-379.

Wenn, Christopher. "Can data protection solve the problem of microtargeting, manipulation of internet users and fake news?" *Entertainment Law Review* 29, no. 7 (2018): 216-218.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Social Media Users Data Access: Russian Legal Approach

Juliya Kharitonova, Larissa Sannikova

Abstract: The article is devoted to the problem of the legal protection of data of users of social networks. Businesses are interested in the data posted by users on their social media pages. Big data from social media users have a high potential commercial value. However, at present, Russian legislation does not provide for the legal possibility of processing data for transferring it to third parties. For the development of digital markets, it is important to find a balance between the personal data protection of social media users and data processing companies. For this purpose, a legal regime for open-access personal data is being introduced in the Russian jurisdiction.

Keywords: personal data protection, sensitive personal data, personal data in the public domain, Big data.

Chapter 1. Introduction

The issue of personal data control in social media networks is directly connected to those legal restrictions regulating the level of privacy given to the data in question[1]. The processing of a users personal data is regulated by Russian legislation, and disclosure of the information provided by the user, including personal data, is only possible at the request of a court, law enforcement agency and in other cases as prescribed by law. However, the interests of these vital companies that analyze and process the vast amounts of data that is acquired by such social networks remain unprotected. This paper analyzes one such case, VKontakte LLC vs DABL LLC, which aims at protecting small businesses (companies) when using publicly available data from open social networking pages for commercial purposes. In the next part of the paper, we explore the legal treatment of user data on social

1 Katharine Sarikakis and Lisa Winter, "Social Media Users' Legal Consciousness About Privacy", *Social Media + Society,* February 2017. https://journals.sagepub.com /doi/pdf/10.1177/2056305117695325

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

media under Russian law. Particular attention is paid to the new legal concept "personal data, which the personal data subject has permitted to disclose".

Chapter 2. VKontakte Case Study

A key case in Russian law was VKontakte LLC vs DABL LLC (No. 40-18827/17-110-180).

VKontakte LLC, the VKontakte social network (VK.com) operator, brought a claim against DABL LLC, claiming that the defendant's actions violated its exclusive rights. DABL extracted and then used user information from the VK.com Database. According to the plaintiff, the Database producer's rights were violated. The parties' attention in the lawsuit was focused on the protection of IP rights. However, the legal community also saw in this dispute a more profound problem about the nature of the existing legal regimes regarding users data in online social networks.

The court concluded that DABL had created Double Search, Social Link, and Social Attributes software. This software is based on its unique technological search methods and algorithms for storing and analyzing social networks data, including VKontakte. As the copyright holder of the above mentioned software, DABL offers to collect and automatically process social network users data on behalf of its clients, in order to assess the creditworthiness of potential and existing debtors who are users of such social networks. Thus, for the first time, Russian litigation has addressed the possibility of manipulating social network user data for commercial purposes.

According to the general idea, the social network consists of hardware, software, and information parts. The social network information part comprises several automated databases, each of which consists of independent elements (materials), systematized in a certain way, allowing finding and processing the elements using the software. One of such databases is a database of social network users, which contains a set of independent elements (user cards) with information about each registered user in the social network. The database is updated with a new standalone element through a given data collection algorithm as a new user registers through the social networking site.

According to the experts who conducted a study of Double Search software, a set of independent elements, presented in the form of individual user cards, was studied for the purpose of analytical processing of informa-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tion resulting from viewing and indexing by the search engine Double Search[2] of VK.com users` publicly available pages.

The defendant's software explored the pages of users who had set suitable privacy settings in the social network for search engines to index their pages. In the system of VK.com settings in the Data Management Rules of the website there is an opportunity for users to set the option "The page is available for indexing by search engines." [3]

The courts disagreed on this point at various stages of the proceedings.

The court of the first instance dismissed VKontakte's claim because the plaintiff had not proved that the database had been created. The defendant was searching for publicly available information. The owners of information presented in the profiles are the users themselves and the information published by them is, by setting the appropriate access mode by the user, closed or public, i.e. open for use by any persons following part 2 of Article 7 of the Federal Law "On Information, Information Technologies, and Information Protection", No. 149-FZ, July 27, 2006.

The court of appeal reviewed the decision in favor of VKontakte LLC, noting that the extraction of content from the database DABL LLC violated social media users' rights. Since the plaintiff had assumed obligations to ensure the protection of data from unauthorized copying, distribution, and reproduction, collection and other actions performed with information from the social network, for commercial purposes or its use in whole or in any part in any way are not permitted without the licensor's (social media user's) consent.

---

2 The defendant is the copyright holder of the following software:
Double Search - a specialized search engine for finding information about people, including social networks;
Social Link - a program for viewing the results of clicking on the links uploaded to it by reflecting on the user's screen the contents of the page/pages to which the links uploaded to Social Link lead. This program can handle any links, both those received from Double Search and those received from other search engines or other sources.
Social Attributes is a program designed to follow the links uploaded to it and display the results of the content analysis of the linked pages on the user's screen, in the form of a system of numerical coefficients assigned to specific groups of information.

3 You can choose who can visit your page, contact you and see what you post on your page. You can even make your profile completely private and protect your personal space from unknown people, leaving your page fully visible only to your friends.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The court of appeal emphasized that "the plaintiff guaranteed users the protection of information about them from outsiders who were not users of the social media network, regardless of whether the information was public or private. At the same time, a disclaimer from the VK.com Privacy Policy stated that they did not apply to third parties' actions and internet resources. Also, The Site Administration bears no liability for the actions of third parties which as the result of using the Internet or the Site Services obtained access to the User information in accordance with the confidentiality level selected by the User, for the consequences of use of the information which, due to the Site nature, is available to any Internet user. (clause 8). The VK.com Privacy Policy stipulates that The Site Administration takes technical, organizational, and legal measures to ensure that the User's personal data are protected from unauthorized or accidental access, deletion, modification, blocking, copying, dissemination, as well as from other unauthorized actions. (clause 7.1).

The VK.com License Agreement8 prohibits any actions (Reproduction, copying, collection, arrangement, storage, and transfer of information from the Social Network for commercial purposes) with the Social Network content without the licensor's consent (clause 5.16). In doing so, it stipulates that the licensee, i.e., the user, consents to the reflection of his data on the Personal page within the SNS functionality and that such data will be considered publicly available unless another mode of access is chosen by the subject (point 5.3).

It is the user who chooses the level of visibility of his/her profile per the VK.com Privacy Policy and accepts the responsibility that the information specified by him/her may be accessed by other users of the website, taking into account the specifics of the website architecture and functionality (point 5.2). Consequently, when a data subject decides to apply any privacy settings, he or she also determines the personal data regime, including its public accessibility, by his or her conclusive actions.

The VKontakte License Agreement applies to all Internet users who may not be users of the social network but access the VKontakte user page.

It is worth mentioning that since August 31, 2018, VKontakte has introduced fully private profiles, information from which is only available to those whom the person has added as "friends." [4] However, even if a user has a "closed" profile type, it can still be visible to all users of the

4 Sultan Suleymanov, "VK users can now close profiles from strangers", *Meduza*, August 31, 2018, https://meduza.io/feature/2018/08/31/vkontakte-pozvolila-zakryvat-profili-ot-postoronnih-v-tom-chisle-ot-politseyskih-kotorye-ischut-ekstremizm

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

"Internet," or to everyone except search services, or only to users of the social network VKontakte. However, a closed profile shows a small image of the person, name, date of birth, workplace, and city if they fill in the relevant fields. Thus, it seems that public accessibility is determined first by profile visibility and then by the categories of information that the user reveals depending on the type of profile. The presence of Public post or Friends only ("Visible to all" or "Visible to all except search sites") options in the visibility settings suggests that the profile is unambiguously accessible to all, as it is not restricted to social network members.

DABL Ltd.'s appeal reasonably stresses that the software processes the publicly available information of the social network profiles, originally intended to be accessible to all users. It is up to the subject to consciously dispose of its data and be aware of the consequences of publishing information in public social network profiles.

If the intellectual property regime is extended to the data published by users, the subject would be deprived of the right to dispose of the information about him/herself (Article 20.2 of the cassation appeal).

In rendering its final judgment in the case on March 22, 2021, the court expressed its opinion that the Respondent's software browses the pages of users expressly authorized for everyone to view them by clicking on links to those pages, the experts involved in the case agreed that Double Search was a specialized search engine. Defendant's customers do not use the information in the index (do not read it, do not analyze it, do not search for something in it, etc.) when they work with Double Search and search information for a user. They receive the results they need from the program in the form of links to users' webpages. The Respondent's software indexes only the pages of those users who have consented to this by using the appropriate privacy settings offered by VKontakte and have set the page to be open to all, and information from pages with other access settings is not indexed. Defendant's software interacts with the VKontakte site only within the rules and for the purposes set by the rights holder itself and only with those user pages that have explicitly expressed their consent (using the VKontakte site functionality) for their pages to be indexed by search engines.

These circumstances enabled the court to reject the plaintiff's claims and, in effect, to allow the disputed software to process, in algorithmic ways, open user data from social networks for commercial purposes.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3. Legal treatment of user data on social media under Russian law

The users themselves mainly provide the data that comes in and is stored in the social media information base. However, the set of such data can be particular in each case.

For example, VK.com has and makes it possible to upload user data to the following extent:

- location;
- registration data (name, surname, date of birth, gender, mobile number, email if provided);
- support service contacts;
- profile details (marital status, place of residence and hometown, education, career, and military service);
- history of visits to VKontakte and data about the device from which you are logged in;
- the automatically obtained information (e.g., when the user has logged in to third-party sites through VK.com and has given access to any information);
- the history of posts and subscriptions;
- messages;
- media files in which the user has been marked;
- payment data;
- information from third parties.

Legally, data collected by a social network is subject to different legal regimes.

Chapter 3.a. Personal data

Article 3, paragraph 1 of the Federal Law of 27.07.2006 No. 152-FZ (rev. 30.12.2020) "On Personal Data" states that personal data includes any information that directly or indirectly concerns a defined or identifiable natural person. The law represents the latter as the subject of personal data.

Russia has adopted the broadest approach, according to which personal data is any information: name, surname, patronymic, year, month, date of birth, place of birth, address, marital status, social status, property status, education, profession, income, other information relating to the subject of personal data (paragraph 2.5 of Federal Service for Supervision of Communications, Information Technology, and Mass Media (Roskomnadzor) Order No 94 dated 30.05.2017 (revised on 30.10.2018) "On approval of

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

methodological recommendations for notifying the competent authority on the beginning of personal data processing and on changes in previously submitted. "

The main legal attributes of personal data are distinguished:

1. information (information, data, reports, etc., following the Information Act).
2. relating directly or indirectly to an individual. Personal data containing direct information about a person (passport series and number, DNA, etc.) can be accurately identified. With indirect information about the person, he or she becomes "identifiable" (e.g., such information includes information about the education received). (Article 6 of the Information Act).
3. the subject of personal data is a human being.
4. the purpose of collecting, storing, and using personal data is to identify a data subject based on specific characteristics.
5. to be legally protected, personal data has to be recorded in a specific storage medium. This is information coming from any source and in any form. The Model Law on Personal Data of October 16, 1999, adopted to unify and harmonize the legislation of the countries of the former USSR, states that the information recorded on a tangible medium is subject to legal protection.

Personal data may be permissible for dissemination (Article 3 of the Federal Act of 27.07.2006 No. 152-FZ "On Personal Data").

The law divides personal data into groups:

1. general: 1) basic, 2) additional
2. special;
3. biometric: 1) physiological, 2) physical, 3) behavioral.

General personal data is data that can be identified with the highest degree of certainty.

General personal data includes basic and supplementary data. General basic data directly refers to a specific person: Full name and other passport details, date of birth, place of registration, and actual residence.

General additional data is, for example, information on education, profession, marital status, telephone number, etc. With the available general basic data, this type of information makes it possible to identify a person with almost absolute certainty.

Special categories of personal data include race, nationality, political views, religious or philosophical beliefs, health conditions, intimate life (clause 2.6 of Roskomnadzor Order No 94 dated 30.05.2017).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Biometric personal data characterizes a person's physiological and biological characteristics that can be used to identify the person. (clause 2.7 of Order of Roskomnadzor of 30.05.2017 N 94). An image of a person (photograph and video recording) that allows identification and is used by the operator for this purpose is considered personal biometric data (Clarifications by Roskomnadzor "On the issues of attributing photo and video images, fingerprints, and other information to personal biometric data and the specifics of their processing").

All of this data is used by the operator (the person who organizes and/or carries out personal data processing and determines the purposes and content of personal data processing) to establish the personal data subjects identity.

The law does not single out the so-called personal sensitive data of a citizen - one of the personal data types - but it does have increased importance to the individual. Personal data, the disclosure of which may cause substantial non-pecuniary damage to an individual. For example, information on race, sexual orientation, religious and political beliefs, criminal record, etc. Article 6 of the Convention for the Protection of Individuals concerning Automatic Processing of Personal Data states that it is unlawful to subject such data to automatic processing. The latter is possible only if the domestic law of the state provides appropriate safeguards (Convention for the Protection of Individuals concerning Automatic Processing of Personal Data. ETS No.108. Strasbourg, 28/01/1981).

In Russia, the principle of indirect identification of the data subject is enshrined, which allows an individual to claim data protection rights in many contentious situations. There is information that can, to a certain extent, identify an individual or a specific range of data subjects or, in conjunction with other personal data, identify a person. This approach is recognized in the doctrine and is also confirmed by international law-making practice. For example, the EU Directive 95/46/EC on personal data protection (GDPR) contains a similar approach.

The approach to understanding personal data established by Russian law is called context-oriented and has formal ambiguity. It is not easy to define precisely what information should be classified as personal data.

Chapter 3.b. Sensitive personal data

However, in general, the business model of most major social networks is built around personal data. The accumulation of personal, susceptible

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

information about the user and encouraging the user to disclose relevant information continuously is at the core of social media functioning.

Russian law does not distinguish a separate concept of sensitive personal data, unlike the GDPR rules, which qualify personal data as "sensitive" on a par with health, political and religious beliefs.

Centrally, it argues that scholars and regulators need to pay attention to the principle of intimacy[5].

M.A. Rozhkova[6] also notes that personal data, in general, is understood instead as data about citizens processed by public authorities. The researcher refers to them as:

1. unique identifiers of a person;
2. an image of a citizen;
3. unique identification numbers (TIN, Insurance Number of Individual Ledger Account);
4. publicly available information about a citizen that is self-published on the internet.

Regarding personal information that an individual puts in the public domain, by Article 152.2.1 of the Russian Civil Code, the subject of such information has no right to prevent its further use without his or her consent. As we can see, the legislator provides for the possibility of processing such information. That requires the subject of such information to place it in the public domain independently. Such information is regulated as publicly accessible.

Part 1 Article 7 of the Information Act determines that publicly accessible information is the information the access to which is unlimited. That establishes the presumption of openness of information, as it is any information to which access is open to everyone. As Article 7 (2) of the Information Act establishes, such information may be used without restrictions, but there are restrictions regarding dissemination.

A particular case of the above rule is the provision in point 2 of paragraph 1 of Article 152.2 of the Russian Civil Code which states that in a situation where information about a citizen's private life has previously been made publicly available or has been disclosed by the individual him-

---

5 Andrew McStay, "Empathic media and advertising: Industry, policy, legal and citizen perspectives (the case for intimacy)", *Big data & society* (December 2016), https://doi.org/10.1177/2053951716666868.

6 Marina Rozhkova, "Personal data: can they be classified as property? (view of a civilist)", Zakon.ru, February 28, 2019, https://zakon.ru/blog/2019/02/28/personalnye_dannye_mozhno_li_otnosit_ih_k_imuschestvu_vzglyad_civilista

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

self or by his will, it will not be considered a violation to collect, store, distribute or otherwise use such information without the citizen's consent. In other words, federal law here enshrines the rule that it is permissible to disseminate the designated publicly available information without the consent of the data subject.

It can be concluded that the posting of personal information on an internet site makes it publicly available. However, to disseminate such information, the procedure set out in Article 9 of the Law on Information must be followed.

Thus, personal data is any information that directly or indirectly relates to an identified or identifiable natural person. Personal data (or "personal sensitive data"), on the other hand, is understood as a particular type of personal data, which implies mainly personal information about an individual.

Chapter 3.c. Personal data in the public domain

Amendments to the Personal Data Law came into force on March 1, 2020, introducing the term "personal data, which the personal data subject has permitted to disclose" and establishing a special legal regime. Essentially, this refers to personal data that is publicly available and that users themselves post via geolocation tags, photos, audio, and video recordings, comments, reposts, participation in groups, polls, etc[7].

According to Article 10.1 of the Personal Data Law, users must expressly consent to the processing of personal data that is publicly available. Social networks and other digital platforms shall provide the user with the possibility to determine the list of personal data for each category of personal data specified in such consent. Consent can be executed directly on the digital platform's website or using a unique information system of the authorized body to protect personal data subjects' rights. Silence or inaction of the user under no circumstances can be considered as consent.

The consent must clearly express the user's will to disseminate the personal data to which the user has access. Otherwise, the user will be deemed not to have consented to the dissemination of their data. In the consent,

---

7 David Hiatt and Young B. Choi, "Role of Security in Social Networking", *(IJACSA) International Journal of Advanced Computer Science and Applications* 7, no. 2 (2016): 12, https://thesai.org/Downloads/Volume7No2/Paper_2-Role_of_Security_in_Social_Networking.pdf

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the user can establish prohibitions and conditions for the processing of the data, except for access to it. Thus, the law establishes a rather strict legal regime for publicly available personal data.

However, at present, social networks have not responded to the legislation changes on personal data. For example, not a single social network has offered its users consent to the processing of publicly available personal data with a list for each category of personal data. It appears that some social networks operating in Russia will be unable to meet these requirements due to their algorithm. For example, the algorithm of the well-known dating network Tinder.com is set up to provide information on the user's age and allow access to geolocation. The availability of this data allows for a more accurate matchmaking process. The digital platform is required by law to ensure that the user can set a ban on sharing personal data about them, but this cannot be easy to implement due to the algorithm in place.

User agreements with social media platforms usually specify that the user also bears third parties' risk using this information. Thus, point 2.1 of "Rules of Protection of Information about Users of VK.com" states directly that the user "understands that the information on the Site posted by the User about himself can become available to other users of the Site and Internet users and can be copied and distributed by such users." The Odnoklassniki social network followed the path of limiting its liability by explicitly stating that it "shall not be held liable for the actions of third parties that gain access to information about the User following the User's chosen level of privacy as a result of using the Internet or the Social Network, for the consequences of using information that, due to the nature of the Social Network, is available to any Internet user." Thus, social networks merely alert their users to the technical possibility of third parties collecting such information for further processing.

The processing of publicly available personal information using data mining systems provides insights into specific user behavior groups. This information is therefore of considerable interest for both commercial and other public purposes. Big Data is recognized as a new digital asset - BigDate - and is in high demand on the market[8].

Article 5(2) of the Personal Data Law stipulates that personal data may only be collected and processed for "specific, predetermined and legitimate purposes". Social networks (VK.com, Odnoklassniki.com, etc.) in their

8 Larisa Sannikova and Juliya Kharitonova, *Digital Assets: A Legal Analysis*, (Moscow, 2020), 58.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

rules specify as such a purpose the execution of an agreement with users. There is no legal possibility of transferring (sellng) the processed data as Big Data to third parties. Based on data processing's strict purpose, social networks cannot collect data for different purposes: fulfillment of user agreements and sale to third parties. Clause 3 of Article 5 of the Personal Data Law expressly prohibits combining databases containing personal data whose processing is incompatible with one another. Thus, at present, Russian social networks are not allowed to collect and process users' information for subsequent sale to third parties.

At the same time, all market participants recognize Big Data's value as a digital asset and the need to legalize its circulation. To date, the problem of legalizing Big Data circulation is closely linked to protecting individual personal data. The law prescribes that personal data must be destroyed or depersonalized once the purpose of its processing has been achieved (Article 5(7) of the Personal Data Law). The law essentially equates depersonalization with destruction.

However, an analysis of regulatory rules shows that these categories are not identical. According to the Methodological Recommendations on the application of Roskomnadzor Order No 996 of September 5, 2013, "On Approval of Requirements and Methods of Personal Data Depersonalisation," depersonalized data refers to data stored in information systems in electronic form that cannot be identified as belonging to a specific personal data subject without additional information.

This recommendation also contains a non-exhaustive list of methods of depersonalization:

> the method of introducing identifiers (replacement of the part of the information (personal data values) with identifiers with the creation of a table (reference book) of identifiers compliance with the original data);
> the method of composition or semantics modification (change of personal data composition or semantics using statistical processing results replacement, summarization, or deletion of part of the data);
> the decomposition method (splitting the set (array) of personal data into several subsets (parts) and then storing the subsets separately);
> shuffling (shuffling of individual records or groups of records in a personal data file).

It should be noted explicitly that Roskomnadzor Order No 996 of September 5, 2013, "On Approval of Requirements and Methods for Personal Data De-identification," makes reversibility a mandatory requirement for the properties of the de-identification method. Reversibility refers to the

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

possibility of de-anonymization, whereby anonymised data can be reduced to its original form, making it possible to determine whether the personal data belongs to a specific subject and eliminate anonymity.

Thus, in depersonalization, as opposed to destruction, the possibility of extracting information about a particular user as a whole is retained, and hence the risk of disclosure of user information is also retained[9].

When choosing a technique, it is crucial to maintain a balance between confidentiality and the data's usefulness[10]. The greater the anonymization of big data, the less accurate information can be gleaned from its analysis. Consequently, the value of such data decreases significantly.

According to the Russian regulator, big data can only be sold if the user has given his or her separate consent. As an additional protective measure, it is proposed to prohibit identifiers to third parties for de-identification procedures. The relevant bill is now in the Russian State Duma. However, a business has been skeptical of the bill, pointing out that the lack of an appropriate legal framework significantly hinders the digital economy's development.

Chapter 4. Conclusion

In general, it can be concluded that a legal regime for open-access personal data is being introduced in the Russian jurisdiction. If the data subject allows everyone to see his or her social media account data, personal data is considered publicly available. Any company can collect, analyse and share this information with its customers for commercial purposes.

Bibliography

Hiatt, David and Choi Young B. "Role of Security in Social Networking" *(IJACSA) International Journal of Advanced Computer Science and Applications* 7, no 2. (2016): 12-15. https://thesai.org/Downloads/Volume7No2/Paper_2-Role_of_Security_in_Social_Networking.pdf

9 Alexander Savelyev, "Problems of the application of legislation on personal data in the era of "Big Data"", *Law. Journal of the Higher School of Economics* 1 (2015).

10 Vladislav Kiselenko, "Anonymization of work in the global computer network Internet", *Vestnik Bauman MSTU Instrument making series* 1 (2005), https://cyberleninka.ru/article/n/anonimizatsiya-raboty-v-globalnoy-kompyuternoy-seti-internet

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Kiselenko, Vladislav. “Anonymization of work in the global computer network Internet.” *Vestnik Bauman MSTU Instrument making series* 1 (2005): 44-51. https://cyberleninka.ru/article/n/anonimizatsiya-raboty-v-globalnoy-kompyuternoy-seti-internet

McStay, Andrew. “Empathic media and advertising: Industry, policy, legal and citizen perspectives (the case for intimacy)”, *Big data & society* (December 2016). https://doi.org/10.1177/2053951716666868.

Rozhkova, Marina. “Personal data: can they be classified as property? (view of a civilist)” Zakon.ru, February 28, 2019. https://zakon.ru/blog/2019/02/28/personalnye_dannye_mozhno_li_otnosit_ih_k_imuschestvu_vzglyad_civilista

Sannikova, Larisa, and Kharitonova Juliya. *Digital Assets: A Legal Analysis*. Moscow: 4 Print, 2020.

Sarikakis, Katharine, and Lisa Winter. “Social Media Users’ Legal Consciousness About Privacy”. *Social Media + Society*, (February 2017), https://journals.sagepub.com/doi/pdf/10.1177/2056305117695325.

Savelyev, Alexander. «Problems of the application of legislation on personal data in the era of "Big Data"» *Law. Journal of the Higher School of Economics* 1 (2015): 43-66.

Suleymanov, Sultan. “VK users can now close profiles from strangers”, *Meduza*, August 31, 2018. https://meduza.io/feature/2018/08/31/vkontakte-pozvolila-zakryvat-profili-ot-postoronnih-v-tom-chisle-ot-politseyskih-kotorye-ischut-ekstremizm

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Hate Speech on Platforms

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Protecting Democratic Expression Online: Canada's Work in Progress

Richard Janda[*]

Abstract: In June, 2021, Canada's federal government finally introduced a part of its promised legislation to combat online hate speech. However, this bill, introduced on the last day Parliament sat before an election, was destined to "die on the order paper". Furthermore, the ambitious goal of creating a regulator for online platforms was postponed, although detailed consultation papers were issued in July, 2021. Thus, it has proved harder than anticipated to strike the balance between dismantling barriers to full democratic expression placed upon groups targeted by hate speech, on the one hand, and ensuring that platforms not be pushed toward zealous takedown practices, on the other. This article reviews Canada's existing legal framework, recent reports that are orienting policy options, the new bill, and the consultations papers issued by the government concerning additional new legislation. It concludes with some observations about how Canada could reinforce online dispute resolution and help shift platform business models that serve to amplify extreme content.

Keywords: online hate, platform governance, duty to act responsibly, social media councils, notice and takedown, online dispute resolution

There is something revealing about the very existence of this article. As texts were being gathered by the editors to present a comparative understanding of platform regulation, Canada was in the midst of formulating new legislation on online hate speech. Indeed, the Minister of Canadian Heritage had been given, on January 15, 2021, a new mandate by the Prime Minister to:[1]

---

* I am grateful to Judit Bayer, Lex Gill, Vivek Kirshnamurthy, and Taylor Owen for their assistance in the preparation of this article, though of course they bear no responsibility for any of its shortcomings.

1 Office of the Prime Minister of Canada, "Minister of Canadian Heritage Supplementary Mandate Letter," January 15, 2021, https://pm.gc.ca/en/mandate-letters/2

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

> Work with the Minister of Public Safety and Emergency Preparedness and the Minister of Justice and Attorney General of Canada to take action on combatting hate groups and online hate and harassment, ideologically motivated violent extremism and terrorist organizations. You will be supported in this work by the Minister of Diversity and Inclusion and Youth, the Minister for Women and Gender Equality and Rural Economic Development and the Minister of Innovation, Science and Industry.

After the March 2019 terror attack in Christchurch, New Zealand, Canada had joined Christchurch Call to Action to address violent extremism online.[2] In 2021, combatting extremism and hate speech was also presented as a matter of increasing priority in the wake of the Capitol Hill insurrection in Washington on January 6, and the Minister made multiple public statements to the effect the legislation would be forthcoming imminently, statements that continued to be made up to the moment when this article was submitted.[3] The author thus fully expected to be writing about the nature and implications of the new Canadian regime.

021/01/15/minister-canadian-heritage-supplementary-mandate-letter. In his 2019 Mandate Letter, the Minister had already been charged to "[c]reate new regulations for social media platforms, starting with a requirement that all platforms remove illegal content, including hate speech, within 24 hours or face significant penalties. This should include other online harms such as radicalization, incitement to violence, exploitation of children, or creation or distribution of terrorist propaganda." Office of the Prime Minister of Canada, "Minister of Canadian Heritage Mandate Letter," December 13, 2019, https://pm.gc.ca/en/mandate-letters/2019/12/13/minister-canadian-heritage-mandate-letter.

2 Office of the Prime Minister of Canada, "Canada joins Christchurch Call to Action to eliminate terrorist and violent extremist content online," May 15, 2019, https://pm.gc.ca/en/news/news-releases/2019/05/15/canada-joins-christchurch-call-action-eliminate-terrorist-and-violent. See Christchurch Call to Eliminate Terrorist and other Extremist Content Online, May 15, 2019, https://www.christchurchcall.com/call.html.

3 See, for example Elizabeth Thompson, "Canada not exempt from social media forces that created U.S. Capitol riot, heritage minister says," *CBC News*, January 29, 2021, https://www.cbc.ca/news/politics/facebook-twitter-canada-regulation-1.5894301, as well as Anja Karadeglija, "New definition of hate to be included in Liberal bill that might also revive contentious hate speech law," *National Post*, March 3, 2021, https://nationalpost.com/news/politics/new-definition-of-hate-to-be-included-in-liberal-bill-that-might-also-revive-contentious-hate-speech-law, and Bill Curry and Menaka Raman-Wilms, "New internet bill on hate crime and revenge porn coming in 'very near future,' Guilbeault says," *Globe & Mail*, June 7, 2021, https://www.theglobeandmail.com/politics/article-new-internet-bill-on-hate-crime-and-revenge-porn-coming-in-very-near/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

And yet it was not to be – at least not entirely. The urgent, imminent legislation to establish a new regulator for online platforms continues to be a chimera. However, on the last day of sitting of the current minority Parliament, the Minister of Justice introduced Bill C-36 which would amend the *Criminal Code* and the *Canadian Human Rights Act* so as to address certain dimension of online hate speech.[4] And just weeks before Canadians were called to vote in a general election, The Minister of Canadian Heritage released a Discussion Guide[5] and Technical Paper,[6] launching a public consultation about proposed legislation to be introduced in the fall of 2021 should the government be re-elected.

As a result, this article seeks to accomplish the following. First, it lays out in general terms the current state of Canadian law, which sets in context why the government – and the public – have concluded that there is need for legislative reform. Second, it describes and analyses reports that have been issued since 2019 aiming to pave the way for new legislation, with some emphasis on the work of the Canadian Commission on Democratic Expression, which issued an ambitious report in January 2021 just on the eve of the supplementary mandate issued to the Heritage Minister. Third, it assesses Bill C-36. Fourth, it gives an account of the Discussion Guide and Technical Paper that map out the approach the current government now proposes to take. Finally, it draws some lessons from the difficulties faced by the Minister in presenting this legislation, offering observations about the limits encountered when a state like Canada endeavours to create a new regulatory agency to control online speech.

This article is entitled "Protecting democratic expression online" rather than "Combatting online hate speech" so as to place emphasis upon the tension at play in seeking to eliminate harmful or dangerous forms of expression. The spread of hatred online can and indeed has transformed democratic expression into the sort of factionalism feared by James Madison, which he defined as arising when a group in "united and actuated

4 Bill C-36, *An Act to amend the Criminal Code and the Canadian Human Rights Act and to make related amendments to another Act (hate propaganda, hate crimes and hate speech)*, 2d sess., 43d Parliament, June 23, 2021, https://www.parl.ca/LegisInfo/BillDetails.aspx?Bill=C36&Language=E&Mode=1&Parl=43&Ses=2.

5 Digital Citizen Initiative, Department of Canadian Heritage, "Discussion Guide," July 26, 2021, https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html.

6 Digital Citizen Initiative, Department of Canadian Heritage, "Technical Paper," July 26, 2021, https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/technical-paper.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

by some common impulse of passion" and the effects of which, if left without response, could turn a republic into a mob.[7] On the other hand, the same James Madison, an architect of the First Amendment to the US Constitution, took a dim view of any prior restraints or *ex post facto* penalties imposed on publications.[8] Madison's opposing concerns help to frame the issues raised here.

---

7 James Madison, *Federalist* No.10, in *The Federalist Papers*, ed. Clinton Rossiter (New York: New American Library, 1961), https://avalon.law.yale.edu/18th_century/fed10.asp. See also Jeffrey Rosen, "America is Living James Madison's Nightmare," *The Atlantic*, October, 2018, https://www.theatlantic.com/magazine/archive/2018/10/james-madison-mob-rule/568351/.

8 In his "Report on the Virginia Resolutions," Madison insisted on constitutional protection against state encroachments upon freedom of the press: "This security of the freedom of the press requires that it should be exempt not only from previous restraint by the Executive, as in Great Britain, but from legislative restraint also; and this exemption, to be effectual, must be an exemption not only from the previous inspection of licensers, but from the subsequent penalty of laws." James Madison, "Report on the Virginia Resolutions," January, 1800, https://press-pubs.uchicago.edu/founders/documents/amendI_speechs24.html. See also David Sentelle, "Freedom of the Press: A Liberty for All or a Privilege for a Few?" *Cato Supreme Court Review* (2014): 25. Nonetheless, it is obviously important to distinguish a late 18th century conception of freedom of the press from the contemporary challenge posed by online platforms. Madison, who himself made sophisticated and influential use of the press, had imagined a "class of literati" who would become "cultivators of the human mind—the manufacturers of useful knowledge—the agents of the commerce of ideas—the censors of public manners—the teachers of the arts of life and the means of happiness." James Madison, "Notes for the National Gazette Essays" (ca. December 19, 1791–March 3, 1792), https://founders.archives.gov/?q=literati%20%22useful%20knowledge%22&s=1111311111&r=1. Given its freedom and elevated by such a class of literati, the press itself would serve to mediate and constrain the possible excesses of speech. See Colleen Sheehan, "The Politics of Public Opinion: James Madison's 'Notes on Government'," *William and Mary Quarterly* 49, no. 4 (1992) 621. Extreme expression on the internet is not mediated by internet literati.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

I. Canada's existing legal framework[9]

Prof. Natasha Tusikov has remarked that Canada "is continuing to outsource regulation to commercial platforms."[10] It is fair to point out that although Canada does have a relatively robust criminal law framework to address hate speech, it has lagged behind other jurisdictions in developing tools to address the online phenomenon.[11] The brief review here of Canada's existing legal framework will touch upon the following elements: a) the absence of an equivalent to Germany's NetzDG regime; b) the all-but non-existent current role for Canada's communications, human rights and privacy agencies; c) the restricted reach of criminal law; and d) the constraints imposed by the Canada-U.S.-Mexico Trade Agreement.

a. No equivalent to NetzDG

Canada does not currently have any functional equivalent to the German NetzDG legislation requiring takedown by platforms of "manifestly unlawful" content.[12] Interestingly, though, Canada does have experience with a quasi-takedown regime in the form of what is called "notice and notice" under the *Copyright Act*.[13] Pursuant to s. 41.26, an internet service provider

---

9 See the excellent review of the "Legal Aspects of Hate Speech" by Lex Gill prepared for the Canadian Commission on Democratic Expression, https://ppforum.ca/wp-content/uploads/2020/07/1.DemX_LegalAspects-EN.pdf. See also Sonja Solomun, Maryna Polataiko, and Helen A. Hayes, "Platform Responsibility and Regulation in Canada: Considerations on Transparency, Legislative Clarity, and Design," *Harvard Journal of Law and Technology (Digest)* 34 (2021): 1-18, https://jolt.law.harvard.edu/digest/platform-responsibility-and-regulation-in-canada-considerations-on-transparency-legislative-clarity-and-design.

10 Natasha Tusikov, "U.K. and Australia move to regulate online hate speech, but Canada lags behind," *National Post*, April 11, 2019, https://nationalpost.com/pmn/news-pmn/u-k-and-australia-move-to-regulate-online-hate-speech-but-canada-lags-behind. See also Natasha Tusikov, *Chokepoints: Global Private Regulation on the Internet*, (Oakland: University of California Press, 2017).

11 The discussion in this section focuses upon hate speech and does not touch upon content inciting violence, terrorist content, child pornography, or non-consensual sharing of intimate images, to which criminal law and administrative law apply.

12 Available in translation as Germany, Network Enforcement Act, section 3, https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2.

13 *Copyright Act*, *Revised Statutes of Canada*, 1985, c. C-42, https://laws-lois.justice.gc.ca/eng/acts/c-42/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

("ISP") that receives notice of a claimed copyright infringement (e.g. illegal downloading of a movie) shall "forward the notice electronically to the person to whom the electronic location identified by the location data specified in the notice belongs and inform the claimant of its forwarding." The ISP shall also "retain records that will allow the identity of the person to whom the electronic location belongs to be determined", typically for a period of six months. There is some evidence that parallel provisions of the *US Digital Millennium Copyright Act* have given rise to "surprisingly high percentages of notices of questionable validity, with mistakes made by both 'bots' and humans."[14] Indeed, the original Canadian version of the notice and notice regime attracted criticism that "notices using threatening language, making outrageous claims of liability, and making offers of settlement that were excessive and required the recipients disclose their personal information" and were thus serving to restrict permitted expression.[15] This ultimately gave rise to amendments in 2018 specifying that notice could not contain such statements and allowing ISPs not to forward notices including pressure of that sort.[16]

Thus, even in advance of establishing any takedown regime for online hate speech, Canada has some experience with the perils of implementing a regime that could lead to an overly broad chilling effect on legitimate postings.

b. All-but non-existent role of government agencies

There are three potential sources of regulatory oversight of online hate speech in Canada: the Canadian Radio-television and Telecommunications Commission (CRTC), provincial and federal human rights commissions, and provincial and federal privacy commissioners. None of these instances have developed a significant role in this domain.

---

14 See Jennifer M. Urban, Joe Karaganis and Brianna L. Schofield, "Notice and Takedown in Everyday
Practice" *UC Berkeley Public Law Research Paper*, No. 2755628, March 24, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628.

15 Michal Jaworski and Athar Malik, "Did You Notice? When A Notice Is Not A Notice Under The Notice And Notice Regime, „March 27, 2019, https://www.mondaq.com/canada/copyright/792094/did-you-notice-when-a-notice-is-not-a-notice-under-the-notice-and-notice-regime.

16 *Copyright Act*, s. 41.25(3).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

For its part, the CRTC has a statement on its website explaining its forbearance from regulating internet content:[17]

> The CRTC does not regulate internet content because consumers can already control access to unsuitable material on the internet using filtering software. Any potentially illegal content on the internet can be addressed with civil action, existing hate crime legislation, and the courts.

It should be noted, however, that the *Broadcasting Distribution Regulations* adopted pursuant the *Broadcasting Act* provide that:[18]

> 8 (1) No licensee shall distribute a programming service that the licensee originates and that contains
> (b) any abusive comment or abusive pictorial representation that, when taken in context, tends to or is likely to expose an individual or group or class of individuals to hatred or contempt on the basis of race, national or ethnic origin, colour, religion, sex, sexual orientation, age or mental or physical disability

In principle, the distribution of such content by a licensee can lead to fines and even to the removal of the license by the CRTC. Although there is currently a bill before Parliament to amend the *Broadcasting Act* that would extend the definition of "broadcasting undertaking" to include "an online undertaking," the CRTC would not be empowered to establish a class of licences for such online undertakings.[19] Thus, the current *Broadcasting Distribution Regulations* would not give the CRTC regulatory authority over hate speech distributed by online undertakings. Nonetheless, the bill would grant discretion to the CRTC to develop conditions applicable to all broadcasting undertakings, including online undertakings, "that the Commission considers appropriate for the implementation of the broadcasting policy set out in subsection 3(1)."[20] In principle the CRTC could

---

17 Canadian Radio-television and Telecommunications Commission (CRTC), "Frequently asked questions," April 1, 2015, https://crtc.gc.ca/eng/faqs.htm.

18 *Broadcasting Distribution Regulations*, SOR/97-555, https://laws.justice.gc.ca/eng/regulations/SOR-97-555/page-3.html#h-1010707.

19 Bill C-10, *An Act to amend the Broadcasting Act and to make related and consequential amendments to other Acts*, 2d sess., 43d Parliament, November 3, 2020, ss. 1(1) and 6(1)(a), https://parl.ca/DocumentViewer/en/43-2/bill/C-10/first-reading#ID0E02B0AA. For a discussion of this proposed legislation see the article by Michael Geist in this collection.

20 Bill C-10, s. 9.1.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

therefore develop a code of conduct parallel s. 8 of the *Broadcasting Distribution Regulations.*

The Canadian Human Rights Commission, for its part, until 2013 oversaw section 13 of the *Canadian Human Rights Act* which then provided (emphasis added):[21]

> 13 (1) It is a discriminatory practice for a person or a group of persons acting in concert to communicate or to cause to be so communicated, repeatedly, in whole or in part by means of the facilities of a telecommunication undertaking within the legislative authority of Parliament, *any matter that is likely to expose a person or persons to hatred or contempt by reason of the fact that that person or those persons are identifiable on the basis of a prohibited ground of discrimination.*
> Interpretation
> (2) For greater certainty, subsection (1) applies in respect of a matter that is communicated by means of a computer or a group of interconnected or related computers, including the Internet, or any similar means of communication, but does not apply in respect of a matter that is communicated in whole or in part by means of the facilities of a broadcasting undertaking.
> Interpretation
> (3) For the purposes of this section, no owner or operator of a telecommunication
> undertaking communicates or causes to be communicated any matter described in subsection (1) by reason only that the facilities of a telecommunication undertaking owned or operated by that person are used by other persons for the transmission of that matter.

This provision was repealed by a private member's bill that notably had the support of the Canadian Civil Liberties' Association largely on the grounds that it had infringed upon free speech.[22] The repeal arose under the previous Conservative government despite the fact that the Federal Court of Appeal had upheld the constitutional validity of the provision.[23]

---

21 *Canadian Human Rights Act*, *Revised Statutes of Canada*, 1985, c. H-6, archived version, https://laws-lois.justice.gc.ca/eng/acts/h-6/section-13-20021231.html#:~:text=13%20(1)%20It%20is%20a,Parliament%2C%20any%20matter%20that%20is.

22 See Joel Webe, "Hate speech no longer part of Canada's Human Rights Act" *National Post*, June 27, 2013, https://nationalpost.com/news/politics/hate-speech-no-longer-part-of-canadas-human-rights-act.

23 *Lemire v. Canada (Human Rights Commission)*, 2014 FCA 18, https://canlii.ca/t/g2x2d.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Since the repeal of s. 13, the Canadian Human Rights Commission has periodically made public statements about the need to address online hate but has not engaged in any enforcement strategy.[24]

Provincial human rights commissions do not have the equivalent to the former s. 13 and have thus had to rely upon general anti-discrimination protections to pursue online hate speech, something they have seldom ever done. In 2015, the then Minister of Justice of Québec proposed legislation that would have added a hate speech regime to the Québec *Charter of Human Rights and Freedoms* giving rise to remedies and enforcement by the Quebec Human Rights Commission.[25] However, the government eventually backed away from the proposal, with the Commission for its part recommending the introduction of such a regime but cautioning against potential overreach.[26] Since that time, and in the wake of a horrific mass shooting at a mosque in Quebec City, the Commission was tasked with conducting a study on hateful acts, notably those motivated by islamophobia, in which it underscored the proliferation of hate speech on the internet.[27] However, the Commission also noted that most cases are not reported to authorities, and for those that are it was typically very difficult to trace the origin of the message to an individual.[28]

For a comment on the consistency of s. 13 with human rights protection, see Pearl Eliadis, "The Controversy Entrepreneurs," *Maisonneuve*, August 20, 2009, https://maisonneuve.org/article/2009/08/20/controversy-entrepreneurs/.

24 See for example, Canadian Human Rights Commission, "Statement – We must do more to curb online hate," January 21, 2021, https://www.chrc-ccdp.gc.ca/en/resources/statement-we-must-do-more-curb-online-hate.

25 Projet de loi 59, *Loi édictant la Loi concernant la prévention et la lutte contre les discours haineux et les discours incitant à la violence et apportant diverses modifications législatives pour renforcer la protection des personnes*, 1er sess., 42e législature, June 10, 2015, http://m.assnat.qc.ca/fr/travaux-parlementaires/projets-loi/projet-loi-59-41-1.html.

26 See Commission des droits de la personne et des droits de la jeunesse, *Mémoire à la Commission des institutions de l'Assemblée nationale*, August 2015, https://www.cdpdj.qc.ca/storage/app/media/publications/memoire_PL59_discours-haineux.pdf.

27 Commission des droits de la personne et des droits de la jeunesse, *Les actes haineux à caractère xénophobe, notamment islamophobe : résultats d'une recherche menée à travers le Québec*, August, 2019, 111, https://www.cdpdj.qc.ca/storage/app/media/publications/etude_actes_haineux.pdf.

28 Commission des droits de la personne et des droits de la jeunesse, *Les actes haineux*, 186 ff. and 249.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

To pursue the Québec example further, section 11 of the Québec *Charter of Human Rights and Freedoms* specifies that:

> No one may distribute, publish or publicly exhibit a notice, symbol or sign involving discrimination, or authorize anyone to do so.

The Québec Human Rights Tribunal has found that a message reading "Landlords Go Home" addressed to persons of Haitian origin violated s. 11 and gave rise to damages.[29] In principle this kind of case suggests a pathway toward civil remedies against online hate if the person at the origin of the message can be identified.

Although Canada's Privacy Commissioner has commented publicly on the issue of online hate speech, the role of the Commissioner's office is currently all but non-existent.[30] However, Lex Gill has noted that to the degree platforms adopt algorithmic approaches to screening online hate:[31]

> content filtering and censorship technology is almost always surveillance technology
> as well. It is therefore rare that the adoption of such measures will not involve at least indirect impacts on users' privacy rights.

Furthermore, the business model of platforms is focused on keeping users engaged (addicted) and collecting as much data from them as possible so as to increase advertising revenue.[32] This business model tends to amplify the spread of extreme content.[33] Under proposed new Canadian legislation, modelled on the *California Consumer Privacy Act*,[34] the Privacy Commissioner and a new Data Protection Tribunal would have significantly augmented enforcement powers, including to impose fines of up to 3% of the organization's yearly gross global revenue. These powers would en-

---

29 *Commission des droits de la personne et des droits de la jeunesse (Coffy et une autre) c. Brisson*, 2009 QCTDP 3, https://canlii.ca/t/22qhm.

30 See Office of the Privacy Commissioner of Canada, "A Data Privacy Day Conversation with Canada's Privacy Commissioner," January 28, 2020, https://www.priv.gc.ca/en/opc-news/speeches/2020/sp-d_20200128/.

31 Gill, "Legal Aspects of Hate Speech," 15.

32 Andrew Burt, "Can Facebook Ever be Fixed?," *Harvard Business Review*, April 8, 2019, https://hbr.org/2019/04/can-facebook-ever-be-fixed.

33 Gilad Edelman, "Social Media CEOs Can't Defend Their Business Model," *Wired*, March 25, 2021, https://www.wired.com/story/social-media-ceo-hearing-cant-defend-business-model/.

34 *California Code*, Title 1.81.5. *California Consumer Privacy Act of 2018*, https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

force a range of requirements touching digital platforms directly, notably that companies must:

- only collect personal information for an appropriate purpose;
- not require consent to the collection of personal information beyond what is needed for the provision of a service;
- not obtain consent by a misleading practice;
- not retain personal information for longer than needed to fulfill the purpose for which it was collected;
- dispose of personal information collected by them from users if users withdraw consent to its use; and
- protect personal information through security safeguards.[35]

The bolstered privacy regime has the potential to disrupt the business model of digital platforms and thus, indirectly, to have an impact on the spread of online hate.[36] This point will be contextualized somewhat further in the fourth part of this article.

c. Criminal law provisions

Canada's *Criminal Code* contains a number of prohibitions touching upon hate speech: advocating genocide (s. 318); publicly inciting hatred (s. 319(1)); and promoting hatred (s. 319(2)).[37] These provisions have been found consistent with constitutional protections of free speech.[38]

Although the dataset is incomplete, Statistics Canada has reported that there were only some 50 police-reported online cases per year across Cana-

---

35 Bill C-11, *Digital Charter Implementation Act, 2020*, 2d sess., 43d Parliament, November 17, 2020, https://parl.ca/DocumentViewer/en/43-2/bill/C-11/first-reading. See also the parallel proposed Quebec legislation, which with respect to the "right to de-indexation" and the "right to be forgotten" goes further than the federal legislation: Projet de loi 64, *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*, 1er sess., 42e legislature, June 12, 2020, http://m.assnat.qc.ca/fr/travaux-parlementaires/projets-loi/projet-loi-64-42-1.html.

36 See Jon Swartz, "California's landmark privacy law is Facebook's next 'nightmare'," *Market* Watch, August 22, 2020, https://www.marketwatch.com/story/californias-landmark-privacy-law-is-facebooks-next-nightmare-2020-08-18.

37 *Criminal Code*, *Revised Statutes of Canada*, 1985, c. C-46, https://laws-lois.justice.gc.ca/eng/acts/c-46/.

38 See notably *R v Keegstra*, [1990] 3 SCR 697, https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/695/index.do.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

da, with no records kept on successful prosecutions.[39] Police forces have noted that they face major resource constraints investigating and prosecuting these crimes, which confront them with significant technical obstacles involving encryption and the generalized use of virtual private networks.[40]

There are two other areas of Canadian criminal law relating to online hate speech. The *Protecting Canadians from Online Crime Act* of 2014 added a new offence prohibiting non-consensual distribution of intimate images including (s. 162.1) as well as complementary amendments to authorize the removal of such images, including child pornography, from the Internet (s. 164.1(5)) and the restriction of the use of a computer or the Internet by a convicted offender (s. 162.2).[41] The *Anti-Terrorism Act of 2015* amended the *Criminal Code* to add a prohibition against counselling another person to commit a terrorism offence (s. 83.221.).[42] Furthermore, a judge may order that "terrorist propaganda" available to the public be deleted from a computer system (s. 83.223).

d. The Canada-U.S.-Mexico Trade Agreement

Article 19.17 of the *Canada-U.S.-Mexico Trade Agreement* (in force 2020) all but extends Section 230 of the *U.S. Communications Decency Act* to Canada by providing:[43]

---

39 House of Commons, Standing Committee on Justice and Human Rights, *Taking Action to End Online Hate*, June, 2019, 21, https://www.ourcommons.ca/Content/Committee/421/JUST/Reports/RP10581008/justrp29/justrp29-e.pdf. A review of the caselaw on www.canlii.org reveals 5 successful reported prosecutions over the last 5 years.

40 See Canadian Commission of Democratic Expression, *Harms Reduction: A Six-Step Program to Protect Democratic Expression Online*, Public Policy Forum, January, 2021, 21, https://ppforum.ca/wp-content/uploads/2021/01/CanadianCommissionOnDemocraticExpression-PPF-JAN2021-EN.pdf.

41 *Protecting Canadians from Online Crime Act*, *Statutes of Canada*, 2014, c. 31, https://laws-lois.justice.gc.ca/eng/annualstatutes/2014_31/FullText.html.

42 *Anti-Terrorism Act of 2015*, *Statutes of Canada*, 2015, c. 20, https://laws-lois.justice.gc.ca/eng/annualstatutes/2015_20/page-3.html#h-20.

43 *Canada-U.S.-Mexico Trade Agreement*, Article 19.17.1, https://www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/cusma-aceum/text-texte/19.aspx?lang=eng. By contrast, *U.S. Communications Decency Act*, *U.S. Code* 47 (2018) § 230(c) provides: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

> no Party shall adopt or maintain measures that treat a supplier or user of an interactive computer service as an information content provider in determining liability for harms related to information stored, processed, transmitted, distributed, or made available by the service, except to the extent the supplier or user has, in whole or in part, created, or developed the information.

This entails that Canada cannot treat online platforms as providing user-generated content and treat them as liable for harms caused by that content.[44]

There is some question as to how much this provision will constrain any new Canadian online hate speech regime. Krishnamurthy and Fjeld have argued that by contrast with Section 230, which bars all causes of action against a platform that treat it as "publisher or speaker" of hosted information, Article 19.17 only excludes it from being held "liable," thus according to them opening the possibility of "equitable" remedies including restraining orders and injunctions.[45] Notice and takedown would thus be possible.

It should be added that a regulatory framework applying to online platforms, including, for example a duty to act responsibly in overseeing its own community standards, which might be accompanied by enforcement powers including fines, should not in principle run afoul of Article 19.17 as long as it does not impose liability on platforms (i) in the same way as it does to content providers and (ii) for steps taken by platforms themselves to control "harmful or objectional" content.

The former point suggests that the CRTC should be cautious before using any new powers acquired pursuant to Bill C-10 simply to extend to online platforms the regime applicable to broadcasting licensees under the *Broadcasting Distribution Regulations*.[46] Even were it to do so, however, it is arguable that those regulations give rise to equitable remedies rather than to a liability regime.

---

44 Article 19.17.4 makes clear that the Article does not apply to the protection of intellectual property rights or to the enforcement of criminal law.

45 For a detailed discussion, see Vivek Krishnamurthy and Jessica Fjeld, "CDA 230 Goes North American? Examining the Impacts of the USMCA's Intermediary Liability Provisions in Canada and the United States," *SSRN*, July 7, 2020, https://ssrn.com/abstract=3645462.

46 See *supra* notes 17 to 20 and accompanying discussion.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

This latter point is underscored by Article 19.17.3:

> No Party shall impose liability on a supplier or user of an interactive computer service on account of:
> (a) any action voluntarily taken in good faith by the supplier or user to restrict access to or availability of material that is accessible or available through its supply or use of the interactive computer services and that the supplier or user considers to be harmful or objectionable; or
> (b) any action taken to enable or make available the technical means that enable an information content provider or other persons to restrict access to material that it considers to be harmful or objectionable.

This provision emphasizes immunity from liability for platforms on what might be called a "Good Samaritan" basis: where platforms make good faith efforts to control harmful or objectional material, they should not be held liable for those actions. However, if for example a regulatory requirement is imposed on platforms to take steps to restrict access to harmful or objectional material, and the platform fails to comply, it could not invoke Article 19.17.3 as a shield. It would no longer be operating in the realm of voluntary corporate social responsibility: it would be subject to legal constraints.

II. Reports on directions for law reform

The review of existing Canadian law makes clear that at best, Canada has a limited range of tools to address online hate speech and nothing resembling an overall legal framework to ensure that democratic expression is not undermined by the existence of filter bubbles that can concentrate and reinforce extreme expression.[47] A survey conducted by the Canadian Race Relations Foundation in January, 2021 found that "93% of Canadians believe that online hate speech and racism are a problem, including 49 percent who believe online hate speech and racism are very serious prob-

47 See Eli Paliser, *The filter bubble: what the Internet is hiding from you*, (New York: Penguin, 2011). See also Daniel Kilvington, "The virtual stages of hate: Using Goffman's work to conceptualise the motivations for online hate," *Media, Culture & Society* 43, no. 2 (2020): 256-272, https://journals.sagepub.com/doi/10.1177/0163443720972318.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

lems."[48] Furthermore, "the majority of Canadians—at least 60 percent—believe that the federal government has an obligation to put forward regulation to prevent the spread of hateful and racist rhetoric and behaviour online," and "nearly 80 percent of Canadians said they would support regulation that would require social media companies to remove hateful or racist content from their platforms within 24 hours of it being posted."

Not surprisingly, therefore, there have been a number of recent prominent reports and consultation papers paving the way for legislative reform. Six of them are singled out here for review: a) *Taking Action to End Online Hate*, the 2019 Report of the House of Commons Standing Committee on Justice and Human Rights; b) C*anada's communications future: Time to act*, the 2020 Report of the Broadcasting and Telecommunications Legislative Review Panel; c) *Defamation Law in the Internet Age*, the 2020 Report of the Law Commission of Ontario; d) *Recommendations to Strengthen Canada's Response to New Digital Technologies and Reduce the Harm Caused by their Misuse*, the 2021 Report of the Canadian Citizens' Assembly on Democratic Expression, and e) the *Harms Reduction: A Six-Step Program to Protect Democratic Expression*, the 2021 companion Report of the Canadian Commission on Democratic Expression.[49]

---

48 Canadian Race Relations Foundation, "Poll demonstrates support for strong social media regulations to prevent online hate and racism," January 25, 2021, https://www.crrf-fcrr.ca/en/news-a-events/media-releases/item/27349-poll-demonstrates-support-for-strong-social-media-regulations-to-prevent-online-hate-and-racism.

49 Two additional relevant reports are not treated in detail here. The wide-ranging 2018 Report of the House of Commons Standing Committee on Access to Information, Privacy and Ethics entitled *Democracy under Threat: Risks and Solutions in the Era of Disinformation and Data Monopoly*, Our Commons https://www.ourcommons.ca/Content/Committee/421/ETHI/Reports/RP10242267/ethirp17/ethirp17-e.pdf touched upon the issue of notice and takedown and recommended (at 3) that "to remove manifestly illegal content in a timely fashion, including hate speech, harassment and disinformation, or risk monetary
sanctions commensurate with the dominance and significance of the social platform, and allowing for judicial oversight of takedown decisions and a right of appeal." See also Jacob Davey, Mackenzie Hart, Cécile Guerin, ed. Jonathan Birdwell, *Interim Report: An Online Environmental Scan of Right-wing Extremism in Canada, The Institute for Strategic Dialogue*, June 19, 2020, https://www.isdglobal.org/wp-content/uploads/2020/06/An-Online-Environmental-Scan-of-Right-wing-Extremism-in-Canada-ISD.pdf. The Report "identified 6,660 right-wing extremists channels, pages, groups and accounts across 7 social media platforms" operating in Canada (at 5).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a. Taking Action to End Online Hate

The Standing Committee on Justice and Human Rights ("Committee"), which was composed of six voting members of the Liberal government, three from the Conservative opposition and one from the NDP opposition, chose to frame its Report with a pointed quotation from the reasons of Rothstein J. in the Supreme Court of Canada's *Walcott* decision:[50]

> Hate speech is not only used to justify restrictions or attacks on the rights of protected groups on prohibited grounds … hate propaganda opposes the targeted group's ability to find self-fulfillment by articulating their thoughts and ideas. It impacts on that group's ability to respond to the substantive ideas under debate, thereby placing a serious barrier to their full participation in our democracy. Indeed, a particularly insidious aspect of hate speech is that it acts to cut off any path of reply by the group under attack. It does this not only by attempting to marginalize the group so that their reply will be ignored: it also forces the group to argue for their basic humanity or social standing, as a precondition to participating in the deliberative aspects of our democracy.

A concern with the rise in hate crimes reported by the police as well as with the connection between online hate and acts of violence led the Committee to initiate a study in March of 2019. The Committee, chaired by Anthony Housefather of the Liberal Party, heard from forty groups and organizations, including Facebook, Twitter and Google, as well as nine individuals. The result was a report culminating in nine recommendations. The Conservative Party members of the Committee dissented from the Report[51] and the New Democratic Party ("NDP") issued a supplementary report essentially endorsing the general direction taken but proposing some further detail.[52] Five of these recommendations concerned ways to improve existing mechanisms for combatting online hate, placing emphasis upon improved funding for training of police, crown attorneys and judges, better collection of data on hate crimes including via a national database on hate crimes and hate incidents, facilitation of reporting and

50 See *Saskatchewan (Human Rights Commission) v. Whatcott*, [2013] 1 SCR 467 at 507, https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/12876/index.do?q=whatcott. Cited by House of Commons, *Taking Action to End Online Hate*, 5.

51 House of Commons, *Taking Action to End Online Hate*, 55-6.

52 House of Commons, *Taking Action to End Online Hate*, 57-61.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

public education. Without recommending specific language, the Committee urged:[53]

> That the Government of Canada formulate a definition of what constitutes 'hate' or 'hatred' that is consistent with Supreme Court of Canada jurisprudence. It is critical that this definition acknowledges persons who are disproportionately targeted by hate speech including but not limited to racial, Indigenous, ethnic, linguistic, sexual orientation, gender identity, and religious groups.

The Committee also recommended that a civil remedy be established, perhaps by reinstating s. 13 of the *Canadian Human Rights Act* or some analogous measure.[54] The Committee favoured the implementation of a timely notice and takedown regime, with platforms required to "make it simple for users to flag problematic content," but did not specify applicable standards.[55] The NDP for its part favoured a "manifestly illegal" standard like that adopted in Germany's NetzDG, as well as "monetary sanctions commensurate with the dominance and significance of the social platform, and allowing for judicial oversight of takedown decisions and a right of appeal."[56] The Report placed considerable emphasis upon transparency, recommending common standards for platform reporting mechanisms and a duty to report regularly to users concerning incidents reported, actions taken, and the speed of response, which significant monetary penalties for failure to report.[57] Finally, the Report signalled support for an effort to enhance the authentication of online content by recommending:[58]

> That online platforms be encouraged to provide optional mechanisms to authenticate
> contributors and digitally sign content, and couple this with visual indicators signifying
> that given user or content is authenticated, and provide users options for filtering nonsigned or non-authenticated content.

---

53 House of Commons, *Taking Action to End Online Hate*, 41.
54 House of Commons, *Taking Action to End Online Hate*, 41.
55 House of Commons, *Taking Action to End Online Hate*, 42.
56 House of Commons, *Taking Action to End Online Hate*, 61. Note that this recommendation tracked the language House of Commons, *Democracy under Threat Report*, 349.
57 House of Commons, *Taking Action to End Online Hate*, 42.
58 House of Commons, *Taking Action to End Online Hate*, 42.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

It is worth noting, finally, as the Report itself underscored, that in its presentation to the Committee, Facebook supported "the establishment of clear baseline standards applicable to all platforms would help to counter online hate" since "people use many different online platforms to communicate".[59]

b. *Canada's communications future: Time to act*

Although the Report of the Broadcasting and Telecommunications Legislative Review Panel was mainly of significance for the preparation of Bill C-10,[60] the Report did address the fact that "[o]nline platforms have … created forums that enable the dissemination of harmful content, fake news and disinformation, and violent and extremist content."[61] One of the core recommendations, taken up in Bill C-10, was to ensure that the CRTC can impose codes of conduct "regarding all media content undertakings" – including online platforms.[62] The Report also formulated a specific recommendation about liability for harmful content:[63]

> We recommend that the federal government introduce legislation with respect to liability of digital providers for harmful content and conduct using digital technologies, separate and apart from any responsibilities that may be imposed by communication legislation. Given that the challenges in this area are global in nature, we also encourage the federal government to continue to participate actively in international fora and activities to develop international cooperative regulatory practices on harmful content.

---

59 House of Commons, *Taking Action to End Online Hate*, 27.

60 Bill C-10.

61 Broadcasting and Telecommunications Legislative Review Panel, Final Report, *Canada's communications future: Time to act*, January, 2020, https://www.ic.gc.ca/eic/site/110.nsf/vwapj/BTLR_Eng-V3.pdf/$file/BTLR_Eng-V3.pdf. The Panel, appointed by the federal government, was chaired by Janet Yale, former CEO of the Canadian Cable television Association and a former Director General of the CRTC. Other panelists included lawyers Peter Grant, Hank Intven, and Monica Song, academics Marina Pavlović and Pierre Trudel, and Monique Simard, who had been CEO of the Société de développement des entreprises culturelles.

62 Broadcasting and Telecommunications Legislative Review Panel, 10 and 34.

63 Broadcasting and Telecommunications Legislative Review Panel, 37 as well as discussion at 190-193.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Finally, it made a parallel recommendation with respect to illegal content and conduct:[64]

> We recommend that the federal government regularly review the efficiency of enforcement mechanisms for monitoring and removing illegal content and conduct found online. Given the diverse range of governing frameworks for these matters in Canada, we encourage the federal government to coordinate with provincial and territorial governments.

The Report has thus set the stage for a greater role for the CRTC in applying codes of conduct to online platforms but has at the same time envisaged a separate "liability" regime for online platforms (Article 19.17 of Canada-U.S.-Mexico Trade Agreement was not referenced) and placed considerable emphasis upon inter-governmental coordination.

c. Defamation Law in the Internet Age

The Law Commission of Ontario spent four years studying "how best to reform defamation law in response to the social and technological revolution in written communications brought about by the internet."[65] There is, of course, some significant overlap between defamatory speech and hate speech. The Commission made clear that it had "explored the role of defamation law in relation to an array of legal tools for regulating online speech in the 21st century" including "myriad laws directed at particular types of harmful speech, such as child pornography and hate speech."[66] Nonetheless, the Commission excluded "direct examination of these related areas of law."[67]

Taken as a whole, the Report contains three chapters with significant implications for the regulation of online hate speech and the protection

---

64 Broadcasting and Telecommunications Legislative Review Panel, 193.

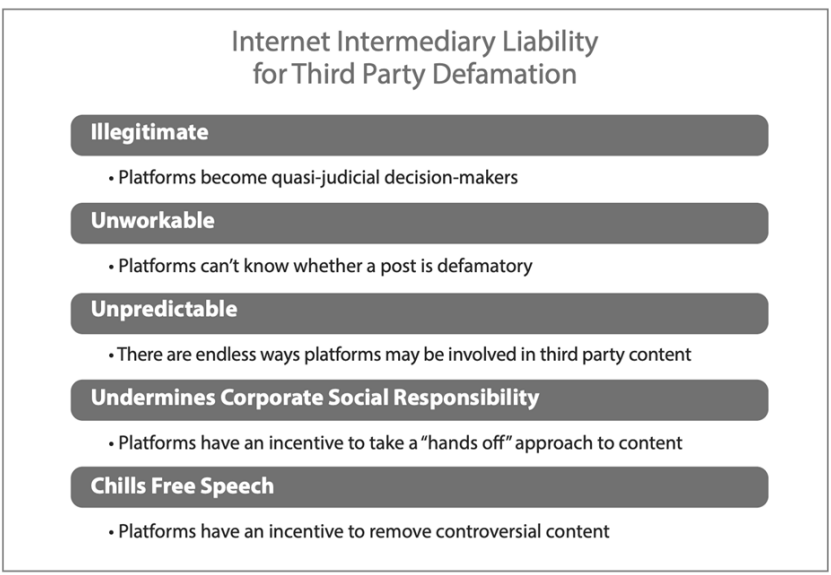
65 Law Commission of Ontario, *Defamation Law in the Internet Age*, March 2020 at 1, https://www.lco-cdo.org/wp-content/uploads/2020/03/Defamation-Final-Report-Eng-FINAL-1.pdf. The Law Commission of Ontario was originally created by the Ontario Ministry of the Attorney General, the Law Foundation of Ontario, the Law Society of Ontario, Osgoode Hall Law School and the Law Deans of Ontario and is now funded by the Law Foundation of Ontario, the Law Society of Ontario, Osgoode Hall Law School, and York University: "Learn about us," https://www.lco-cdo.org/en/learn-about-us/.

66 Law Commission of Ontario, *Defamation Law in the Internet Age*, 15.

67 Law Commission of Ontario, *Defamation Law in the Internet Age*, 15.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of democratic expression: chapters on new legal responsibilities for intermediary platforms, notice and takedown, and online dispute resolution. As to legal responsibilities, the Commission came to the conclusion that the existing common law framework imposing liability for defamation upon publishers is "clumsy and overinclusive when applied to the unique functioning of the internet."[68] The Commission produced an interesting table summarizing the reasons why platforms should not be liable for defamation:[69]



Arguably each of these rationales applies as well to other forms of harmful content including hate speech. It is striking that the Commission was unimpressed by existing and evolving quasi-judicial processes established by platforms such as the Facebook Oversight Board.[70] In the end, it rec-

68 Law Commission of Ontario, *Defamation Law in the Internet Age*, 74.
69 Law Commission of Ontario, *Defamation Law in the Internet Age*, 77.
70 See Facebook Oversight Board, https://oversightboard.com/. The Commission observed that "Although Facebook's Oversight Board contemplates some adjudicative elements, it is probably not a promising model of [online dispute resolution] ODR in the absence of a more direct focus on the interests of the parties. Furthermore, it does not contemplate any supervisory role for government." Law Commission of Ontario, *Defamation Law in the Internet Age*, 103.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ommended that online platforms should be excluded from the category "publisher" by defining that term to "to require an intentional act of communicating a specific expression."[71]

The Commission also recommended that there be "a takedown obligation on intermediary

platforms hosting third party content available to users in Ontario."[72] Notice of complaint would be forwarded by the platform to the publisher of the allegedly defamatory material. The publisher would then have two days to respond in writing. Where a response was received by the platform within the prescribed period, that response would be forwarded to the complainant and no further action would be taken, since the platform would be given no role in assessing the merits of the complaint. On the other hand, if no response was forthcoming, the platform would be required to take down the specific language that is alleged to be defamatory. Notice of takedown would be provided to the publisher, who could require put-back if "there is evidence that the publisher failed to receive the notice or unintentionally missed the deadline and where it is technologically reasonable to do so."[73] Regulations would specify an administrative fee that platforms could charge to complainants. Failure by the platform to comply with takedown requirements would entitle the complainant to statutory damages. The Ontario Superior Court of Justice, the court of general jurisdiction, would enforce the notice and takedown regime.

Interestingly, the Commission specified that the regime would only apply to platforms hosting content available in Ontario and recommended excluding search engines from its ambit. Since the publication of the Report, Google has come under considerable pressure to intervene to prevent websites from running a successful business involving the publication of defamatory material that appears high in Google searches, in turn allowing these websites to charge thousands of dollars to victims to take the posts down.[74] Google has announced that it will change its search algorithm to prevent predatory websites from appearing in the list of results when

71 Law Commission of Ontario, *Defamation Law in the Internet Age*, 80 and 109.
72 Law Commission of Ontario, *Defamation Law in the Internet Age*, 96 and 109.
73 Law Commission of Ontario, *Defamation Law in the Internet Age*, 96 and 109.
74 See Adam Krolik and Kashmir Gill, "The Slander Industry," *New York Times*, April 24, 2021, https://www.nytimes.com/interactive/2021/04/24/technology/online-slander-websites.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

someone searches for a person's name, and that it has created a "known victims" service for those who report having been attacked by sites that charge for the removal of posts.[75] After a report to the "known victims" service, Google will suppress similar content when someone searches for a victim's name. This represents an important departure for Google, since it had heretofore taken the position, parallel to that of the Commission, that "[w]e never touch search, no way, nohow."[76]

The Commission was aware that its recommendations on notice and takedown arose in a context where there was increasing pressure on the federal government to enact a takedown regime for "manifestly illegal content," which could include defamation.[77] While the Commission eschewed taking any position on the merits of proposals such as those contained in the *Taking Action to End Online Hate* Report, it did signal the relevance to the debate of proposals to create a statutory duty of care, underscoring in particular the importance of the 2019 UK *White Paper on Online Harms*.[78]

Finally, while the Commission expressed deep skepticism about the development of online dispute resolution by platforms themselves, characterizing them as having "few of the hallmarks of procedural fairness and none of the authoritativeness of a judicial decision,"[79] it did consider a "co-regulatory approach"[80] as well as the possibility of creating social media councils to be a "multi-stakeholder accountability mechanism for platform

75 See Kashmir Hill and Daisuke Wakabayashi, "Google Seeks to Break Vicious Cycle of Online Slander," *New York Times*, June 10, 2021, https://www.nytimes.com/2021/06/10/technology/google-algorithm-known-victims.html.

76 Kashmir Hill and Daisuke Wakabayashi.

77 Law Commission of Ontario, *Defamation Law in the Internet Age*, 95.

78 Law Commission of Ontario, *Defamation Law in the Internet Age*, 95. See United Kingdom, *White Paper on Online Harms*, April 6, 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

79 Law Commission of Ontario, *Defamation Law in the Internet Age*, 102.

80 Law Commission of Ontario, *Defamation Law in the Internet Age*, 103. The Commission referenced the "right to be forgotten" regime of the European Union, Article 29 Data Protection Working Party, "Guidelines on the Implementation of the Court of Justice of the European Union Judgment on '*Google Spain and Inc v Agencia Española de Protección de Datos* (*AEPD*) *and Mario Costeja González*'," November 26, 2014, C-131/12, https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=64437. It also referenced the EU *Directive on Copyright in the Digital Single Market*, April 17, 2019), 2019/790, https://eur-lex.europa.eu/eli/dir/2019/790/oj.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

content decisions."[81] In the end the Commission took a favourable view of social media councils because:[82]

- A social media council would likely be able to address multi-jurisdictional dispute more effectively than a government-created [online dispute resolution] ODR tribunal;
- A social media council would operate within the contractual relationship between platforms and their users, thereby binding publishers to the process.
- Techno-legal remedies such as red-flags and the modulation of views could be directly implemented by the platform.

Nonetheless, because it judged that the subject of social media councils went beyond the scope of its mandate, Commission chose to make a recommendation only calling for the future exploration by the Ontario government of online dispute resolution, including by means of social media councils or other regulatory models.[83]

d. Report of the Citizens' Assembly on Democratic Expression

In 2020, the Public Policy Forum with Funding from the McConnell Foundation and the Government of Canada launched an ambitious three-year initiative to study how to strengthen Canadian democracy in response to the ubiquitous presence of online technologies. In its first year, the goal was to develop a plan on how to mitigate the negative effects on Canadian democracy of online hate, disinformation and other forms of harmful content while encouraging the broadest possible application of the freedom of expression in Canada's *Charter of Rights and Freedoms*. The first year of the initiative involved two parallel and innovative processes: creating a blue-ribbon Canadian Commission on Democratic Expression, the Report of which is discussed in the next section of this article; and convening a Canadian Citizen's Assembly on Democratic Expression, made up of

81 Law Commission of Ontario, *Defamation Law in the Internet Age*, 103. The Commission referenced in particular the work of Article 19, *The Social Media Councils: Consultation Paper*, June, 2019, 7, https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

82 Law Commission of Ontario, *Defamation Law in the Internet Age*, 104.

83 Law Commission of Ontario, *Defamation Law in the Internet Age*, 104 and 110.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

representative body of 42 Canadians, which authored its own Report that came out almost contemporaneously with that of the Commission.[84]

Over 12,000 Canadians were invited to serve in the Assembly, nearly 400 volunteered and in the end 42 were selected at random to "represent the widest possible range of voices and perspectives."[85] The Assembly was convened in March 2020 and its members met eighteen times, hearing from over a dozen experts and from senior representative of Google and Facebook.[86]

The Report of the Assembly contains notably a set of guiding values,[87] a set of key concerns,[88] and 33 recommendations grouped largely around the key concerns.

The recommendations were far-reaching and served to demonstrate perhaps that informed non-experts can generate fresh proposals that at the very least provided a valuable stress test for the proposals developed in parallel by the Canadian Commission on Democratic Expression.[89] Key among them were creating a new digital platforms regulator, becoming more savvy at international cooperation, establishing user ownership of personal data, introducing user-friendly standardized descriptions of terms of service across platforms, and making anonymous users accountable for their actions.[90]

---

84 Canadian Citizens' Assembly on Democratic Expression, *Recommendations to Strengthen Canada's Response to New Digital Technologies and Reduce the Harm Caused by their Misuse*, Public Policy Forum, January, 2021, https://ppforum.ca/wp-content/uploads/2021/01/CanadianCitizens%E2%80%99AssemblyOnDemocraticExpression-PPF-JAN2021-EN.pdf.

85 Canadian Citizens' Assembly on Democratic Expression, 5. An overview of the representativeness of the Assembly is provided at 9 and the assembly process is described in detail at 22-28.

86 Canadian Citizens' Assembly on Democratic Expression, 7.

87 Canadian Citizens' Assembly on Democratic Expression, 32-33. The Assembly sought an Internet that is 1) accessible; 2) accountable; 3) reliable; 4) safe and secure; and 5) amplifies diverse voices.

88 Canadian Citizens' Assembly on Democratic Expression, 34-37. The Assembly focussed on 1) lack of oversight, transparency and accountability of digital platforms; 2) the spread of misinformation; 3) protecting digital rights and user control; and 4) harms to vulnerable persons and minority groups.

89 The Assembly presented its recommendations to the Commission in advance of the Commission issuing its Report. See Canadian Citizens' Assembly on Democratic Expression, at 28.

90 Canadian Citizens' Assembly on Democratic Expression, 37-42.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

As regards the digital platforms regulator, the Assembly wanted it to:

- develop a national code of online conduct;
- require compliance with principles of responsible algorithmic development and algorithmic transparency;
- levy fines for contravention of relevant laws and regulations;
- require independent compliance audits;
- require enhancement of content moderation policies;
- regulate the use and labelling of bots;
- establish e-courts to adjudicate complaints of harmful speech;
- regulate the collection, storage and sale of data related to underage users; and
- create mechanisms for public participation, including citizens committees.

As regards international cooperation, the Assembly urged strategic cooperation with democratic countries to establish common practices, adoption by Canada of certain existing frameworks, such as the EU GDPR, and the enhancement of collaborative competition law enforcement.

As regards user ownership of personal data, the Assembly challenged platforms *inter alia*:

- to grant users more control over settings influencing content, notably the option only to display content from verified users and credible sources;
- to seek consent for continued collection of data regularly with the option to download and/or fully delete all user data; and
- to delete user data when consent is not obtained or after a set period of time.

It also sought to enshrine user ownership of data in law, policies and regulations.

As regards standardized interface and descriptions of terms of service across platforms, the Assembly sought that these include clear descriptions of i) user rights, ii) information being collected and iii) how it is used and stored, as well as iv) data controls and permissions.

Finally, as regards anonymous users, the Assembly affirmed that anonymity is not a right. The Assembly sought the development of policies, laws and regulations to ensure that anonymity cannot be used to shield individuals from the consequences of producing harmful, hateful, or defamatory speech.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

e. Canadian Commission on Democratic Expression

The seven distinguished members including of the Canadian Commission on Democratic Expression included Beverley McLachlin, former Chief Justice of the Supreme Court of Canada.[91] The Commission at a number of points drew explicitly upon the work of the Assembly and indeed met twice with the members of the Assembly during the course of its own deliberations.[92]

The Commission considered and, in the end, recommended against following either a hands-off self-regulation approach or the German NetzDG notice and takedown approach. It opted instead for a regulatory regime that would include six interconnected elements.

The **first element** was a new legal duty placed upon platforms to act responsibly. The virtue of this was said to be that it provides a "regulatory focus on systemic issues" rather than having regulatory intervention into thousands of content disputes.[93] The idea was drawn in significant degree from the UK *White Paper on Online Harms*, which has proposed imposing a statutory duty of care on platforms.[94] The idea was to impose on companies the onus to fulfil this legal duty, with the regulator positioned to set out how to do this in codes of practice.

However, the UK White Paper also recommended that "If companies want to fulfil these duties in a manner not set out in the codes, they will have to explain and justify to the regulator how their alternative approach will effectively deliver the same or greater level of impact."[95] It was not

---

91 The other members were Rick Anderson, Principal, Earnscliffe Strategy, Julie Caron-Malenfant, Director General, Institut du Nouveau Monde, Adam Dodek, Dean, Faculty of Law (Common Law Section), University of Ottawa, Amira Elghawaby, Journalist and Human Rights Advocate, Jameel Jaffer, Executive Director, Knight First Amendment Institute at Columbia University, and Jean La Rose, Former CEO, Aboriginal Peoples Television Network. See Canadian Commission of Democratic Expression, 46-47.

92 Canadian Commission of Democratic Expression, 51.

93 Canadian Commission of Democratic Expression, 31-2.

94 See United Kingdom, *White Paper on Online Harms*, 7. The Commission noted that "[t]he United Kingdom, within its own legal traditions, is currently advancing a similar type of duty of care for online platforms": Canadian Commission of Democratic Expression, 31. Since Québec is not a common law jurisdiction and does not include duty of care analysis as part of the law of extra-contractual obligations, it is understandable that the Commission adopted the idea of a "duty to act responsibly," which arguably can be applied within both common law and civil law contexts.

95 See United Kingdom, *White Paper on Online Harms*, 7.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

made clear in Commission's Report whether platforms would be allowed to depart from codes of conduct.

It should be noted that the proposed new statutory duty gave rise to a partially dissenting minority report. Commission member Jameel Jaffer wrote:[96]

> I find it difficult to endorse the proposed Duty to Act Responsibly when the content of that duty is left almost entirely to Parliament and the new regulator to decide. Defining the duty will require difficult tradeoffs, not only between free speech and other values—for example, privacy, equality, and due process—but also between different conceptions of free speech.

This point is particularly striking in light of Heritage Minister Steven Guilbeault's perhaps incautious public statement that "hurtful" speech could be included within the scope of what is to be regulated in an eventual bill.[97]

The **second element** in the Commission's proposed regime was a new regulator to oversee and enforce the duty to act responsibly. The goal of creating such a regulator would be to move content moderation and platform governance beyond the exclusive preserve of the platforms. "The regulator would oversee a code of conduct to guide the actions of parties under its supervision, while recognizing that not all platforms can be treated in the same manner."[98] The Commission also sought to ensure that the regulator would be able to impose significant fines and even pursue imprisonment for platform executives.[99]

The **third element** was a social media council to serve as an accessible forum in reducing harms and improving democratic expression on the internet. The social media council would be conceived as "an independent, stakeholder-based body with dedicated professional support that is attached to the regulator."[100] It would serve as a consultative body for the regulator on codes of conduct and on how changing technology, business models and user experience affect policy.

---

96 Canadian Commission of Democratic Expression, 48.

97 Michael Geist, "The real consequences of Steven Guilbeault's battle with the web giants," *Maclean's*, May 3, 2021, https://www.macleans.ca/opinion/the-real-consequences-of-steven-guilbeaults-battle-with-the-web-giants/.

98 Canadian Commission of Democratic Expression, 9.

99 Canadian Commission of Democratic Expression, 33.

100 Canadian Commission of Democratic Expression, 33.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The idea of a social media council had been raised by the Law Commission of Ontario[101] and could perhaps interact with online dispute resolution already being established by the platforms themselves and in some cases across platforms (such as the Global Internet Forum to Counteract Terrorism).[102] A difficult question concerns how the social fractures that are evident on social platforms would be represented on the social media council. Surely it would damage the credibility of such a council to have its membership swing radically according to political winds.

The **fourth element** was a "world-leading transparency regime" to provide the flow of necessary information to the regulator and social media council.[103] The Commission envisaged i) periodic public risk assessment reports from the platforms, ii) power granted to the regulator to compel access to information, notably to the black box of platform algorithms, iii) disclosure rules on data sharing, iv) rules on advertising transparency, v) public labelling and registration of bots, and vi) disclosure of the ownership structure behind those disseminating user-generated and other third-party content.[104]

The **fifth element** was an e-tribunal to facilitate and expedite dispute resolution and a process for addressing complaints swiftly and lightly before they become disputes. The key idea would be to: [105]

> allow for the resolution of Canadian disputes within Canada. Currently, with content moderation under the control of platform companies, the training and domicile of the content moderators is a black box. Meanwhile, Facebook's new Oversight Board hears only a handful of global cases and has no Canadian member.

This was another point on which Jameel Jaffer dissented, writing:[106]

> I am not persuaded, though, that establishing a new tribunal system with a broad mandate would be preferable to requiring large platforms themselves to establish, at their own expense, review and appeals processes that are more efficient and transparent than the ones some of them have already established. Before endorsing the proposed e-tribunals, I would want to know more about their mandate, and also

101 See *supra* notes 81 to 83 and accompanying discussion.
102 See Global Internet Forum to Counteract Terrorism, https://gifct.org/.
103 Canadian Commission of Democratic Expression, 9.
104 Canadian Commission of Democratic Expression, 35-6.
105 Canadian Commission of Democratic Expression, 38.
106 Canadian Commission of Democratic Expression, 48.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

about what relationship the proposed tribunals would have to the processes that some of the platforms have already established.

The **sixth element** was a mechanism for quick removal of content that presents an imminent threat to a person. This would constitute an exception to the general avoidance of a notice and takedown mechanism. It would involve a "quick-response system" within 24 hours under the authority of the regulator to ensure the rapid removal– even temporarily – of content that creates a reasonable apprehension of an imminent threat to the safety of the targeted party. The Commission insisted that such decisions should be subject to judicial sanction before either the e-tribunal or the courts.[107]

In addition to these six main recommendations, Commission also flagged a number of further issues including legal liability and fines, law enforcement resources, the interaction of the proposed regime with the Canada-US Mexico Agreement, and the need for periodic review of any legislation eventually adopted.

III. Bill C-36

Although Bill C-36 was not adopted by Parliament before the 2021 election, the fact that the government revealed the proposed language for amendments to the *Criminal Code* and to the *Canadian Human Rights Act* merits discussion. Should the government be re-elected, there is a strong likelihood that the bill will be reintroduced.

a. Definition of hate speech and hatred

The *Taking Action to End Online Hate* Report had called on the government to define hate or hatred so as to acknowledge "persons who are disproportionately targeted by hate speech including but not limited to racial, Indigenous, ethnic, linguistic, sexual orientation, gender identity, and religious groups."[108] This indeed is the approach adopted as far as defining "hate speech" for the purposes of the *Canadian Human Rights Act* is concerned. The bill provides that "***hate speech*** means the content of a

107 Canadian Commission of Democratic Expression, 39.
108 House of Commons, *Taking Action to End Online Hate*, 41.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

communication that expresses detestation or vilification of an individual or group of individuals on the basis of a prohibited ground of discrimination."[109] It also provides that "communication does not express detestation or vilification… solely because it expresses mere dislike or disdain or it discredits, humiliates, hurts or offends."

The bill takes a somewhat different though parallel approach for the purposes of the *Criminal Code*, defining "hatred" as "the emotion that involves detestation or vilification and that is stronger than dislike or disdain" and specifying that "the communication of a statement does not incite or promote hatred … solely because it discredits, humiliates, hurts or offends."[110] Since the *Criminal Code* already includes sanctions against inciting and promoting hatred "against any identifiable group," the Minister presumably concluded that the only purpose of a *Criminal Code* definition was to follow the guidance of the Supreme Court of Canada and orient the courts as to the intensity of the emotion communicated through hate propaganda.[111]

b. Peace bond

Bill C-36 introduces a new peace bond to help forestall hate crimes.[112] Someone who reasonably fears that they could be a target of hate propaganda or criminal mischief could apply for a peace bond to be imposed on an individual to deter that person from committing the crime. Such a peace bond could involve the imposition of conditions including wearing an electronic monitoring device, a curfew, prohibition against consuming drugs or alcohol together with a requirement to provide samples for testing, a prohibition against communicating with any person, and a prohibition against possession of firearms. Surprisingly enough, no specific mention is made of prohibiting visits to or participation in online fora known to convey hate propaganda.

A breach of the proposed peace bond would carry a maximum penalty of four years' imprisonment, the same penalty that exists for breaches of other peace bonds. Consent by the appropriate Attorney General would

---

109 Bill C-36, s. 13.

110 Bill C-36, s. 2. This in effect codifies the approach taken by the Supreme Court of Canada in *Saskatchewan (Human Rights Commission) v. Whatcott* (para. 41).

111 *Criminal Code*, s 319.

112 Bill C-36, s. 3, adding a new s. 810.012 to the *Criminal Code.*

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

be required before the peace bond could be used, as is the case with some existing peace bonds.

c. Canadian Human Rights Act

Rather than simply restoring the former s. 13 of the *Canadian Human Rights Act*, Bill C-36 includes a revised version of it.[113] In addition to the new definition of "hate speech" already signalled, under the proposed legislation the Canadian Human Rights Tribunal would gain new powers 1) to order the party complained against to cease the hate speech and provide redress in consultation with the Canadian Human Rights Commission, 2) to pay damages of up to $20,000 to each victim personally identified in the communication, and 3) to pay a fine of up to $50,000.[114] The bill excludes "private communication" such as private emails or direct messages from the scope of hate speech, and does not apply to online communication service providers. Indeed, in the materials accompanying the release of the bill, the Department of Justice underscored that online platforms "are the focus of upcoming engagement by Canadian Heritage, which will outline a proposed approach to regulating social media and harmful content, including hate speech, online."[115]

The bill would empower the Commission to prevent the disclosure of the identity of the complainant to the person against whom the complaint is filed,[116] and gives further scope to the Tribunal to conduct confidential inquiries where there is a real and substantial risk that a complainant or witness "will be subjected to threats, intimidation or discrimination."[117] Violation of such confidentiality orders made by the Commission or Tribunal would be subject to a fine of up to $50,000.[118]

Since it had been objected that the former s. 13 caused the Commission to flood the Tribunal with an unmanageable caseload, the bill gives stricter

---

113 Bill C-36, s. 13.

114 Bill C-36, s. 19.

115 Department of Justice, "Combatting hate speech and hate crimes: Proposed legislative changes to the *Canadian Human Rights Act* and the *Criminal Code*," June 23, 2021, https://www.justice.gc.ca/eng/csj-sjc/pl/chshc-lcdch/index.html.

116 Bill C-36, s. 14.

117 Bill C-36, s. 17.

118 Bill C-36, s. 20. Unlike other penal offences envisaged under the *Canadian Human Rights Act*, prosecution for these offences would not require prior approval of the Attorney-General of Canada, implying that prosecution should be undertaken as a matter of course.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

guidance to the Commission not to pursue cases where "the complaint indicates no *hate speech*."[119] The bill creates an additional caseload management tool through the possibility of awarding costs for abuse of process.[120] It would also expand the Tribunal so as to address the anticipated increase in workload, adding two to five new members for a maximum of seventeen and eventually twenty members.[121]

d. Ideas not retained

Two ideas raised in the Minister of Justice's *Consultation Paper* were not retained.[122] Individuals will not be empowered to pursue complaints themselves before the Tribunal and thus will rely upon the Commission to initiate complaints. Nor does Bill C-36 remove the requirement that the appropriate Attorney-General provide express consent to prosecutions for alleged wilful promotion of hatred pursuant to s. 319(2) of the *Criminal Code*.

IV. The Digital Citizen Initiative Consultation Papers

The Discussion Guide and Technical Paper issued by the Digital Citizen Initiative of the Ministry of Canadian Heritage on July 26, 2021 have the appearance of internal documents drafted in preparation of legislation. It seems clear that the Minister was in the end reluctant to put forward legislation without a prior public consultation. Thus, the public now has access to the road map offered to the Minister by his civil service together with a narrow set of options presented to him. It remains to be seen whether the public consultation will raise issues and concerns going beyond the scope of the consultation documents, especially given the broader range of options canvassed in the various reports discussed above.

The Discussion Guide is meant to background and justification for the Technical Paper and uses lay language. The Technical Paper has the same structure as the Discussion Paper but uses formulations that could

---

119 Bill C-36, s. 15.
120 Bill, C-36, s.19.
121 Bill C-36, s. 16.
122 Minister of Justice and Attorney General of Canada, *Consultation Paper: Online Hate*, July 14, 2020. https://ocla.ca/wp-content/uploads/2020/07/2020-07-14-Consultation-paper-Online-hate.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

find their way into legislation. This summary focuses primarily on the Technical Paper.

The Technical Paper is divided into two "Modules" that could presumably become two parts of the legislative framework: 1) a new legislative and regulatory framework for social media platforms; and 2) modifications to existing legislation. Not surprisingly, Module 1 is the more elaborate and is itself divided into four parts: A) a general framework of purposes, interpretation and application; B) new rules and obligations; C) new regulators; and D) new regulatory powers and enforcement mechanisms.

Module 1(A) puts forward a set of "premises" about the benefits and potentially harmful impacts of "Online Communication Services" (OCSs).[123] An OCS is defined as "a service that is accessible to persons in Canada, the primary purpose of which is to enable users of the service to communicate with other users of the service, over the internet," but would "exclude services that enable persons to engage only in private communications."[124] The Discussion Paper explains that the definition "is intended to capture major platforms, (e.g., Facebook, Instagram, Twitter, YouTube, TikTok, Pornhub), and exclude products and services that would not qualify as online communication services, such as fitness applications or travel review websites."[125] Nonetheless, it would appear that "private" communication services owned by major platforms, such as Facebook's WhatsApp, and which can be used to spread harmful content, are *prima facie* to be excluded from the application of the legislation. The proposed legislation would give power to the cabinet, upon consultation with the new Digital Safety Commissioner described below, to narrow or extend the definition of OCS by regulation. The legislation would apply to OCSs and to "the closest legal entity to a regulated OCS", called an Online Communication Service Provider (OCSP).[126]

123 Digital Citizen Initiative, "Technical Paper," Module 1, para. 1.
124 Digital Citizen Initiative, "Technical Paper," Module 1, para. 2.
125 Digital Citizen Initiative, "Discussion Guide," "Who and what would be regulated".
126 Digital Citizen Initiative, "Technical Paper," Module 1, para. 6.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The legislation would concern "harmful content" falling into five defined categories:[127]

- child sexual exploitation, 1) as specified in the *Criminal Code*, including child pornography, and 2) material related to child sexual exploitation ("e.g., screen shots of videos that do not include the criminal activity but refer to it obliquely; up-to-date photos of adults who were exploited/ abused as children being posted in the context of their exploitation and abuse as children");
- terrorist content "that actively encourages terrorism and which is likely to result in terrorism";
- content that incites violence, namely "that actively encourages or threatens violence and which is likely to result in violence";
- hate speech as defined in Bill C-36 and "communicated in a context in which it is likely to cause harms identified by the Supreme Court of Canada and in a manner identified by the Court in its hate speech jurisprudence"; and
- non-consensual sharing of intimate images as defined in the *Criminal Code* "with the intent to capture the communication of an intimate image of a person that the person depicted in the image or video did not give their consent to distributing, or for which it is not possible to assess if a consent to the distribution was given by the person depicted in the image or video."

It would thus not extend to defamatory speech. Nor would it extend to misinformation or other "awful but lawful" content.

Module 1(B) proposes creating a new obligation that "an OCSP must take all reasonable measures, which can include the use of automated systems, to identify harmful content that is communicated on its OCS and that is accessible to persons in Canada."[128] It thus comes close to the duty to act responsibly proposed by the Canadian Commission on Democratic Expression. The obligation would extend to abiding by regulations prescribed by the Digital Safety Commissioner and would also require that OCSPs ensure that the measures they take not give rise to differential treatment of any group based on a prohibited ground of discrimination.

Furthermore, "an OCSP must address all content that is flagged by any person in Canada as harmful content" within 24 hours.[129] This means

127 Digital Citizen Initiative, "Technical Paper," Module 1, para. 8.
128 Digital Citizen Initiative, "Technical Paper," Module 1, para. 10.
129 Digital Citizen Initiative, "Technical Paper," Module 1, para. 11.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

either responding to that person that the content does not meet the definition of harmful content or taking down the content. Thus, in this respect the proposed approach follows the NetzDG model despite the recommendation of the Canadian Commission on Democratic Expression not to do so.

Module 1(B) also includes significant transparency and procedural requirements for OCSPs.[130] The flagging mechanism must be "accessible and easy-to-use," as must be the opportunity to make representations and compel prompt review and reconsideration by the OCSP. Upon reconsideration, notice must be given of the recourse to the new Digital Recourse Council of Canada, discussed below. The OCSP must publish "clear content-moderation guidelines," and "must generate and provide reports on a scheduled basis to the Digital Safety Commissioner [also discussed below] on Canada-specific data." The content of these latter reports is to be prescribed in significant detail, including, for example, information from the OCSPs about "how they monetize harmful content". Regulations are envisaged to determine what records OCSPs must keep.

The Technical Paper leaves open two options as to how OCSPs should meet a mandatory notification requirement for law enforcement agencies. The options involve differing thresholds of potential harm. Option (a) would require notification to the Royal Canadian Mounted Police where the OCSP has reasonable grounds to believe that defined "harmful content reflects an imminent risk of serious harm to any person or to property." Option (b) would set the notification requirements by regulation. As the Discussion guide notes: "The legal thresholds (reasonable suspicion, reasonable grounds to believe) for reporting this content… could differ based on the category. For example, the threshold for reporting potentially terrorist and violent extremist content could be lower than that for potentially criminal hate speech."[131] Option (b) would include mandatory reporting of potential terrorist activity to the Canadian Security Intelligence Service. At stake is that police forces are seeking to require platforms to inform them when they take down illegal content and to provide the deleted content to the police as evidence for possible further criminal investigation. The RCMP seeks to hold these materials in a database. Other government partners are resisting this approach, although apparently all stakeholders agree with requiring the platforms to keep the content they

130 Digital Citizen Initiative, "Technical Paper," Module 1, paras. 12-15.

131 Digital Citizen Initiative, "Discussion Guide," "Engaging law enforcement and CSIS".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

remove for a year. Option (a) is a compromise position involving narrower disclosure.

Module 1(C) proposes the establishment of four new bodies: the Digital Safety Commissioner, the Digital Recourse Council of Canada, an Advisory Board and the Digital Safety Commission. The Commissioner would handle the basic administration of the new legislation, including giving general advice to OCSPs (though not about specific content-moderation decisions),[132] establishing an Incident Response Protocol for potential terrorist activity,[133] receive complaints from the public about OCSP non-compliance,[134] and have the power to issue regulations,[135] subject to binding directions from cabinet.[136]

The Digital Recourse Council of Canada would be designed to provide an independent recourse in response to OCSP decisions and would arise only upon exhaustion of remedies available with the OCSP. It is not entirely clear how this would interact, say, with the Facebook Oversight Board, which does not necessarily provide a timely remedy. It is not made entirely clear whether this body would operate as an eCourt. Compliance orders issued by the Digital Safety Commissioner would be appealed to the Personal Information and Data Protection Tribunal created under the proposed overhaul of data privacy legislation.[137]

The Commissioner and Council would be counselled by a new, external Advisory Board having the characteristics of the recommended social media council.[138] Its members would be drawn from civil society, academia, and cultural groups. Its role would be to inform the Council and the Commissioner, both of which would be appointed by the government but independent from ministerial oversight.

The Commissioner, Council and Advisory Board would all operate supported by an umbrella Digital Safety Commission of Canada, which would have a Chief Executive Officer who is not the Commissioner.[139] The entire

---

132 Digital Citizen Initiative, "Technical Paper," Module 1, para. 16.
133 Digital Citizen Initiative, "Technical Paper," Module 1, paras. 18-19.
134 Digital Citizen Initiative, "Technical Paper," Module 1, paras. 40-44.
135 Digital Citizen Initiative, "Technical Paper," Module 1, para. 17.
136 Digital Citizen Initiative, "Technical Paper," Module 1, para. 39.
137 See Bill C-11. See also Digital Citizen Initiative, "Technical Paper," Module 1, para. 81.
138 See Digital Citizen Initiative, "Technical Paper," Module 1, paras. 71-75.
139 See Digital Citizen Initiative, "Technical Paper," Module 1, paras. 60-65.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

apparatus would operate on a cost-recovery basis through charges imposed on the OCSPs.[140]

Module 1(D) proposes a new set of powers and enforcement remedies, including compliance orders against OCSPs and broad inspection powers for the Commissioner, and Administrative Monetary Penalties for non-compliance issued by the Personal Information and Data Protection Tribunal.[141] It would also be an offence for an OCSP to fail to comply with a compliance agreement, adhere to an order issued by the Council or Commissioner, resist or obstruct and inspection, or knowingly make false or misleading statements to the Commissioner or Council.[142] For the most serious offences, the maximum penalty would be a fine not exceeding five percent of gross global revenues in the financial year that precedes the date of sentencing or $25,000,000, whichever is higher. If an OCSP "repeatedly demonstrates persistent non-compliance" with respect to orders for removing content relating to child sexual exploitation or terrorism, the Commissioner could apply to the Federal Court for an order requiring telecommunications service providers to block access to the offending OCS in Canada.[143]

Module 2 proposes certain amendments to existing Canadian legislation. As regards child pornography, the *Mandatory Reporting Act*[144] would be amended so as to extend its application to OCSPs and other internet services, centralize reporting with the National Child Exploitation Crime Centre of the Royal Canadian Mounted Police (RCMP) and generally strengthen its administration.[145] An unresolved issue concerning the reporting of clear child pornography offences has to do with transmission data (i.e., Internet protocol address, date, time, type, origin, destination of the material) or basic subscriber information (BSI) (i.e., customer's name, address, phone number, billing information associated with the IP address). One option is simply to require that such information be shared with the police. The other, stricter, option would require the police to seek

---

140 See Digital Citizen Initiative, "Technical Paper," Module 1, paras. 66-70.
141 See Digital Citizen Initiative, "Technical Paper," Module 1, paras. 81-82, 88-114.
142 See Digital Citizen Initiative, "Technical Paper," Module 1, paras. 119.
143 See Digital Citizen Initiative, "Technical Paper," Module 1, paras. 120.
144 An Act respecting the mandatory reporting of Internet child pornography by persons who provide an Internet service 2011, S.C. c. 4.
145 See Digital Citizen Initiative, "Technical Paper," Module 2, Mandatory Reporting Act, paras. 1-11.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a production order from the court to obtain BSI.[146] A parallel option being considered is to allow the Canadian Security Information Service easier access to BSI in the case of terrorist content, since currently CSIS must seek a warrant, a process that can take four to six months.[147]

V. *Final critical observations*

In my concluding remarks, I would like to signal what I believe are two important conceptual challenges raised by the emerging Canadian approach. The first of these was signalled by Jameel Jaffe in his dissenting report to the Commission.[148] How should a national regulatory approach properly interconnect with the emerging online dispute resolution regimes developed by platforms themselves, notably the Facebook Oversight Board? The second challenge was signalled by the Commission when quoting Prof. Taylor Owen to the effect that the negative effects of social media companies are "baked into their business models."[149] How can regulators retool business models that rely on algorithms that amplify the propagation of extreme content as well on the sweeping collection of personal data that allows platforms to target users with recommended, sometimes extreme, content?

Let me make some affirmations designed to provoke debate. The establishment of the Facebook Oversight Board gives rise to the counter-intuitive conclusion that national regulatory responses should seek to strengthen and widen the reach of online dispute resolution offered by platforms regime rather than simply to substitute for it.[150] I take the point that as of August 19, 2021 the Oversight Board had only rendered fifteen decisions in its inaugural year. But in addition to the much-discussed decision on Donald Trump,[151] the Zwarte Piet decision handed down on April 13, 2021 by the Oversight Board illustrates that a transnational body perhaps

146 See Digital Citizen Initiative, "Technical Paper," Module 2, Mandatory Reporting Act, paras. 7-8.
147 See Digital Citizen Initiative, "Technical Paper," Module 2, Canadian Security Intelligence Service Act, paras. 1-6.
148 Canadian Commission of Democratic Expression, 48.
149 Canadian Commission of Democratic Expression, 12.
150 See a detailed discussion of the Facebook Oversight Board in chapter 1.6. of this volume (Schultz, Mårten: Six Problems with Facebook's Oversight Board).
151 Facebook Oversight Board, Case decision 2021-001-FB-FBR, https://oversightboard.com/decision/FB-691QAMHJ/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

can accomplish something important that a national regulator might be less capable of achieving.[152] That case concerned the removal of a 17 second video showing a child meeting three adults, one dressed to portray "Sinterklaas" (the Dutch version of Santa Claus) and two portraying "Zwarte Piet," also referred to as "Black Pete," who in the Dutch Christmas tradition accompanies Sinterklaas during the Feast of Saint Nicholas, distributing sweets. The video in question was posted to document this event. The two adults portraying Zwarte Piet had their faces painted black, wore Afro wigs under hats and colourful renaissance-style clothes. All the adults and the child in the video appeared to be white, including those with their faces painted black. The Board conducted a sophisticated analysis grounded notably in the UN Guiding Principles on Business and Human Rights and in human rights standards including Article 19 of the International Covenant on Civil and Political Rights. It considered the Zwarte Piet "tradition" in a comparative perspective and noted that it could be practiced without blackface or racial stereotypes. The majority upheld the removal of the content but concluded that Facebook had not sufficiently notified users about its community standard.

The point is that a purely Dutch body may or may not have been willing to put the Dutch tradition in comparative context and to consider the matter from the vantage point of how the stereotypes depicted were to be perceived in the context of global communications. Consequently, I would suggest that Canada consider the following points in addition to the direction being apparently being considered:

1) Promulgate a standard promoting the establishment of internal appeal bodies parallel to the Oversight Board for other OCSPs;
2) Promulgate a standard ensuring balanced membership in the instances established for content moderation and reconsideration processes at the OCSP level and promoting a role for Canada and other members of the Freedom Online Coalition in ensuring the representativeness, expertise and commitment to online freedom of OCSP online dispute settlement bodies as a whole[153];
3) Ensure coordination and consultation between the new Council and "appellate" online dispute settlement bodies such as the Oversight Board; and

---

152 Facebook Oversight Board, Case decision 2021-002-FB-UA, https://oversightboard.com/decision/FB-S6NRTDAJ/.

153 See Freedom Online Coalition, https://freedomonlinecoalition.com/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

4) Promulgate a standard promoting whistleblower access to "appellate" online dispute settlement bodies such as the Oversight Board. Facebook moderators have often been highly dissatisfied with the internal implementation of Facebook community standards and have encountered heavy-handed control from senior management.[154] One thinks as well of the controversy around the Google AI Ethics Unit.[155] Internal dissent at the platforms on these issues should be harnessed and subject to independent oversight.

The largest challenge to producing an online speech environment conducive to democratic expression is surely the need to adjust the underlying business models of the platforms. How does regulation get to the algorithmic ghost inside the machine? Here I have only more speculative ideas to offer, but ones that show some signs of promise. First, a duty to "take all reasonable measures to identify harmful content," conjoined with the fiduciary duty that companies owe to shareholders and stakeholders, could give rise to pressure on business models, notably as advertisers also attract scrutiny for having their messages accompany hateful content. Second, as discussed earlier,[156] the adoption of a more robust privacy regime with powers given to the Privacy Commissioner to constrain the collection of personal information, if acted upon together with other like-minded jurisdictions, could have significant impact on the business case of some platforms. Indeed, one is already seeing the impact of Apple's new privacy drive on Facebook's advertising relationships.[157] Third, and somewhat more ambitiously, perhaps a group of like-minded states could adopt legislation akin to the proposed *Protecting Americans from Dangerous Algorithms Act* designed to hold platforms liable for algorithms designed to amplify extreme content.[158] Finally, and most ambitiously of all, momentum could

154 Andrew Marantz, "Why Facebook Can't Fix Itself," *New Yorker*, October 12, 2020, https://www.newyorker.com/magazine/2020/10/19/why-facebook-cant-fix-itself.

155 Shirin Ghaffary, "Google says it's committed to ethical AI research. Its ethical AI team isn't so sure," *Vox*, June 2, 2021, https://www.vox.com/recode/22465301/google-ethical-ai-timnit-gebru-research-alex-hanna-jeff-dean-marian-croak.

156 See *supra* note 32 to 36 and accompanying discussion.

157 Laura Forman, "Facebook and Its Advertisers Feel Pinch of Apple's Privacy Drive," *Wall Street Journal*, June 12, 2021, https://www.wsj.com/articles/facebook-and-its-advertisers-feel-pinch-of-apples-privacy-drive-11623502980.

158 U.S. Congress, House, *Protecting Americans from Dangerous Algorithms Act*, H.R. 8636 116th Congress 2d Session, introduced in House October 20, 2020, https://www.congress.gov/bill/116th-congress/house-bill/8636/text. See also Tom Malinowski, "Reps. Malinowski and Eshoo Reintroduce Bill to Hold Tech Platforms

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

gather among leading jurisdictions to overhaul competition law so as to make it easier to break up platforms and reshape the way they operate.[159]

Bibliography

Anti-Terrorism Act of 2015. *Statutes of Canada*. 2015, c. 20. https://laws-lois.justice.gc.ca/eng/annualstatutes/2015_20/page-3.html#h-20.

Arnold, Brent. "Online Harms: Federal Government Announces New Rules and Regulator." *Gowlings WLG*, March 31, 2021. https://gowlingwlg.com/en/insights-resources/articles/2021/federal-government-announces-new-rules/.

Article 19. *The Social Media Councils: Consultation Paper*. June, 2019. https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

An Act respecting the mandatory reporting of Internet child pornography by persons who provide an Internet service. 2011, S.C. c. 4.

Bill C-10. *An Act to amend the Broadcasting Act and to make related and consequential amendments to other Acts*, 2d sess., 43d Parliament, November 3, 2020. https://parl.ca/DocumentViewer/en/43-2/bill/C-10/first-reading#ID0E02B0AA.

Bill C-11. *Digital Charter Implementation Act, 2020*, 2d sess., 43d Parliament, November 17, 2020. https://parl.ca/DocumentViewer/en/43-2/bill/C-11/first-reading.

Bill C-36. *An Act to amend the Criminal Code and the Canadian Human Rights Act and to make related amendments to another Act (hate propaganda, hate crimes and hate speech)*, 2d sess., 43d Parliament, June 23, 2021. https://www.parl.ca/LegisInfo/BillDetails.aspx?Bill=C36&Language=E&Mode=1&Parl=43&Ses=2.

Broadcasting and Telecommunications Legislative Review Panel. Final Report. *Canada's communications future: Time to act*. January, 2020. https://www.ic.gc.ca/eic/site/110.nsf/vwapj/BTLR_Eng-V3.pdf/$file/BTLR_Eng-V3.pdf.

Broadcasting Distribution Regulations. SOR/97-555. https://laws.justice.gc.ca/eng/regulations/SOR-97-555/page-3.html#h-1010707.

Burt, Andrew. "Can Facebook Ever be Fixed?" *Harvard Business Review*. April 8, 2019, https://hbr.org/2019/04/can-facebook-ever-be-fixed.

California Code. Title 1.81.5. *California Consumer Privacy Act of 2018*. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

Accountable for Algorithmic Promotion of Extremism" March 24, 2021, https://malinowski.house.gov/media/press-releases/reps-malinowski-and-eshoo-reintroduce-bill-hold-tech-platforms-accountable.

159 Cecilia Kang, "Lawmakers, Taking Aim at Big Tech, Push Sweeping Overhaul of Antitrust," *New York Times*, June 11, 2020, https://www.nytimes.com/2021/06/11/technology/big-tech-antitrust-bills.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Canada-U.S.-Mexico Trade Agreement. Article 19.17. https://www.international.gc.ca/trade-commerce/trade-agreements-accords-commerciaux/agr-acc/cusma-aceum/text-texte/19.aspx?lang=eng.

Canadian Citizens' Assembly on Democratic Expression. *Recommendations to Strengthen Canada's Response to New Digital Technologies and Reduce the Harm Caused by their Misuse*. Public Policy Forum, January, 2021. https://ppforum.ca/wp-content/uploads/2021/01/CanadianCitizens%E2%80%99AssemblyOnDemocraticExpression-PPF-JAN2021-EN.pdf.

Canadian Commission of Democratic Expression. *Harms Reduction: A Six-Step Program to Protect Democratic Expression Online*. Public Policy Forum. January, 2021. https://ppforum.ca/wp-content/uploads/2021/01/CanadianCommissionOnDemocraticExpression-PPF-JAN2021-EN.pdf.

Canadian Human Rights Act. *Revised Statutes of Canada.* 1985, c. H-6, archived version. https://laws-lois.justice.gc.ca/eng/acts/h-6/section-13-20021231.html#:~:text=13%20(1)%20It%20is%20a,Parliament%2C%20any%20matter%20that%20is.

Canadian Human Rights Commission. "Statement – We must do more to curb online hate." January 21, 2021. https://www.chrc-ccdp.gc.ca/en/resources/statement-we-must-do-more-curb-online-hate.

Canadian Race Relations Foundation. "Poll demonstrates support for strong social media regulations to prevent online hate and racism." January 25, 2021. https://www.crrf-fcrr.ca/en/news-a-events/media-releases/item/27349-poll-demonstrates-support-for-strong-social-media-regulations-to-prevent-online-hate-and-racism.

Canadian Radio-television and Telecommunications Commission (CRTC). "Frequently asked questions." April 1, 2015. https://crtc.gc.ca/eng/faqs.htm.

Christchurch Call to Eliminate Terrorist and other Extremist Content Online. May 15, 2019. https://www.christchurchcall.com/call.html.

Commission des droits de la personne et des droits de la jeunesse. *Mémoire à la Commission des institutions de l'Assemblée nationale*. August 2015. https://www.cdpdj.qc.ca/storage/app/media/publications/memoire_PL59_discours-haineux.pdf.

Les actes haineux à caractère xénophobe, notamment islamophobe: résultats d'une recherche menée à travers le Québec, August, 2019. https://www.cdpdj.qc.ca/storage/app/media/publications/etude_actes_haineux.pdf.

Commission des droits de la personne et des droits de la jeunesse (Coffy et une autre) c. Brisson. 2009 QCTDP. https://canlii.ca/t/22qhm.

Copyright Act. *Revised Statutes of Canada*. 1985, c. C-42. https://laws-lois.justice.gc.ca/eng/acts/c-42/.

Criminal Code. Revised Statutes of Canada. 1985, c. C-46. https://laws-lois.justice.gc.ca/eng/acts/c-46/.

Curry, Bill and Raman-Wilms, Menaka. "New internet bill on hate crime and revenge porn coming in 'very near future,' Guilbeault says." *Globe & Mail*, June 7, 2021. https://www.theglobeandmail.com/politics/article-new-internet-bill-on-hate-crime-and-revenge-porn-coming-in-very-near/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Davey, Jacob, Mackenzie Hart, Cécile Guerin, ed. Jonathan Birdwell. *Interim Report: An Online Environmental Scan of Right-wing Extremism in Canada*, *The Institute for Strategic Dialogue*. June 19, 2020. https://www.isdglobal.org/wp-content/uploads/2020/06/An-Online-Environmental-Scan-of-Right-wing-Extremism-in-Canada-ISD.pdf.

Department of Justice. "Combatting hate speech and hate crimes: Proposed legislative changes to the *Canadian Human Rights Act* and the *Criminal Code*." June 23, 2021. https://www.justice.gc.ca/eng/csj-sjc/pl/chshc-lcdch/index.html.

Digital Citizen Initiative, Department of Canadian Heritage. "Discussion Guide." July 26, 2021. https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/discussion-guide.html.

"Technical Paper." July 26, 2021. https://www.canada.ca/en/canadian-heritage/campaigns/harmful-online-content/technical-paper.html.

Edelman, Gilad. "Social Media CEOs Can't Defend Their Business Model." *Wired*, March 25, 2021. https://www.wired.com/story/social-media-ceo-hearing-cant-defend-business-model/.

Elghawaby, Amira. "Canada is Bringing in New Legislation to Stop the Spread of Online Hate. Here's How It Can Work." *Press Progress*, April 7, 2021. https://presprogress.ca/canada-is-bringing-in-new-legislation-to-stop-the-spread-of-online-hate-heres-how-it-can-work/.

Eliadis, Pearl. "The Controversy Entrepreneurs." *Maisonneuve*, August 20, 2009. https://maisonneuve.org/article/2009/08/20/controversy-entrepreneurs/.

European Union. Article 29 Data Protection Working Party. "Guidelines on the Implementation of the Court of Justice of the European Union Judgment on '*Google Spain and Inc v Agencia Española de Protección de Datos* (*AEPD) and Mario Costeja González*'." November 26, 2014, C-131/12. https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=64437.

Directive on Copyright in the Digital Single Market. April 17, 2019), 2019/790. https://eur-lex.europa.eu/eli/dir/2019/790/oj.

Facebook Oversight Board. https://oversightboard.com/.

Case decision 2021-001-FB-FBR. https://oversightboard.com/decision/FB-691QAMHJ/.

Case decision 2021-002-FB-UA. https://oversightboard.com/decision/FB-S6NRTDAJ/.

Forman, Laura. "Facebook and Its Advertisers Feel Pinch of Apple's Privacy Drive." *Wall Street Journal*, June 12, 2021. https://www.wsj.com/articles/facebook-and-its-advertisers-feel-pinch-of-apples-privacy-drive-11623502980.

Freedom Online Coalition. https://freedomonlinecoalition.com/

Ghaffary, Shirin. "Google says it's committed to ethical AI research. Its ethical AI team isn't so sure." *Vox*, June 2, 2021. https://www.vox.com/recode/22465301/google-ethical-ai-timnit-gebru-research-alex-hanna-jeff-dean-marian-croak.

Geist, Michael. "The real consequences of Steven Guilbeault's battle with the web giants." *Maclean's*, May 3, 2021. https://www.macleans.ca/opinion/the-real-consequences-of-steven-guilbeaults-battle-with-the-web-giants/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Germany. *Network Enforcement Act*, English translation. https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2.

Gill, Lex. "Legal Aspects of Hate Speech." *Canadian Commission on Democratic Expression*, June, 2020. https://ppforum.ca/wp-content/uploads/2020/07/1.DemX_LegalAspects-EN.pdf.

Global Internet Forum to Counteract Terrorism. https://gifct.org/.

Hill, Kashmir and Daisuke Wakabayashi. "Google Seeks to Break Vicious Cycle of Online Slander." *New York Times*, June 10, 2021. https://www.nytimes.com/2021/06/10/technology/google-algorithm-known-victims.html.

House of Commons. Standing Committee on Justice and Human Rights. *Taking Action to End Online Hate*. June, 2019. https://www.ourcommons.ca/Content/Committee/421/JUST/Reports/RP10581008/justrp29/justrp29-e.pdf.

Standing Committee on Access to Information, Privacy and Ethics. *Democracy under Threat: Risks and Solutions in the Era of Disinformation and Data Monopoly*. December, 2018. https://www.ourcommons.ca/Content/Committee/421/ETHI/Reports/RP10242267/ethirp17/ethirp17-e.pdf

Jaworski, Michal and Athar Malik. "Did You Notice? When A Notice Is Not A Notice Under The Notice And Notice Regime." March 27, 2019. https://www.mondaq.com/canada/copyright/792094/did-you-notice-when-a-notice-is-not-a-notice-under-the-notice-and-notice-regime.

Kang, Cecilia. "Lawmakers, Taking Aim at Big Tech, Push Sweeping Overhaul of Antitrust." *New York Times*, June 11, 2020. https://www.nytimes.com/2021/06/11/technology/big-tech-antitrust-bills.html.

Karadeglija, Anja. "New definition of hate to be included in Liberal bill that might also revive contentious hate speech law." *National Post*, March 3, 2021. https://nationalpost.com/news/politics/new-definition-of-hate-to-be-included-in-liberal-bill-that-might-also-revive-contentious-hate-speech-law.

Kilvington, Daniel. "The virtual stages of hate: Using Goffman's work to conceptualise the motivations for online hate." *Media, Culture & Society* 43, no. 2 (2020): 256-272. https://journals.sagepub.com/doi/10.1177/0163443720972318.

Krishnamurthy, Vivek and Jessica Fjeld. "CDA 230 Goes North American? Examining the Impacts of the USMCA's Intermediary Liability Provisions in Canada and the United States." *SSRN*, July 7, 2020. https://ssrn.com/abstract=3645462.

Krolik Adam, and Kashmir Gill. "The Slander Industry." *New York Times*, April 24, 2021. https://www.nytimes.com/interactive/2021/04/24/technology/online-slander-websites.html.

Law Commission of Ontario. *Defamation Law in the Internet Age*, March 2020 at 1. https://www.lco-cdo.org/wp-content/uploads/2020/03/Defamation-Final-Report-Eng-FINAL-1.pdf.

"Learn about us." https://www.lco-cdo.org/en/learn-about-us/.

Lemire v. Canada (Human Rights Commission). 2014 FCA 18. https://canlii.ca/t/g2x2d.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Madison, James. *Federalist* No.10, in *The Federalist Papers*. ed. Clinton Rossiter (New York: New American Library, 1961). https://avalon.law.yale.edu/18th_century/fed10.asp.

"Notes for the National Gazette Essays." (ca. December 19, 1791–March 3, 1792). https://founders.archives.gov/?q=literati%20%22useful%20knowledge%22&s=1111311111&r=1.

"Report on the Virginia Resolutions." January, 1800. https://press-pubs.uchicago.edu/founders/documents/amendI_speechs24.html.

Marantz, Andrew. "Why Facebook Can't Fix Itself." *New Yorker*. October 12, 2020. https://www.newyorker.com/magazine/2020/10/19/why-facebook-cant-fix-itself.

Malinowski, Tom. "Reps. Malinowski and Eshoo Reintroduce Bill to Hold Tech Platforms Accountable for Algorithmic Promotion of Extremism." March 24, 2021. https://malinowski.house.gov/media/press-releases/reps-malinowski-and-eshoo-reintroduce-bill-hold-tech-platforms-accountable.

Minister of Justice and Attorney General of Canada. *Consultation Paper: Online Hate*. July 14, 2020. https://ocla.ca/wp-content/uploads/2020/07/2020-07-14-Consultation-paper-Online-hate.pdf.

Office of the Prime Minister of Canada. "Canada joins Christchurch Call to Action to eliminate terrorist and violent extremist content online." May 15, 2019. https://pm.gc.ca/en/news/news-releases/2019/05/15/canada-joins-christchurch-call-action-eliminate-terrorist-and-violent.

"Minister of Canadian Heritage Mandate Letter." December 13, 2019. https://pm.gc.ca/en/mandate-letters/2019/12/13/minister-canadian-heritage-mandate-letter.

"Minister of Justice and Attorney General of Canada Mandate Letter." December 13, 2019. https://pm.gc.ca/en/mandate-letters/2019/12/13/minister-justice-and-attorney-general-canada-mandate-letter.

"Minister of Canadian Heritage Supplementary Mandate Letter." January 15, 2021. https://pm.gc.ca/en/mandate-letters/2021/01/15/minister-canadian-heritage-supplementary-mandate-letter.

"Minister of Justice and Attorney General of Canada Supplementary Mandate Letter." January 15, 2021. https://pm.gc.ca/en/mandate-letters/2021/01/15/minister-justice-and-attorney-general-canada-supplementary-mandate.

Office of the Privacy Commissioner of Canada. "A Data Privacy Day Conversation with Canada's Privacy Commissioner." January 28, 2020. https://www.priv.gc.ca/en/opc-news/speeches/2020/sp-d_20200128/.

Paliser, Eli. *The filter bubble: what the Internet is hiding from you*, (New York: Penguin, 2011).

Projet de loi 59. *Loi édictant la Loi concernant la prévention et la lutte contre les discours haineux et les discours incitant à la violence et apportant diverses modifications législatives pour renforcer la protection des personnes*, 1er sess., 41e législature, June 10, 2015. http://m.assnat.qc.ca/fr/travaux-parlementaires/projets-loi/projet-loi-59-41-1.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Projet de loi 64. *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels* 1er sess., 42e legislature, June 12, 2020. http://m.assnat.qc.ca/fr/travaux-parlementaires/projets-loi/projet-loi-64-42-1.html.

Protecting Canadians from Online Crime Act. *Statutes of Canada*, 2014, c. 31. https://laws-lois.justice.gc.ca/eng/annualstatutes/2014_31/FullText.html.

R v Keegstra. [1990] 3 SCR 697. https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/695/index.do.

Rosen, Jeffrey. "America is Living James Madison's Nightmare." *The Atlantic*, October, 2018. https://www.theatlantic.com/magazine/archive/2018/10/james-madison-mob-rule/568351/.

Saskatchewan (Human Rights Commission) v. Whatcott. [2013] 1 SCR 467. https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/12876/index.do?q=whatcott.

Sentelle David. "Freedom of the Press: A Liberty for All or a Privilege for a Few?" *Cato Supreme Court Review* (2014): 15-34.

Sheehan, Colleen. "The Politics of Public Opinion: James Madison's 'Notes on Government'." *William and Mary Quarterly* 49, no. 4 (1992) 609-27.

Solomun, Sonja Maryna Polataiko, and Helen A. Hayes. "Platform Responsibility and Regulation in Canada: Considerations on Transparency, Legislative Clarity, and Design." *Harvard Journal of Law and Technology (Digest)*) 34 (2021): 1-18. https://jolt.law.harvard.edu/digest/platform-responsibility-and-regulation-in-canada-considerations-on-transparency-legislative-clarity-and-design.

Swartz, Jon. "California's landmark privacy law is Facebook's next 'nightmare'." *Market Watch*, August 22, 2020. https://www.marketwatch.com/story/californias-landmark-privacy-law-is-facebooks-next-nightmare-2020-08-18.

Thompson, Elizabeth. "Canada not exempt from social media forces that created U.S. Capitol riot, heritage minister says." *CBC News*, January 29, 2021. https://www.cbc.ca/news/politics/facebook-twitter-canada-regulation-1.5894301.

Tusikov, Natasha. "U.K. and Australia move to regulate online hate speech, but Canada lags behind." *National Post*, April 11, 2019. https://nationalpost.com/pmn/news-pmn/u-k-and-australia-move-to-regulate-online-hate-speech-but-canada-lags-behind.

Chokepoints: Global Private Regulation on the Internet, (Oakland: University of California Press, 2017).

United Kingdom. White Paper on Online Harms. April 6, 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.

Urban, Jennifer M., Joe Karaganis and Brianna L. Schofield. "Notice and Takedown in Everyday Practice" *UC Berkeley Public Law Research Paper*, No. 2755628, March 24, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2755628.

U.S. Congress. House. *Protecting Americans from Dangerous Algorithms Act*, H.R. 8636 116th Cong., 2d sess. Introduced in House October 20, 2020. https://www.congress.gov/bill/116th-congress/house-bill/8636/text.

U.S. Communications Decency Act. *U.S. Code* 47 (2018).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Webe, Joel. "Hate speech no longer part of Canada's Human Rights Act." *National Post*, June 27, 2013. https://nationalpost.com/news/politics/hate-speech-no-longer-part-of-canadas-human-rights-act.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Lessons learned from the first years with the NetzDG

Maximilian Hemmert-Halswick

Abstract: German lawmakers have done pioneering work with the Network Enforcement Act. This law makes specific compliance requirements for social networks in order for them to remove illegal content more quickly and reliably. Criticism was particularly strong at the beginning, but the law now seems to have taken its place in the field of platform regulation. Three years after its enactment, the law underwent an amendment process. The amendment is based primarily on the experience gained up to that point. In essence, the aim is to eliminate identified deficiencies, which is mainly achieved by extending user rights and tighter regulatory control.

Keywords: Content Regulation; Hate Speech; Social Networks; Overblocking; Compliance approach; Censorship

I. Introduction - Balance between State Sovereignty and Economic Freedom

"[A]s Nostradamus said: across the sea they will come like locusts, but they will not be animals... how right the man was...". This comment refers to the refugee crisis in Germany and reveals the author's displeasure or contempt for the policy and the people who have fled. Are the confines of civilized, permissible discourse being left behind here? This is certainly not a simple question. The court that had to rule on this also had a hard time with the decision, but ruled that the deletion of the content was lawful.[1]

When it comes to the legally mandated deletion of content on social networks, such cases inevitably take center stage. There will hardly be any discussion about extreme, blatant cases, i.e. where there is broad consensus that such social interaction is unacceptable. Think, for example, of a call to commit murder; or even dishonorable and false allegations about other people that can seriously damage the person. In contrast, borderline cases show the difficulty of the matter. One must be aware of this, and it is

1 OLG Stuttgart, 6.9.2018 – 4 W 63/18.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

against this background that the German "Law to Improve Law Enforcement in Social Networks" – in short: NetzDG[2] – must be seen. The law was passed just under 4 years ago. The aim is to protect public safety, in particular the general right to protection of one' s personality, and to ensure rational discourse. To this end, it requires providers of social networks to delete unlawful comments more quickly and reliably.[3] In 2020, the legislator began the process for amending the NetzDG. With three years of experience applying the law, it is hoped to address some shortcomings and provide improved enforcement. A year later, the amendments were approved by all legislative bodies.

This chapter will present the beginnings and the experiences that have led to the amendments of the NetzDG. The lines of development clearly speak for a learning effect on the part of the German legislator, which is also likely to be of international interest, since parallels can also be found in other countries, even if they have not rushed ahead with a law like the NetzDG. However, a great deal has happened in the industry during this time: content regulation has often been in the media spotlight; Twitter has blocked accounts of highly public figures[4], and Facebook has set up an "Oversight Board" which serves as a quasi-court. Against the backdrop of these events, the tendency can probably be discerned that it has now become clear that states must become involved in content regulation in one way or another.

II. The Approach of the NetzDG

The law was the first of its kind worldwide. It was subject to harsh criticism, especially during the legislative process.[5] Yet, a modus vivendi seems

2 Netzwerkdurchsetzungsgesetz, September 1, 2017 (BGBl. I S. 3352).

3 Cf. BT-Drs. 18/12356, 11 (German Parliament Document).

4 The most noted case was certainly the blocking of the account of then-President Donald Trump in response to the storming of the Capitol, and later, for example, of Mike Lindell ("MyPillow Guy") for spreading disinformation concerning the results of the 2020 presidential election.

5 Spindler, Gerald, "Der Regierungsentwurf zum Netzwerkdurchsetzungsgesetz – europarechtswidrig?", *Zeitschrift für Urheber- und Medienrecht ZUM* (2017): 473; Gersdorf, Hubertus, "Hate Speech in sozialen Netzwerken," *MMR Zeitschrift für IT-Recht und Recht der Digitalisierung* (2017): 439. Guggenberger, Nikolas, "Das Netzwerkdurchsetzungsgesetz – schön gedacht, schlecht gemacht," *Zeitschrift für Rechtspolitik ZRP* (2017): 98.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to have been found now. While the law was described as "innovative"[6] on the one hand, criticism was also voiced that the law constituted a violation of freedom of expression. In the following, the approach of the law will be described, and the points of criticism will be explained and put into context.

1. Compliance approach

The law has a compliance approach, which means that it just wants to make sure that the social networks act according to the law, and it does not – at least directly – address the users in any way.

The regulatory approach of the NetzDG is to make social network providers comply to a greater extent with their deletion obligations, which existed already without the NetzDG. In this respect, the NetzDG sets compliance requirements for how social network providers are to set up their complaint management, i.e. how they are supposed to deal with user complaints. Among other things, it is supposed to guarantee the deletion of illegal content within seven days or – if the illegality is "obvious" – within 24 hours. This focus on time constraints resulted from previous experience, as high-ranking politicians, among others, had been victims of online aggression and the content was not deleted for several days despite complaints. Subsequently, the German government determined in surveys that the social networks in fact only very rarely comply with their deletion obligations.[7] This deficiency is therefore to be remedied by setting specific time limits.

In this approach, it is of course noticeable that the network providers take on the active role; ultimately, they are the ones who decide which content will be deleted. The state retreats to a position of an observer. Criticism was voiced during the legislative process that law enforcement was being "privatized".[8] However, this criticism does not really hold up, since the NetzDG also emphasizes that network operators do not have any active monitoring obligations. One alternative, of course, is for the state

---

6 Holznagel, Bernd, "Das Compliance-System des Entwurfs des Netzwerkdurchsetzungsgesetzes," *Zeitschrift für Urheber- und Medienrecht ZUM* (2017): 615; Bundestag Protokoll-Nr. 18/153, 21.

7 BT-Drs. 18/12356, 11.

8 Wimmers, Jörg and Heymann, Britta, "Zum Referentenentwurf eines Netzwerkdurchsetzungsgesetzes (NetzDG) – eine kritische Stellungnahme," *AfP Zeitschrift für das gesamte Medienrecht* (2017): 98.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

to take over monitoring. Conversely, there is the option of not imposing strict requirements on network operators. With the compliance approach, the German legislator wants to strike a balance between a too far-reaching encroachment on corporate freedom and an ineffective regulatory regime.

2. *Terminology*

Just as in any law, it is of particular importance to define the essential terms in order to ensure its applicability in the first place. In many areas of law, reference can be made to familiar terms (that are already in use (in other laws). In the case of the NetzDG, this applies only to a limited extent, since content regulation on social networks is practically a new phenomenon. First and foremost, the definitions must clarify what social networks are and what content is to be covered by the NetzDG.

a) Addressees: social networks

Regarding the term “social network”, perhaps the "you know it when you see it"-approach would be a viable option, especially for younger Internet users who have a more or less clear idea of what is meant by a social network without ever having taken a closer look at the individual criteria. Ultimately, the question of definition is related to what the NetzDG is intended to achieve. The legislator was primarily concerned with the perpetuation effect associated with publication on a network used by many people. The fact that only the big ones are to be covered means that various parameters have to be set in a way to exclude the smaller ones. The number of members plays a role – but does it only depend on registered users? What about messenger services, where the groups can sometimes be so large that they resemble “classic” social networks? What about comment functions on ,for example, newspaper sites? In order not to cover these aspects, the German legislator has chosen the following definition:

"Telemedia service providers who operate platforms on the Internet with the intention of making a profit, which are intended for users to share any content with other users or make it accessible to the public."[9] Thematically limited social networks such as Linkedin are excluded by the arbitrariness criterion (“to share *any* content”). Furthermore, only

9 § 1 sec. 1 NetzDG.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

networks with a user base of more than 2 million users are included. Journalistic content is excluded. Gated communities such as messenger services are also explicitly excluded, as this is not directly clear from the definition.[10]

b) The most important term: illegal content

One of the most important points to consider when passing a law that targets content on social networks is what content is covered. What content is at issue in the first place? What content is so important or harmful that it should be removed from the platform as quickly as possible. If the guidelines are too far-reaching, the accusation of censorship can legitimately be made. However, if too few specifications are made, the effectiveness can be questioned and the scope of application will be relatively small. In addition, the delimitation of content must be as precise as possible so that it is clear which content is covered - so that, on the one hand, network providers know which content is to be deleted. On the other hand, it is just as important for users to know what they are allowed to post.

According to the NetzDG, illegal content is content that fulfills one of the criminal offenses listed in § 1 of the NetzDG, which are primarily those that roughly cover the phenomenon of "hate speech," i.e., the crimes of defamation and the crimes against public order such as incitement to hatred. The fact that the NetzDG is limited to criminal offenses makes it clear that only relatively vile statements are covered by it, although it goes without saying that difficulties of demarcation cannot be avoided here either. However, by this the legislator avoids defining the term hate speech or disinformation, as this is hardly possible.[11] The criminal offense of "insult" is included in the NetzDG. This is, of course, also an offense that is open to interpretation, and in this respect is similarly vague as the term hate speech. Social networks, on the other hand, define hate speech and disinformation in their guidelines, but can do so comparatively easily because, unlike democratically constituted states, they – generally speaking – do not have to meet balanced legal requirements. Here, however, it also

10 BT-Drs. 18/12356, 12 (German Parliament Document).

11 Hoven, Elisa and Krause, Melena, "Die Strafbarkeit der Verbreitung von ‚Fake News'," *Juristische Schulung (JuS)* (2017): 1167.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

becomes obvious that the question of definitional sovereignty is of utmost importance.[12]

3. Complaints management

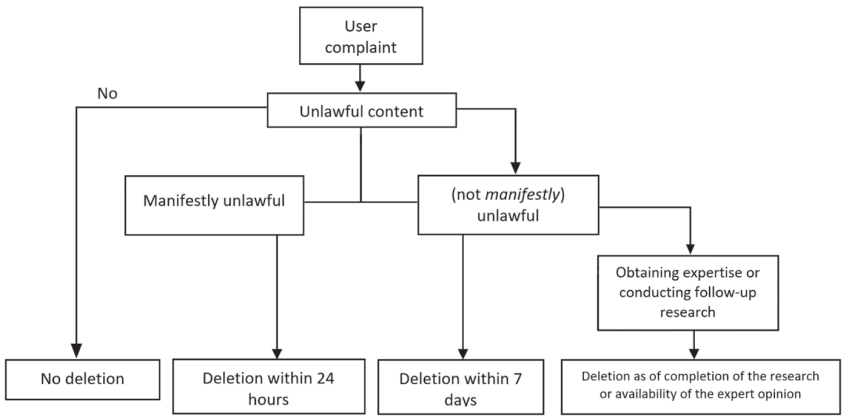
a) Establishment of a complaint management system

The most important element is that the social networks must implement a proper complaints management system as outlined in § 3 NetzDG: a functional system of how they handle the complaints against hate speech content. The submission of complaints must be as simple as possible. Ideally, the complaint option is set up directly next to the post in the news feed. It must then be ensured that the network provider takes note of the complaint quickly. This is then followed by an assessment within the appropriate deletion period. Setting up a complaints management system also involves providing regular training for employees to ensure that they are up to the task; it has been found ot that the task of deleting illegal content is very stressful.

12 The subject of a current debate is whether network providers may also delete content if the content is not constitutionally objectionable. This has to do with the indirect third-party effect of fundamental rights that exists in Germany. This means that under certain circumstances, private parties such as big companies must also comply with fundamental rights requirements in their actions. In relation to the case of content regulation on social networks, this means that Facebook and Co. must respect users' freedom of expression. However, the extent to which they must respect users' fundamental rights has not yet been conclusively decided. The answer is particularly relevant to the question of whether network providers may also delete lawful content. If there is a strong connection to fundamental rights, this is likely to be negated.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Illustrative diagram of the complaints management system, according to § 3 NetzDG:



b) The Danger of Overblocking

The main goal of the NetzDG is that the networks get illegal content quickly off their platforms. However, since a fine can be imposed for violations of the NetzDG, there is a risk that the networks will, in case of doubt, delete content rather than leave it on the platform. The consequence of this situation might be an overblocking. That was the main critique by the experts during the legislative hearing.[13] The hope was that an assessment of whether there really was overblocking could be made on the basis of the transparency reports that providers were required to produce under § 2 of the NetzDG. Unsurprisingly, however, the transparency reports hardly allowed any conclusions to be drawn in this regard. The bare figures on complaints received and deletions made say practically nothing about the

13 Cf. protocol of the parliamentary expert hearing protocol Nr. 18/153, 16, 30, 38; Guggenberger, Nikolas, "Das Netzwerkdurchsetzungsgesetz in der Anwendung, "TRENNUNG *Neue Juristische Wochenschrift (NJW)* (2017): 2577; Schwartmann, Rolf, "Verantwortlichkeit Sozialer Netzwerke nach dem Netzwerkdurchsetzungsgesetz,"TRENNUNG *GRUR-Prax Praxis im Immaterialgüter- und Wettbewerbsrecht* (2017): 317.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

existance of overblocking.[14] However, a recent study has come to the conclusion that overblocking is taking place.[15]

Closely related to this is the accusation of insufficient consideration of the interests of the authors of deleted posts. In principle, the NetzDG does not provide for the author to be heard at any point in the proceedings or to obtain a reversal of the deletion. There may be various reasons for not involving the author, for example because this would prolong the proceedings. After all, expeditiousness in cancellation is the key purpose of the law. A possible starting point is the establishment of an option for the author to complain after the deletion.

c) Establishment of regulated self-regulation

In response to this criticism, which was voiced during the legislative process, another mechanism was introduced to ensure the accuracy of the decisions on the one hand and to take the pressure off the social network providers on the other. To this end, in § 3 sec. 6-9 NetzDG, the option was created for the social networks themselves to establish independent bodies to review complaints on behalf of the social networks: establishment of regulated self-regulation.[16] The first and most prominent of its kind is the *Freiwillige Selbstkontrolle Multimedia-Dienste* (FSM[17]). The social networks can therefore then always decide whether to forward a complaint to this body. They must then accept the decision of this body and make a deletion accordingly. This regulation has existed from the beginning, but the first trials were not carried out until 2020 , as the establishment of the body had taken some time.[18]

---

14 Löber, Lena Isabell and Roßnagel, Alexander, "Das Netzwerkdurchsetzungsgesetz in der Umsetzung," *MMR* (2019): 73; Ladeur, Karl-Heinz, "Ist der Regierungsentwurf eines NetzDG 2.0 vom 19.2.2020 netzgerecht?", *Kommunikation und Recht K&R* (2020): 250; Heindorf, Manon, „Das Netzwerkdurchsetzungsgesetz in der Umsetzung: Zwei Jahre NetzDG – eine Bilanz," *Verwaltungsrundschau VR* (2020): 117.

15 Liesching, Marc, et al., *Das NetzDG in der praktischen Anwendung* (Berlin: Carl Grossmann Verlag, 2021), 143, doi:10.24921/2021.94115953.

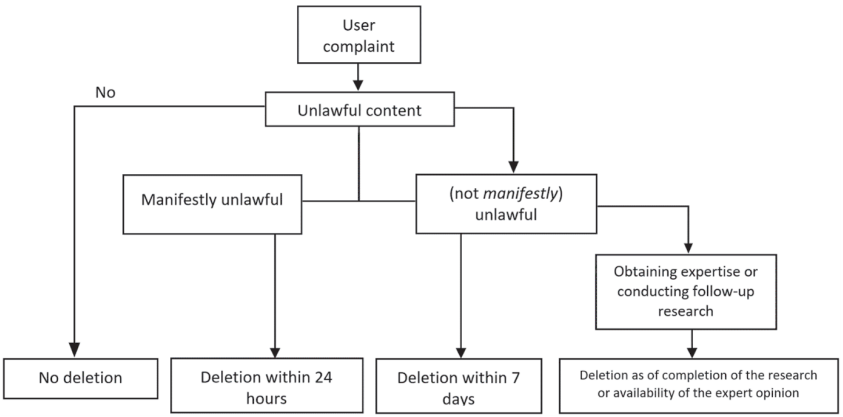
16 This is modeled after voluntary monitoring in the film industry (*Freiwillige Selbstkontrolle der Filmwirtschaft*).

17 In English: Association for Voluntary Self-Regulation of Digital Media Service Providers.

18 See chapter in this book *Holznagel/Kalbhenn*.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Illustrative diagram of the process with the FSM[19]:



4. *Transparency obligations: Conflict between NetzDG and community standards – Facebook case study*

As already mentioned, the obligations of the NetzDG apply primarily to the network providers. They have to set up complaint management, they have to check the content, and they have to carry out deletions if necessary. State authorities stay out of this practice. However, since the state does not want to leave the network providers to act idly and unsuspectingly, the NetzDG provides for comprehensive transparency obligations. Network providers must report on how they implement the requirements of the NetzDG. In terms of the role of the state in the structure of the NetzDG, this is the most important regulation, so that the state does not completely relinquish responsibility. The transparency obligations provide the state and the public with information about how many deletions are made, how many complaints are received by the networks, and what type of content the complaints focus on. It should be mentioned that the transparency reports essentially only provide quantitative information on how complaints are handled. In qualitative terms, the network operators only have to describe how they have structured their complaints management and what they base their decisions on. Examples of deletions are – unfortunately – not to be included in the transparency reports.

19 On the basis of the diagram at "NetzDG," FSM, accessed July 6, 2021, https://www.fsm.de/de/netzdg.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

It is particularly noteworthy that Facebook – i.e. the largest social network – is attempting to circumvent the transparency obligation by setting up two complaints procedures. One is the NetzDG complaint procedure, which is very hidden and can only be reached via several clicks. The other is Facebook's own complaint procedure, which is based on Facebook's community standards and is located directly next to a post. Facebook does not list any information about the latter complaint procedure in the transparency report, although the vast majority of complaints are made via this procedure: approximately 2,000 complaints are listed in the NetzDG Transparency Report[20], whereas several million complaints are counted with regard to the Community Standards, albeit globally.[21] The Federal Office of Justice has imposed a fine of two million euros on Facebook for this. The courts are currently reviewing whether this was lawful. It is not clear what the answer will be, because the NetzDG only stipulates transparency obligations in relation to the NetzDG complaints procedure. Rather, the legislator had probably not even considered this as an option, and had simply assumed that the network providers would report on *any* deletions. However, it is also clear that Facebook's approach is a deliberate circumvention of the law. On the other hand, it may not always be so easy for the social networks to determine exactly whether a complaint is applicable to German law.

5. Conclusion

So far, the basic requirements of the NetzDG have been presented. It became clear that there was potential for improvement. For example, the danger of overblocking could be further mitigated with a counter-appeal procedure, in which the person affected by the deletion also presents his or her point of view. Also, the users had little say in the matter. However, the lack of inclusion of network-internal complaints was an obvious shortcoming. It was also stated that the legislator had simply not thought of this and subsequently realized that the NetzDG should also include these com-

20 "NetzDG Transparenzbericht Januar 2021," Facebook, accessed July 6, 2021, https://about.fb.com/de/wp-content/uploads/sites/10/2021/01/Facebook-NetzDG-Transparenzbericht-Januar-2021.pdf (July-December 2020: 4.211 complaints, 29% deleted).

21 "Community Standards Enforcement Report, Third Quarter 2020," Facebook, accessed July 6, 2021, https://about.fb.com/news/2020/11/community-standards-enforcement-report-nov-2020/ (22 Mio. regarding hate speech, 95% deleted).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

plaints. The fact that there is a fundamental misunderstanding is also clear from the fact that in the Summer of 2019, the Federal Minister of Justice, *Christine Lambrecht*, had to make clear that the community standards of the networks are "not above the law".[22]

III. Amending the NetzDG

Three things can be identified that fueled the amendment process: First, the lawmaker still want to make sure that there is no overblocking ocurring. Secondly, they want to know about the complaints handling in more detail, as in enhance the transparency of deletion decisions. And thirdly, the power struggle in which the state wants to be in control of the rules that govern online speech. The new provisions that address these findings are presented below.

1. Countercomplaints procedure

The countercomplaints procedure, which is introduced in the new § 3b of the NetzDG, obliges the providers of social networks to have a procedure in place that allows users to state their opinion on a decision made by the provider and to seek a re-examination. Both the author of a content and the complaining user can take the initiative. According to the new provision, the provider must "promptly subject its initial decision to reconsideration by a person not involved with the initial decision." Only if the provider wishes to remedy the objection the other party must be given the opportunity to state its position; in this way the new law takes into account that objections can also be raised abusively. The procedure ends with the review decision, which must be forwarded to the two users concerned.

With this, the legislator wants to address the criticism regarding overblocking and the lack of procedural participation of the author. Too much deletion is to be prevented by allowing the authors of content to take action against deletions in an internal network procedure. Conversely,

22 "Bundesamt für Justiz (BfJ) erlässt Bußgeldbescheid gegen Facebook," Federal Ministry of Justice and Consumer Protection, Press release from July 3, 2019, accessed July 6, 2021, https://www.bmjv.de/SharedDocs/Artikel/DE/2019/070319_Facebook.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

it should be easier to have a review of content that has been objected to by users but not removed.

The countercomplaints procedure is in particular a response to the widespread called for a "put-back procedure".[23] However, the law precisely does not stipulate the right to reinstatement of the unlawfully deleted content. Indeed, it does not say anything about what is to be done after the provider's repeated decision. Presumably, the lawmaker implicitly assumes that this will lead to a put-back, in case of the lawfulness of the previously deleted content. For example, at one point the explanatory memorandum talks about the provider reporting "in how many cases the counterproposal was remedied."[24] It is difficult to imagine a remedy that does not involve a put-back.

It has been argued in the literature that the legislator is reluctant to establish an explicit put-back procedure, as this would entail a right to publication. Whether such a right could be upheld constitutionally is still being debated. In this respect, the legislative reluctance can be explained by the legal situation, which is still developing.

Another criticism goes against the timing of the countercomplaints procedure, or rather that the author of a deleted content must first accept the deletion; he or she is not heard before the deletion. It is argued here that the protection of freedom of expression requires that the author be heard beforehand and given the opportunity to respond to the complaint. The "Stadium Ban Decision" of the Federal Constitutional Court is being referenced for this argument.[25] In this case, the court suggests that a hearing be held prior to the exclusion of a socially significant event (attending a football match).[26] Facebook and other social networks are – in a way – comparable to visiting a football stadium in terms of social relevance... yet it does not seem necessary to require a pre-removal cross-appeal on constitutional grounds: Deleting a piece of content is not comparable to an exclusion - blocking the account would be – furthermore, not removing an unlawful piece of content weighs heavier than removing a lawful piece of content (that is, especially, if the countercomplaints procedure can ensure a speedy restoration).

---

23 Löber, and Roßnagel, "Das Netzwerkdurchsetzungsgesetz," 75; Peukert, Alexander, "Gewährleistung der Meinungs- und Informationsfreiheit in sozialen Netzwerken," *MMR* (2018): 572; Schwartmann,"Verantwortlichkeit sozialer Netzwerke," 318.

24 BT-Drs. 19/18792, 8 (§ 2 (2) No. 11).

25 Niggemann, Sandra, "Die NetzDG-Novelle," *Computer und Recht CR* (2020): 329.

26 BVerfGE 148, 267.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2. *Transparency rules*

Since shortcomings in the transparency regulations also became apparent relatively quickly, the legislator had to take action here as well. In particular, the circumvention by Facebook was a deficiency with an easy remedy. Here, the bill on combating right-wing extremism and hate crime brings about the most significant innovation. Therein it is regulated that, for example, complaints and deletions in connection with the Community Directives are also listed in the transparency report. This is achieved by broadening the definition of "complaints" to now include complaints under the Community Standards. This certainly makes sense, as do the stricter requirements for user-friendliness of reporting channels included in the amendment. A dichotomy of reporting channels, as explained at the beginning with regard to Facebook, should then no longer exist.

The amendment expands the reporting obligations, requiring reporting on the use of procedures for automated detection of illegal content. And also on whether and to what extent academia has been given access to information from the network provider to enable evaluation.

These additions are all reasonable, but the decision-making practice of the network providers should be even more transparent. *All* decisions about content should be published, as some have already pointed out.[27] As a response to this criticism, the legislature added another amendment "at the last minute".[28] § 5a now gives "researchers" a right to access information from social network providers about "the use and specific mode of operation of procedures for the automated recognition of content" as well as "the circulation of content which has been the subject of complaints about illegal content or which has been removed or blocked by the provider". The scope of those entitled to make a claim extends to "any ... person conducting scientific research" (paragraph 1). An excessive level of claims will be prevented by the fact that a concept for the protection of the data received must be submitted to both the network provider and the supervisory authority along with the claim. In addition, the right to information will be mitigated by giving the provider grounds for refusing to provide information ("if his interests worthy of protection significantly outweigh the public interest in the research"). One obstacle to asserting the claim may

27 Eifert, Martin, "Rechenschaftspflichten für soziale Netzwerke und Suchmaschinen," *Neue Juristische Wochenschrift* (NJW) 2017: 1453; Löber, and Roßnagel, "Das Netzwerkdurchsetzungsgesetz," 75.

28 BT-Drs. 19/29392 (German Parliament Document).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

also be that the network provider is entitled to reimbursement for the costs incurred in providing the information; as a general rule, the law limits the reimbursement costs to EUR 5,000.

3. *Regulatory supervision*

a) Powers of intervention

§ 4a introduces a new supervisory regime. According to this, the Federal Office of Justice monitors compliance with the provisions of the NetzDG. From now on, this agency can take the necessary measures against providers in the event of violations of the NetzDG. Previously, the agency could only take repressive action in the form of fines. With its supervisory authority, the Federal Office of Justice can impose a forward-looking obligation on providers to put an end to a violation of the NetzDG without having to initiate fine proceedings.

The more flexible and constructive form of regulatory supervision now impleemented is to be welcomed. The threat of fines - one of the main points of criticism of the NetzDG - will thus lose its intimidating effect.

b) Duty to cooperate – Duty to report

A provision contained in the amendment that obliges network providers to report certain content to the law enforcement authorities has been met with much skepticism. This obligation applies to certain criminal offenses enlisted in § 1 NetzDG; this involves relatively serious offenses in whose prosecution there is a great interest, such as incitement to hatred, the formation of terrorist organizations, or the preparation of a serious act of violence that endangers the state. The idea is that law enforcement agencies will no longer be able to keep track of all the crimes committed on the Internet. Legislators hope that this will not only improve law enforcement, but also have a deterrent effect on users.

In the first draft, the procedure was designed in such a way that the network providers report such content to the Federal Criminal Police Office (*Bundeskriminalamt*); the IP address and port number are also reported; the Federal Criminal Police Office then assesses whether the content is actually relevant to criminal law. If the answer is affirmative, the full data identifying the person is requested from the social network. The

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

data is then forwarded to the law enforcement authorities by the Federal Criminal Police Office. There has been much criticism of this regulation. An expert opinion from the Bundestag's academic service pointed to data protection concerns.[29] Another expert opinion commissioned by the Green Party pointed to problems with regard to the timing of the data transfer. According to the opinion, the Federal Constitutional Court had only shortly before ruled that content should not be transmitted at the same time as the traffic data.[30] The legislative package was therefore not signed by the President (as the final required legislative act). In February 2021, the bill was updated[31] and approved a few month later[32].

Furthermore, this new regulation is also criticized for promoting the "privatization" of law enforcement.[33] It would practically make the social networks responsible for law enforcement. Of course, it cannot be denied that the social networks decide which content they forward to the Federal Criminal Police Office and which they do not. However, it must be seen that ultimately only the social networks are in a position to comprehensively monitor the content and install a comprehensive complaints management system, which makes them aware of the damaging content. In this respect, the entire problem of content regulation can be illustrated with the following: *The state wants to, but cannot - the social networks can, but do not want to*. Of course, the latter is only true as long as it is not financially profitable, but state paternalism is probably always an evil for businesses such as social networks.

---

29 Scientific service of the Bundestag, "Die Vereinbarkeit der Meldepflicht nach § 3a Abs. 4 NetzDG n.F. mit dem Recht der Sitzländer der Anbieter von sozialen Netzwerken und das Verhältnis der verschiedenen Einrichtungen der Entscheidungskontrolle nach NetzDG und JMStV," September 11, 2020 Document No. WD 10 - 3000 - 043/20.

30 Bäcker, Matthias, "Folgerungen aus dem zweiten Bestandsdatenbeschluss des BVerfG für die durch das Gesetz zur Bekämpfung des Rechtsextremismus und der Hasskriminalität geschaffenen Datenverarbeitungsregelungen," Opinion for the Green Party from September 16, 2020: 3.

31 Bundestag Drucksache 19/25294 (German Parliament Document).

32 Bundestag Drucksache 19/29392 (German Parliament Document).

33 Stefan Krempl, "NetzDG-Reform: Gesetzgeber verstrickt sich in unauflösbare Widersprüche," *heise online*, last modified June 17, 2020, https://www.heise.de/news/NetzDG-Reform-Gesetzgeber-verstrickt-sich-in-unaufloesbare-Widersprueche-4786964.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

4. Out-of-court conciliation

The Amending Act also contains a provision for the introduction of an out-of-court dispute resolution. This is intended to enable alternative dispute resolution on a low-key basis. The dispute resolution bodies are not required to have any special qualifications; the goal is merely to make a serious attempt to resolve disputes possible. The parties involved are to be made aware of this possibility after the complaint procedure and also the counter-proposal procedure have been carried out; it thus serves to settle a dispute at a rather late stage. Together with the regulated self-regulation, this represents an attempt by the legislator to also involve other actors from civil society in the issues of content management on social networks. It is unclear how such models will evolve. The state is using these approaches primarily for the purpose of dispelling the impression of state censorship, but also to limit the power of social networks.

IV. Outlook

The NetzDG has already done pioneering work and paved the way for similar laws on content regulation. The amendment makes meaningful additions and will likely allow for better law enforcement on social networks. Of course, the NetzDG has the insurmountable shortcoming that it only applies to Germany, which generally puts obstacles in the way of law enforcement. The European Digital Services Act is intended to put a lid on this under EU law and set standards throughout the Union.[34] Until then, the NetzDG fulfills its role as an experimental field; since Germany is also big enough to stand up to Facebook and Co. in some respects. So, here the famous saying from American constitutional law comes into play, that the states are the "laboratories" for the entire federal system.[35] The NetzDG has made clear what kind of rules are to be found in such regulations. Complaint management is at the heart of this, and transparency regulations enable control by the state and the public. One lesson that Germany had to learn only in the course of application is the inclusion of the network's own deletion guidelines.

---

34 See chapter in this book *Holznagel/Kalbhenn*.

35 New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Bibliography

Bäcker, Matthias. "Folgerungen aus dem zweiten Bestandsdatenbeschluss des BVerfG für die durch das Gesetz zur Bekämpfung des Rechtsextremismus und der Hasskriminalität geschaffenen Datenverarbeitungsregelungen." Opinion for the Green Party from September 16, 2020.

Eifert, Martin. "Rechenschaftspflichten für soziale Netzwerke und Suchmaschinen." *Neue Juristische Wochenschrift* (NJW) 2017, 1450-1454.

Facebook. "Community Standards Enforcement Report, Third Quarter 2020." Accessed July 6, 2021. https://about.fb.com/news/2020/11/community-standards-enforcement-report-nov-2020/.

Facebook. "NetzDG Transparenzbericht Januar 2021." Accessed July 6, 2021. https://about.fb.com/de/wp-content/uploads/sites/10/2021/01/Facebook-NetzDG-Transparenzbericht-Januar-2021.pdf.

Federal Ministry of Justice and Consumer Protection. "Bundesamt für Justiz (BfJ) erlässt Bußgeldbescheid gegen Facebook." Press release from July 3, 2019. Accessed July 6, 2021. https://www.bmjv.de/SharedDocs/Artikel/DE/2019/070319_Facebook.html.

FSM. "NetzDG." Accessed July 6, 2021. https://www.fsm.de/de/netzdg.

Gersdorf, Hubertus. "Hate Speech in sozialen Netzwerken." *MMR Zeitschrift für IT-Recht und Recht der Digitalisierung* (2017): 439-447.

Guggenberger, Nikolas. "Das Netzwerkdurchsetzungsgesetz – schön gedacht, schlecht gemacht." *Zeitschrift für Rechtspolitik ZRP* (2017): 98-101

Guggenberger, Nikolas. "Das Netzwerkdurchsetzungsgesetz in der Anwendung." *Neue Juristische Wochenschrift (NJW)* (2017): 2577-2582.

Heindorf, Manon. „Das Netzwerkdurchsetzungsgesetz in der Umsetzung: Zwei Jahre NetzDG – eine Bilanz." *Verwaltungsrundschau VR* (2020): 113-118.

Holznagel, Bernd. "Das Compliance-System des Entwurfs des Netzwerkdurchsetzungsgesetzes." *Zeitschrift für Urheber- und Medienrecht ZUM* (2017): 615-624.

Hoven, Elisa and Melena Krause. "Die Strafbarkeit der Verbreitung von ‚Fake News'." *Juristische Schulung (JuS)* (2017): 1167-1170.

Krempl, Stefan. "NetzDG-Reform: Gesetzgeber verstrickt sich in unauflösbare Widersprüche." *heise online*. Last modified June 17, 2020. https://www.heise.de/news/NetzDG-Reform-Gesetzgeber-verstrickt-sich-in-unaufloesbare-Widersprueche-4786964.html.

Ladeur, Karl-Heinz. "Ist der Regierungsentwurf eines NetzDG 2.0 vom 19.2.2020 netzgerecht?" *Kommunikation und Recht K&R* (2020): 248-253.

Löber, Lena Isabell and Alexander Roßnagel. "Das Netzwerkdurchsetzungsgesetz in der Umsetzung." *MMR* (2019): 71-76.

Liesching, Marc, Chantal Funke, Alexander Hermann, Christian Kneschke, Carolin Michnik, Linh Nguyen, Johanna Prüßner, Sarah Rudolph, Vivien Zschammer. *Das NetzDG in der praktischen Anwendung*. Berlin: Carl Grossmann Verlag, 2021. doi:10.24921/2021.94115953.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Niggemann, Sandra. "Die NetzDG-Novelle." *Computer und Recht CR* (2020): 326-331.

Peukert, Alexander. "Gewährleistung der Meinungs- und Informationsfreiheit in sozialen Netzwerken," *MMR* (2018): 572-578.

Schwartmann, Rolf. "Verantwortlichkeit Sozialer Netzwerke nach dem Netzwerkdurchsetzungsgesetz," *GRUR-Prax Praxis im Immaterialgüter- und Wettbewerbsrecht* (2017): 317-319.

Scientific service of the Bundestag. "Die Vereinbarkeit der Meldepflicht nach § 3a Abs. 4 NetzDG n.F. mit dem Recht der Sitzländer der Anbieter von sozialen Netzwerken und das Verhältnis der verschiedenen Einrichtungen der Entscheidungskontrolle nach NetzDG und JMStV." September 11, 2020 Document No. WD 10 - 3000 - 043/20.

Spindler, Gerald. "Der Regierungsentwurf zum Netzwerkdurchsetzungsgesetz – europarechtswidrig?" *Zeitschrift für Urheber- und Medienrecht ZUM* (2017): 473-487.

Wimmers, Jörg and Britta Heymann. "Zum Referentenentwurf eines Netzwerkdurchsetzungsgesetzes (NetzDG) – eine kritische Stellungnahme." *AfP Zeitschrift für das gesamte Medienrecht* (2017): 93-102.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Platform Governance at the Periphery: Moderation, Shutdowns and Intervention

Giovanni De Gregorio, Nicole Stremlau

Abstract: After illustrating how the spread of dangerous content has led to troubling consequences beyond digital boundaries, this chapter describes how online hate speech has become criminalised in the global south. It analyses Internet shutdowns to understand their socio-legal consequences, and explores the applicability of public international law and the humanitarian doctrine to information interventions.

Keywords: platform governance; global south; Africa; hate speech; internet shutdowns; information intervention; content moderation; disinformation; media; online speech

1. Introduction

The spread of online hate and disinformation is increasingly provoking dramatic and troubling consequences beyond digital boundaries. False information about health treatments during the Covid-19 pandemic,[1] or the use of social media in mobilizing actors for the attack on Capitol Hill,[2] are some prominent examples of how online speech can affect the general public. But offline harms are far broader and often less explicitly tied to online speech. Our focus here is on areas of the world that have not been considered 'priorities' by social media companies.[3] For example, in

1 Julie Posetti and Kalina Bontcheva, *Disinfodemic: deciphering COVID-19 disinformation. Policy brief 1*. (2020), https://en.unesco.org/covid19/disinfodemic.

2 Joan Donovan, Brian Friedberg and Emily Dreyfuss, "The Capitol siege was the biggest media spectacle of the Trump era," *The Guardian*, January 11 (2021) https://www.theguardian.com/commentisfree/2021/jan/11/the-capitol-siege-was-the-biggest-media-spectacle-of-the-trump-era.

3 In April 2021 the Guardian published an except from an email by a top Facebook executive explaining that the company should address concerns of abuse online by focusing on "top countries, top priority areas… and try to somewhat work our way down [to peripheral countries, or those that are seen as less strategic and driving

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the Central African Republic, online hate speech has contributed to mass atrocities between Christians and Muslims,[4] and in Sri Lanka, rumours on social media have led to a number of religious attacks, including the 2019 Easter Sunday church and hotel bombings,[5] while the use of Facebook in inciting violence against Myanmar's minority Muslim population has elevated concerns about the role of online platforms in perpetrating genocides.[6]

The *fil rouge* connecting these examples is the role of social media companies[7] in governing speech on a global scale.[8] Online platforms that process content rely on a mix of human moderators and artificial intelligence systems that define which content must be removed according to non-transparent standards and without explanation, providing very few opportunities for remedies.[9] As the global pandemic has altered working arrangements for human coders (along with many office workers) it has also made the implementation of artificial intelligence systems in content moderation more urgent for companies.[10] But this has brought to the fore

---

news]". *See:* https://www.theguardian.com/technology/2021/apr/12/facebook-loophole-state-backed-manipulation.

4 Office of the High Commissioner for Human Rights, *Preventing incitement to hatred and violence in the Central African Republic* (2019) https://www.ohchr.org/EN/NewsEvents/Pages/PeacekeepersDay2019.aspx.

5 Newley Purnell, "Sri Lankan Islamist Called for Violence on Facebook Before Easter Attacks," *Wall Street Journal*, April 30, 2019 https://www.wsj.com/articles/sri-lankan-islamist-called-for-violence-on-facebook-before-easter-attacks-11556650954.

6 Fanny Potkin and Poppy McPherson, "Spreading like Wildfire: Facebook Fights Hate Speech before Myanmar Poll," *Reuters*, November 5, 2020, https://www.reuters.com/article/myanmar-election-facebook-idUSL4N2HQ3QU.

7 When referring to 'social media' we are primarily speaking of user-generated content on large platforms such as Facebook, Twitter, YouTube or TikTok.

8 Hannah Bloch-Wehba, "Global platform governance: private power in the shadow of the state," *SMU L. Rev.* 72 (2019): 27; Tarleton Gillespie. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media* (Yale University Press, 2018), Kate Klonick, "The new governors: The people, rules, and processes governing online speech." *Harv. L. Rev.* 131 (2017): 1598; Luca Belli, David Erdos, Maryant Fernandez Perez, Pedro Augusto P. Francisco, Krzysztof Garstka, Judith Herzog, Krisztina Huszti-Orban et al., *Platform regulations: how platforms are regulated and how they regulate us* (Leeds, 2017).

9 Sarah T Roberts, *Behind the screen* (Yale University Press, 2019).

10 Sana Ahmad, "COVID-19 and the Future of Content Moderation," Coronavirus and its Societal Impact-Highlights from WZB Research, 2020. https://www.wzb.eu/en/research/corona-und-die-folgen/covid-19-and-the-future-of-content-moderation.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

just how problematic AI can be when it comes to effectiveness; during the pandemic there have been significant cases when the involvement of human moderators was restricted and an over-reliance on the automated system led to the spread of disinformation and blocking of accounts that were actually countering disinformation.[11]

Against this opaque framework of governance and fragmented responses by social media companies, a variety of actors, from non-governmental organizations to various public authorities around the world have tried to tackle the harm produced by the spread of hate and disinformation online.[12] Governments have reacted in different ways, particularly in poorer and less geopolitically influential countries. They have accused online platforms of disseminating hate and disinformation online, criminalised the spread of hate and disinformation,[13] have used platforms for surveillance, worked to push alternative narratives (sometimes flooding platforms with disinformation), and have attempted to censor content.[14] The spread of hate on social media has also been one of the primary reasons why governments have increasingly justified the use of Internet shutdowns, which can involve a range of tools from slowing down the internet (making it practically unusable) to completely switching it off.[15] Whereas only a few years ago such forms of censorship would have been seen as a grave violation of freedom of expression, increasingly they being seen to be one of the few mechanisms available for addressing online speech and offline harms in a moment of crisis.

The role of media in contributing to disseminate hate and violence is not new.[16] In some cases of violence and mass atrocities, international actors, including the United Nations (UN), have relied on "information

---

11 Statt Nick, "How Facebook Is Using AI to Combat COVID Misinformation and Detect 'Hateful Memes,'" The Verge, May 12, 2020, https://www.theverge.com/2020/5/12/21254960/facebook-ai-moderation-covid-19-coronavirus-hateful-memes-hate-speech.

12 Roxana Radu, Fighting the 'Infodemic': Legal Responses to COVID-19 Disinformation. *Social Media+ Society*, *6*(3), 2020.

13 Dickens Olewe, "Kenya, Uganda and Tanzania in "anti-fake news campaign," *BBC News*, 16 May 2018.

14 Adrian Shahbaz, "The Rise of Digital Authoritarianism: Freedom on the Net 2018," *Freedom House*, October (2018) https://freedomhouse.org/report/freedom-net/2018/rise-digital-authoritarianism.

15 De Gregorio, Stremlau, " Internet Shutdowns and the Limits of Law".

16 Robert Edwin Herzstein. *The war that Hitler won: The most infamous propaganda campaign in history* (Putnam Publishing Group, 1978). Nicole Stremlau. *Media, Conflict and the State in Africa* (Cambridge University Press, 2018).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

interventions," an expression developed in the 1990s in response to the conflict in Rwanda and the Balkans.[17] While information interventions have been applied to traditional media outlets, we ask whether such a response could be relevant for social media, particularly when online platforms have a leading role in disseminating content directly associated with mass atrocities.

Within this framework, this chapter explores the challenges raised by online hate and disinformation in areas of the world that are less of a business priority for large social media companies. By focusing on content moderation as an expression of platform governance, we underline how the spread of online hate and disinformation have led to troubling consequences beyond digital boundaries. In the first part, we focus on the criminalisation of online hate and disinformation as a response to the consequences this content produces in the online and offline world. The second part explores how the spread of online hate speech and disinformation has provided governments with further justifications, that are increasingly becoming internationally acceptable (or at least understood), to censor the Internet for protecting national security or other public interests. The third part focuses on the role of international actors in addressing the spread of online hate and disinformation by looking at the applicability of the doctrine of information intervention to social media.

Our focus in this chapter is on the variety of legal and censorship responses to online hate. We recognize that there are considerable efforts on the part of governments to address online speech with different techniques ranging from attempting to shift narratives through flooding social media with specific content (as seen with the role of Cambridge Analytica), or using surveillance and both online and offline coercion or harassment to silence certain voices. In this chapter, however, our emphasis is on the intersection of concerns around content moderation and the use of law or force to address these concerns.

17 Monroe E Price and Mark Thompson, eds. *Forging peace: intervention, human rights, and the management of media space* (Indiana University Press, 2002); Jamie F. Metzl, "Information intervention: When switching channels isn't enough," *Foreign Affairs* (1997): 15-20.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2. An initial response: Criminalising online hate and disinformation

Social media companies have a critical role in determining the standards of protection of online speech on a global scale. Although these companies do not always have offices in the country where hate and mass atrocities are perpetrated, they exercise broad discretion in determining the rules according to what information circulates online and, therefore, how content is shared between communities.[18] And this does not change even in situations of conflicts or violence where these actors have determine how to moderate hate and disinformation according to their ethical, business and legal framework.

This process, with its strengths and limitations, was evident during the Arab Spring.[19] As observed by Zeitoff,[20] communications in conflicts have typically been defined in two ways: "elite-level communication" focused on tactical and logistical aims; and "mass-based appeals" aimed at coordinating or inhibiting public behaviour through control of the narrative and manipulating mass channels of communication.[21] Social media provide a new paradigm, transforming users into active creators of content whose standard of protection is defined by private companies. This increasing degree of protection of online speech can empower users in authoritarian regimes while affecting social tensions and conflicts.[22] The disintermediation of traditional media outlets allows individuals to challenge elite-dominated discourse, especially in authoritarian regimes, which tend to exercise public control over traditional media outlets. Information spread on social media can be immediately shared with other communities of users, potentially going viral. The digital spaces provided by social media have encouraged access to diverse information online, promoting a plurality of voices and sharing of opinions. In particular, the possibility to use these channels to contest central authorities and spread disinformation has

18 Dimitra Dimitrakopoulou, Georgios Tzogopoulos and Alexandra Nikolakopoulou, *The Role of Social Media in Violent Conflict* (INFOCORE Working Paper 2014/05).

19 Philip N. Howard and Muzammil M. Hussain, *Democracy's Fourth Wave?: Digital Media and the Arab Spring* (OUP 2013).

20 Thomas Zeitzoff, "How social media is changing conflict," *Journal of Conflict Resolution* 61, no. 9 (2017): 1970-1991.

21 Philip N. Howard. *The digital origins of dictatorship and democracy: Information technology and political Islam* (Oxford University Press, 2010).

22 Peter Dahlgren, "The Internet, public spheres, and political communication: Dispersion and deliberation," *Political communication* 22, no. 2 (2005): 147-162.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

encouraged governments to censor online speech or even use social media as instrument of surveillance.[23]

The use of automated technologies for moderating content also produces effects that extend beyond domestic boundaries.[24] These channels of communication allow information to be disseminated more widely and with greater speed, especially in cases involving strong messages of hate or dissent. Algorithmic content moderation contributes to driving people to online hate and disinformation,[25] which can also lead to discrimination.[26] As underlined by Tufekci, "YouTube may be one of the most powerful radicalizing instruments of the 21st century".[27] In areas characterised by tensions and conflicts, this can inflame and escalate violence and conflicts - the lack of language training in certain languages makes content moderation less effective in detecting online hate speech. The Myanmar genocide has underlined the inability of Facebook to detect and limit the spread of hate speech.[28] The spread of hate speech on Facebook supported ethnic cleansing in Myanmar, but this went mostly unchecked due to the lack of moderation tools and human moderators fluent in Burmese. While Facebook significantly expanded its team of Burmese speakers to create a data set of hate and violent expressions, the international pressure to act also led to overreactions including the ban of some armed groups.[29]

Given these challenges, the first reaction by many governments has been to criminalise the spread of online hate and disinformation by users and social media. In May 2019, Singapore adopted the Protection from

23 Evgeny Morozov, *"The net delusion: The dark side of Internet freedom," PublicAffairs*, 2012.

24 Jack M. Balkin, "Free speech in the algorithmic society: Big data, private governance, and new school speech regulation," *UCDL Rev.* 51 (2017): 1149

25 Jack Nicas, "How YouTube Drives People to the Internet's Darkest Corners" *Wall Street Journal*, Feb. 7 2018 https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478.

26 Safiya Umoja Noble, *Algorithms of oppression: How search engines reinforce racism* (NYU Press, 2018).

27 Zeynep Tufecki, "YouTube. The Great Radicalizer" *New York Times*, May 10, 2018 https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

28 Steve Stecklow, "Why Facebook is losing the war on hate speech in Myanmar" *Reuters*, Aug. 15 2018 https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/.

29 Jeffrey Sablosky, "Dangerous organizations: Facebook's content moderation decisions and ethnic visibility in Myanmar" *Media, Culture & Society* (2021).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Online Falsehoods and Manipulation Act.[30] The scope of this legislation covers content that is false or misleading, whether wholly or in part and/or there are reasons to believe it affects public interest. The prohibition of communication of "false statements of fact" in Singapore applies to both individuals and online intermediaries applying a fine from S$ 20,000 (12,000 euro) up to S$ 100,000 (62,000 euro) and/or imprisonment from 1 to 10 years, whereas for intermediaries they generally range between S$ 500,000 (310,000 euro) to S$ 1 million (622,000 euro).

Malaysia similarly followed the path towards the criminalization of online disinformation even if the government decided to repeal the legislation after its adoption.[31] Nonetheless, in March 2021, the Perikatan National Government enacted an emergency ordinance using powers conferred by a January 2021 Emergency Proclamation to face the spread of online disinformation about Covid-19 or the proclamation of the emergency.[32] This measure introduces new criminal offences relating to the creation, publication, or dissemination of so-called 'fake news' and the failure to take down publications containing content deemed as 'fake news'. This conduct is sanctioned with up to three years imprisonment. Furthermore, individuals and internet platforms which do not remove content within 24 hours based on an order coming from public officials, not necessarily courts, can be sanctioned with a fine of up to 100,000 Malaysian ringgit (20,000 euro) and, in the case of a continuing offense, up to 300,000 ringgit (60,000 euro) for every day in which the content is available.

Moving from Asia to Africa, Ethiopia passed a law sanctioning the spread of online hate by Internet users and platforms providing up to three years of imprisonment and a fine of up to 100,000 birrs (2,900 euro).[33] In justifying this legislation, reference was made to the central role of hate speech and electoral related violence in neighboring Kenya as well as the introduction of the Network Enforcement Act (NetzDG) in Germany which is regarded as an ambitious legislation requiring platforms to re-

30 Protection from Online Falsehoods and Manipulation Bill (2019), https://sso.agc.gov.sg/Bills-Supp/10-2019/Published/20190401?DocDate=20190401.

31 Anti-fake News Act 2018, https://perma.cc/Y5H3-D6G8.

32 Malaysia's king declares state of emergency to curb spread of COVID-19, ABC News, https://www.abc.net.au/news/2021-01-12/malaysia-king-declares-state-of-emergency-to-curb-covid-spread/13051642.

33 Proclamation No. 1185 /2020 Hate Speech and Disinformation Prevention and Suppression Proclamation, https://chilot.me/wp-content/uploads/2020/04/HATE-SPEECH-AND-DISINFORMATION-PREVENTION-AND-SUPPRESSION-PROCLAMATION.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

move hate speech within 24 hours or face fines up to 50 million euros.[34] In the case of Nigeria, the fight against online hate speech has been even more radical. In 2019, two bills were proposed to increase government powers to shut down the internet, punish government critics and sanction hate speech with capital punishment.[35]

Among other approaches, the political choice of Uganda concerning how to restrict free speech online has been different. Since July 2018, a new Ugandan tax charges citizens 5 US cents a day for the use of 60 mobile apps, including Facebook, Twitter, Skype and WhatsApp. This social-media tax was passed as part of a bill that also includes taxes on mobile transactions and was seen as a way of attempting to reduce the use of these platforms. Many Ugandans, however, chose to access them from other internet connections (rather than mobile data) or use VPNs to get around the restrictions in place by the mobile operators.

These measures are just a small part of the array of new laws attempting to shape speech on social media but not necessarily limiting access to the internet in its entirety. The next section explores a blunter and far reaching tool of public censorship as a second reaction to the spread of online hate and disinformation. As the next section underlines, governments are increasingly relying on Internet shutdowns, thus, leading to a process of normalisation of these practices as a reaction to the spread of online hate and disinformation on social media.

3. Internet shutdowns and the control of narratives

The spread of online hate and disinformation on social media is increasingly considered by some governments to be a justification (or legitimate aim) to censor speech and shut down the Internet. This is often perceived as the only immediately effective remedy to deal with the escalation of violence in the context of company-led discretion in responding and moderating content. Even though there is very limited evidence about the effects of these practices in tackling the misinformation and hate they purport to address, shutdowns have been implemented to curtail online

34 Network Enforcement Act, 2018, https://germanlawarchive.iuscomp.org/?p=1245.
35 Nigeria bill aims at punishing hate speech with death, https://www.dw.com/en/nigeria-bill-aims-at-punishing-hate-speech-with-death/a-51419750.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

speech, and particularly content that is seen to be provoking violence or promoting dissent.[36]

Internet shutdowns have increased in scale and scope over several years[37], particularly in Asia and Africa.[38] From India, where there have been many localized Internet shutdowns,[39] to Cameroon, a country that brazenly blocked access in half of the country for more than 230 days between 2017 and 2018,[40] shutting down the Internet (either partially or entirely) appears to be used by governments when they want to act quickly, particularly to quell perceived or potential civil unrest, and might have limited capacity for other mechanisms of online control. The rise of internet shutdowns also reflects a frustration on the part of some governments with their inability to intervene in the governance of the digital platforms that are often controlled by businesses in another continent. In the absence of concerted cooperation with companies, shutting down the entire network or specific digital spaces has become increasingly popular. While the ire and frustration coming from countries such as New Zealand, Germany, or France toward Facebook or Twitter's inability to control disinformation and hate speech has been significant, they have also found more engagement at company headquarters. This may be because poorer countries and those that typically resort to Internet shutdowns have far less leverage over the large American companies. It is helpful to keep in mind that the GDP of a country like Burundi is approximately 3 billion USD while the value of Facebook is roughly 240 times that at 720 billion

---

36 Statista. "Government Justifications for Internet Shutdowns Worldwide 2019." Accessed March 25, 2021, https://www.statista.com/statistics/1096316/government-justifications-for-internet-shutdowns/#:~:text=Official%20government%20justifications%20for%20internet%20shutdowns%20worldwide%202019&text=Fake%20news%20and%20hate%20speech.

37 In 2020 it was estimated that there were at least 155 shutdowns in 29 countries, down from 213 incidents in 2019 (https://www.accessnow.org/keepiton/).

38 For an overview of trends on internet shutdowns in Africa see: Eleanor Marchant and Nicole Stremlau, "The Changing Landscape of Internet Shutdowns in Africa", International Journal of Communication, 14(2020), 4216-4223 and Eleanor Marchant and Nicole Stremlau, "A Spectrum of Shutdowns: Reframing Internet Shutdowns from Africa" International Journal of Communications 14(2020): 4327-4342.

39 Megha Bahree, "India leads the world in the number of Internet shutdowns: Report", *Forbes*, November 12, 2018 https://www.forbes.com/sites/meghabahree/2018/11/12/india-leads-the-world-in-the-number-of-internet-shutdowns-report/.

40 Abdi Latif Dahir, "Africa Internet shutdowns grow longer in Cameroon, Chad, Ethiopia," Quartz Africa, November 19, 2018. https://qz.com/africa/1468491/africa-internet-shutdowns-grow-longer-in-cameroon-chad-ethiopia/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

USD. Given these severe inequalities it is not surprising that complaints from countries in Africa have scarce reception in Silicon Valley. In fragile states, the lack of negotiating powers of governments in respect of social media underline the power that these actors can exercise, thus, making shutdowns an apparent necessity to censor online speech.

When Internet shutdowns occur, they are usually met with condemnation by free speech advocates and Internet freedom groups such as Access Now.[41] The effects of Internet shutdowns by virtue of the role of the digital environment in today's society cannot be neglected. Domestic deterrents, such as arguments around potential economic costs, appear to have little impact (particularly if governments are weighing up the comparative economic costs of protests or unrest), and advocacy groups that focus on publicly shaming governments have not reduced the use of shutdowns. The Internet is not only relevant from a technical or economic perspective,[42] but also for the exercise of democratic values such as assembly and freedom of expression and, therefore, as a crucial source of information and knowledge.

A polarized debate has emerged with governments grasping for ways to control flows of misinformation and hate speech, sometimes with legitimate concerns and frustration over their inability to control the vast amount of user generated content, the tepid engagement or responses social media companies to address this issue, and the forceful (and unbending) condemnation of Internet shutdowns by advocacy groups and the human rights community. This can make it difficult to have a nuanced conversation about when and under what circumstances shutdowns might be justified. While there is a lack of transparency and accountability of states when shutting down the Internet, including justification of the reasons or the procedures on which these restrictive measures are implemented, there have been some efforts to map the reasons governments have provided. The majority of explanations reference national security, including political mobilization or protest.[43] Election periods are another

41 Access Now, "The state of Internet shutdowns around the world the 2020", *#KEEPITON Report*, 2021 https://www.accessnow.org/cms/assets/uploads/2021/03/KeepItOn-report-on-the-2020-data_Mar-2021_3.pdf.

42 The Organisation for Economic Co-operation and Development, "OECD digital economy outlook 2017," https://www.oecd.org/sti/ieconomy/oecd-digital-economy-outlook-2017-9789264276284-en.htm.

43 Lynsey Chutel, "Zimbabwe's government shut down the Internet after fuel price protests turned deadly," *Quartz Africa,* January 15, 2019 https://qz.com/africa/1524405/zimbabwe-protest-internet-shut-down-military-deployed-5-dead/; Peter Micek

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

highly contested period.[44] In some cases, targeted shutdowns have been regionally specific whereby governments have tried to marginalize specific groups that may, for example, be attempting publicize human rights violations or may be protesting the absence of government service delivery in peripheral regions. And Internet shutdowns have also been implemented for more benign seeming issues, such as before school exams to prevent cheating.[45]

Unlike social media which are not bound to respect human rights according to international human rights law, states have an obligation to respect human rights according to covenants and customary international law that protects the right to freedom of expression limiting the shutting down of the digital environment. In January 2020, the Supreme Court of India recognised that freedom expression online enjoys constitutional protection,[46] even if this decision has not changed the general approach to Internet shutdowns in India. In January 2019, a Zimbabwean court ruled that government's internet shutdown as an answer to protests was illegal.[47] Similarly, in June 2020, the Economic Community of West African States (ECOWAS) Community Court decided that, by shutting down the Internet during the anti-government protests in 2017, the Togolese government violated human rights.[48] According to the court, the arguments based on

---

and Deji Olukotun, "Internet disrupted in Bahrain around protests as wrestling match sparks shutdown in India," *Access Now*, 24 June 2016 https://www.accessnow.org/internet-disrupted-bahrain-around-protests-wrestling-match-sparks-shutdown-india/; Philip N. Howard, Sheetal D. Agarwal, and Muzammil M. Hussain, "When do states disconnect their digital networks? Regime responses to the political uses of social media," *The Communication Review* 14, no. 3 (2011): 216-232.

44 Hilary Matfess, "More African countries are blocking internet access during elections," *Quartz Africa*, June 1, 2016 https://qz.com/africa/696552/more-african-countries-are-blocking-internet-access-during-elections/; Deji Olukotun, Peter Micek, and Gustav Bjorksten, "Vietnam blocks Facebook and cracks down on human rights activists during Obama visit," *Access Now*, 23 May 2016. https://www.accessnow.org/vietnam-blocks-facebook-human-rights-obama/.

45 Nour Youssef, "Algeria's answer to cheating on school exams: Turn off the Internet," *The New York Times*, June 21, 2018 Retrieved from https://www.nytimes.com/2018/06/21/world/africa/algeria-exams-cheating-internet.html.

46 Anuradha Bhasin vs Union of India & Ors. Writ Petition (Civil).

47 MacDonald Dzirutwe, Zimbabwe court says internet shutdown illegal as more civilians detained https://www.reuters.com/article/us-zimbabwe-politics/zimbabwe-court-says-internet-shutdown-during-protests-was-illegal-idUSKCN1PF11M.

48 Amnesty International et al. v. The Togolese Republic, 2020, https://www.accessnow.org/cms/assets/uploads/2020/07/ECOWAS_Togo_Judgement_2020.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

national security could not justify the internet shutdown according to local or international law. This, however, does not mean that states cannot rely on legitimate interests to rely on shutdowns for example in cases of self-defence. Although there are different nuances of freedom of expression in regional human rights instruments and areas of the world, the Universal Declaration of Human Rights (UNDHR) and the International Covenant on Civil and Political Rights (ICCPR) are the primary structures to take into account for the three step-test based on legality, legitimacy and proportionality of the actions public authorities may take. Together, they can have a role in mitigating the rise of Internet shutdowns.

Despite the potential relevance of these legal procedures, the law has limitations when applied to Internet shutdowns. The scope of applicable regulation and legitimate interests could shape this framework which can be broadly exploited for political purposes. These concerns are particularly relevant in authoritarian regimes since the limits of the law in relation to Internet shutdowns are not only about the boundaries of the three-step test but also concern the scrutiny of these practices. The challenges posed by Internet shutdowns is also due to the lack of a common international enforcement mechanism that allows for both the transparent implementation of processes and procedures for when shutdowns might be justified, as well as the scrutiny of when shutdowns might be applied inappropriately.

4. Building consensus on interventions

This challenge of international coordination and the legitimacy, or illegitimacy, of shutdowns (even in cases when online speech is connected with extreme violence such as genocide) brings us to our third area of focus. Internet shutdowns cannot be a general remedy due to the violations of international human rights law, and even if these violations were not the case, shutdowns would still not be a preferred tool. The growing prominence of social media in spreading hate and inciting violence prompts questions about whether, and to what extent, international law and cooperation can offer new options. The role of media in disseminating hate and violence has been a longstanding aspect of violent conflict.[49] In the last thirty years, such (mis)use of media has exacerbated numerous wars

49 Cees Jan. Hamelink. *Media and conflict: Escalating evil* (Routledge, 2015); Thompson and Price, *Forging peace: intervention, human rights, and the management of media space*, (2002).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

and violent conflicts, and in some cases even genocide like in Rwanda and Bosnia.[50] In the past, international actors, including the United Nations, have intervened in the media environment by implementing measures under the broad umbrella of "information intervention."[51] Information interventions are strategic efforts to interfere in (whether disrupting, manipulating or altering) a communications environment within a community, region or state afflicted by mass atrocities, in order to prevent or counter the dissemination of violence-inciting speech. The intervention can take place at various stages of a conflict and it can involve subsidizing or countering messages (through, for example, counter-narratives or providing support to certain media outlets- so-called 'peace media') or it may involve the direct closure of particular outlets such as the bombing of radio towers or the shuttering of a newspaper.

Information interventions are complex and political endeavours as much as legal ones. They must navigate international law, particularly the principle of non-intervention as expression of national sovereignty, the protection of human rights (i.e. freedom of expression). Such interventions, however, would get their legitimacy from humanitarian norms advancing the responsibility to protect (R2P),[52] and Chapter VII of the United Nations Charter with respect to the threats to the peace and acts of aggression. While the boundaries of the non-intervention principle raise the question of whether information interventions can be justified when seeking to prevent mass atrocities provoked by hate speech and disinformation, international law does not preclude the UN Security Council deciding what kind of speech or incitement satisfies the threshold required to trigger the Chapter VII mechanism. Therefore, while the spread of hate speech and disinformation may lead to conflicts and mass atrocities, the degree of danger may not be considered a threat to international peace and security.

Further challenges are political – gaining consensus on an information intervention is likely to be challenging. The responsibility to protect a regime does address whether and to what extent the international community should intervene in situations where state actors fail (voluntary or involuntary) to protect their population from mass atrocities or genocide.[53] In the absence of UN authorization, interventions cannot be legally based

50 Article 19, "Broadcasting genocide Censorship, propaganda & state-sponsored violence in Rwanda, 1990-1994," *Article* 19, London (United Kingdom, 1996).
51 Metzl, *Information intervention: When switching channels isn't enough*, 1997.
52 Alex J. Bellamy, *Responsibility to protect: A defense* (OUP Oxford, 2014).
53 Ibid.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

on R2P and/or humanitarian reasons thus constituting the most relevant challenge to information intervention. This authorization constitutes the ultimate safeguard to avoid that compelling reasons (or excuses) are used to interfere with states' sovereignty. But in recent years, stemming from the challenging (and ultimately failed) intervention in Libya in 2011, the responsibility to protect has been criticized as being a cover for politically motivated interventions and advocates for invoking interventions based on the responsibility to protect have struggled to get traction within the UN Security Council.

Despite these challenges, information interventions have been applied to traditional media outlets, and in the current climate is important to consider its potential relevance to online communications, and social media in particular. As already underlined, Chapter VII of the UN Charter can be used to authorise international interventions in the media environment of a target state without violating the principle of non-intervention. In cases where social media are involved in the escalation of violent conflicts, particularly mass atrocities such as genocide (as we have seen in Myanmar), the UN Security Council could, in theory, authorise an intervention under Chapter VII due to a breach of international peace and security. In this situation, an independent international body (which we will refer to as an Information Intervention Council) could be involved in limiting access to social media as part of its response to addressing mass atrocities and, as a remedy of last resort, shutting down the Internet.

At first glance, UN authorization could provide a way to extend the doctrine of information intervention to social media promoting online hate and disinformation. Nonetheless, any information intervention measure must take into consideration the network architecture and modalities through which it is possible to limit dissemination of online hate and violence with specific regard to Internet shutdowns. In this case, the cooperation between the international community and social media is critical. For example, social media could remove content or block accounts based on the recommendation of the Information Intervention Council. This would help to foster a more positive framework of content moderation, with greater safeguards to avoid arbitrary internet shutdowns as well as greater care on the part of social media actors to avoid having their activities shut down by the intervention of the external Council.

However, moving towards information interventions risks collateral censorship, particularly in conflict-affected countries where citizens may have significant needs for accurate and plural information sources. Unlike traditional media outlets, which operate within a specific region and have an important role in providing information to those in that area, inter-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

national social media platforms are driven by business incentives. As a consequence, social media companies may be motivated to cease operating in riskier regions where information interventions might be enacted which may lead to financial and reputational losses.

Information interventions are political as much as they are legal. There are, of course, risks with any intervention and particularly with one interfering with a information space. The line between information intervention and censorship can become blurred, with the real test being whether or not the measures address the responsibility to protect.

5. Conclusion

The governance of online speech is increasingly being shaped by a mix of public and private policies in an ad hoc and (often) arbitrary manner. Efforts by social media platforms have demonstrated the challenges of governing speech transnationally, particularly as their approach to moderating content is driven by business purposes rather than human rights norms. This leads to a clash between private interests focusing on profit and public values and the tension between protecting free speech while balancing conflicting rights and freedoms.

The offline harms associated of hate speech are a central justification as to why governments have proposed to criminalise online hate and disinformation, and have, at times, turned to blunter mechanisms, such as internet shutdowns, to regulate content online. Escalating concerns between online content and offline harms calls for urgent action, particularly by independent bodies such as the United Nations. The doctrine of information intervention offers one starting point to think about the potential role and responsibilities of international actors to intervene and address the most severe, or egregious cases, where online speech is leading to mass atrocities and human rights abuses such as genocide.

Bibliography:

Access Now, The state of Internet shutdowns around the world the 2020 #KEEPITON Report. (2021). https://www.accessnow.org/cms/assets/uploads/2021/03/KeepItOn-report-on-the-2020-data_Mar-2021_3.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Ahmad, Sana. *COVID-19 and the Future of Content Moderation*. Coronavirus and its Societal Impact- Highlights from WZB Research, 2020. Accessed March 25, 2021. https://www.wzb.eu/en/research/corona-und-die-folgen/covid-19-and-the-future-of-content-moderation.

Article 19, London (United Kingdom); *Broadcasting genocide Censorship, propaganda & state-sponsored violence in Rwanda, 1990-1994*. 1996.

Bahree, Megha. "India leads the world in the number of Internet shutdowns: Report." *Forbes*, November 12, 2018.

Balkin, Jack M. "Free speech and hostile environments." *Columbia Law Review* (1999): 2295-2320.

Balkin, Jack M. "Free speech in the algorithmic society: Big data, private governance, and new school speech regulation." *UCDL Rev.* 51 (2017): 1149.

Barendt, Eric. "Freedom of expression in the United Kingdom under the Human Rights Act 1998." *Ind. LJ* 84 (2009): 851.

Bellamy, Alex J. *Responsibility to protect: a defense*. OUP Oxford, 2014.

Belli, Luca, David Erdos, Maryant Fernandez Perez, Pedro Augusto P. Francisco, Krzysztof Garstka, Judith Herzog, Krisztina Huszti-Orban et al. *Platform regulations: how platforms are regulated and how they regulate us*. Leeds, 2017.

Bloch-Wehba, Hannah. "Global platform governance: private power in the shadow of the state." *SMU L. Rev.* 72 (2019): 27.

Bozdag, Cigdem. "Managing Diverse Online Networks in the Context of Polarization: Understanding How We Grow Apart on and through Social Media." *Social Media+ Society* 6, no. 4 (2020): 2056305120975713.

Cheng, Sage, Felicia Anthonio and Berhan Taye. #KeepItOn: Internet shutdowns put lives at risk during COVID-19. *Access Now*, 26 May 2020. https://www.accessnow.org/keepiton-internet-shutdowns-put-lives-at-risk-during-covid-19/.

Chutel, Lynsey. "Zimbabwe's government shut down the Internet after fuel price protests turned deadly." *Quartz Africa*, January 15, 2019.

Cohen, Julie E. *Between truth and power: The legal constructions of informational capitalism*. Oxford University Press, 2019.

Dahir, Abdi Latif. "Africa Internet shutdowns grow longer in Cameroon, Chad, Ethiopia." *Quartz Africa*, November 19, 2018.

Dahir, Abdi Latif. "Half the world is now connected to the internet–driven by a record number of Africans." *Quartz Africa*, December 11, 2018.

Dahlgren, Peter. "The Internet, public spheres, and political communication: Dispersion and deliberation." *Political communication* 22, no. 2 (2005): 147-162.

De Gregorio, Giovanni, and Nicole Stremlau. "Internet Shutdowns and the Limits of Law." *International Journal of Communication* 14 (2020): 20.

Dimitrakopoulou, Dimitra, Georgios Tzogopoulos and Alexandra Nikolakopoulou. *The Role of Social Media in Violent Conflict*, INFOCORE Working Paper 2014/05, (2014).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Donovan Joan, Brian Friedberg and Emily Dreyfuss. “The Capitol siege was the biggest media spectacle of the Trump era,” *The Guardian*, January 11 (2021) https://www.theguardian.com/commentisfree/2021/jan/11/the-capitol-siege-was-the-biggest-media-spectacle-of-the-trump-era.

Festinger, Leon. *A theory of cognitive dissonance*. Vol. 2. Stanford University Press, 1962.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press, 2018.

Hamelink, Cees Jan. *Media and conflict: Escalating evil*. Routledge, 2015.

Herzstein, Robert Edwin. *The war that Hitler won: The most infamous propaganda campaign in history*. Putnam Publishing Group, 1978.

Howard, Philip N. *The digital origins of dictatorship and democracy: Information technology and political Islam*. Oxford University Press, 2010.

Howard, Philip N., Sheetal D. Agarwal, and Muzammil M. Hussain. "When do states disconnect their digital networks? Regime responses to the political uses of social media."*The Communication Review* 14, no. 3 (2011): 216-232.

Klonick, Kate. "The new governors: The people, rules, and processes governing online speech."*Harv. L. Rev.* 131 (2017): 1598.

Marchant, Eleanor and Nicole Stremlau. “The Changing Landscape of Internet Shutdowns in Africa”, International Journal of Communication, 14(2020), 4216-4223

Marchant, Eleanor and Nicole Stremlau. “A Spectrum of Shutdowns: Reframing Internet Shutdowns from Africa” International Journal of Communications 14(2020): 4327-4342.

Matfess, Hilary. "More African countries are blocking internet access during elections”, *Quartz Africa*, June 1, 2016.

Metzl, Jamie F. "Information intervention: When switching channels isn't enough." *Foreign Affairs* (1997): 15-20.

Micek, Peter and Deji Olukotun. "Internet disrupted in Bahrain around protests as wrestling match sparks shutdown in India." *Access Now,* 24 June 2016.

Morozov, Evgeny. The net delusion: The dark side of Internet freedom. *PublicAffairs*, 2012.

Nicas, Jack. “How YouTube Drives People to the Internet’s Darkest Corners" *Wall Street Journal*, Feb. 7 2018 https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478.

Noble, Safiya Umoja. *Algorithms of oppression: How search engines reinforce racism*. NYU Press, 2018.

The Organisation for Economic Co-operation and Development“OECD digital economy outlook 2017, (2017). https://www.oecd.org/sti/ieconomy/oecd-digital-economy-outlook-2017-9789264276284-en.htm.

Office of the High Commissioner for Human Rights. *Preventing incitement to hatred and violence in the Central African Republic*. https://www.ohchr.org/EN/NewsEvents/Pages/PeacekeepersDay2019.aspx.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Olewe, Dickens. "Kenya, Uganda and Tanzania in “anti-fake news campaign.” *BBC News*, 16 May 2018.

Olukotun, Deji, Peter Micek and Gustav Bjorksten. "Vietnam blocks Facebook and cracks down on human rights activists during Obama visit." *Access Now*, 23 May 2016.

Potkin, Fanny and Poppy McPherson. “‘Spreading like Wildfire: Facebook Fights Hate Speech before Myanmar Poll,” *Reuters*, November 5, 2020 https://www.reuters.com/article/myanmar-election-facebook-idUSL4N2HQ3QU.

Posetti, Julie and Kalina Bontcheva. "Disinfodemic: deciphering COVID-19 disinformation. Policy brief 1," 2020. https://en.unesco.org/covid19/disinfodemic.

Price, Monroe E. and Mark Thompson, eds. *Forging peace: intervention, human rights, and the management of media space*. Indiana University Press, 2002

Purnell, Newley. "Sri Lankan Islamist Called for Violence on Facebook Before Easter Attacks."*Wall Street Journal*, April 30, 2019. https://www.wsj.com/articles/sri-lankan-islamist-called-for-violence-on-facebook-before-easter-attacks-11556650954.

Radu, Roxana. Fighting the ‘Infodemic’: Legal Responses to COVID-19 Disinformation. *Social Media+ Society*, *6*(3), 2020. 2056305120948190.

Roberts, Sarah T. *Behind the screen*. Yale University Press, 2019.

Sablosky, Jeffrey. “Dangerous organizations: Facebook’s content moderation decisions and ethnic visibility in Myanmar” *Media, Culture & Society* (2021).

Shahbaz, Adrian. "The Rise of Digital Authoritarianism: Freedom on the Net 2018." *Freedom House,* October 2018.

Statista. “Government Justifications for Internet Shutdowns Worldwide 2019.” Statista. https://www.statista.com/statistics/1096316/government-justifications-for-internet-shutdowns/#:~:text=Official%20government%20justifications%20for%20internet%20shutdowns%20worldwide%202019&text=Fake%20news%20and%20hate%20speech. Accessed March 25, 2021.

Statt, Nick. “How Facebook Is Using AI to Combat COVID Misinformation and Detect ‘Hateful Memes.” *The Verge*, May 12, 2020. Accessed March 25, 2021. https://www.theverge.com/2020/5/12/21254960/facebook-ai-moderation-covid-19-coronavirus-hateful-memes-hate-speech.

Stecklow, Steve. “Why Facebook is losing the war on hate speech in Myanmar” *Reuters*, Aug. 15 2018 https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/.

Stremlau, Nicole. *Media, Conflict and the State in Africa*. Cambridge University Press (2018).

Tufecki, Zeynep. “YouTube. The Great Radicalizer” *New York Times*, May 10, 2018 https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.

Youssef, Nour. "Algeria’s answer to cheating on school exams: Turn off the Internet." *The New York Times*, June 21, 2018.

Zeitzoff, Thomas. "How social media is changing conflict." *Journal of Conflict Resolution* 61, no. 9 (2017): 1970-1991.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Protecting the Freedom of Expression in an Era of "Platformization:" Paving a Road to Censorship?

Jacob Mchangama, Natalie Alkiviadou

Abstract: To tackle the problems that arise with the horizontalization of content moderation and the resulting ramifications on free speech, this chapter proposes International Human Rights Law (IHRL) as a framework of first reference to re-imagine the current process of moderating contentious speech such as hate speech. Further, it looks at South African jurisprudence which adopts a nuanced and substantiated approach to the free speech – hate speech question, jurisprudence which can serve as an interpretational aide for IHRL provisions. Whilst the chapter recognizes the weakness of marrying IHRL with practices of private companies which are not bound by it, the chapter explains and concludes that IHRL can and should be developed into a workable solution for private companies in the ambit of content moderation of contentious speech.

Keywords: human rights, freedom of expression, hate speech, South Africa, global platforms.

Chapter 1. Introduction

In the 1990s, the Internet was seen as an unstoppable force for the globalization of freedom of expression. As Stanford professor Lawrence Lessig put it: "Nations wake up to find that their telephone lines are tools of free expression, that e-mail carries news of their repression far beyond their borders, that images are no longer the monopoly of state-run television stations but can be transmitted from a simple modem."[1]

1 Lawrence Lessig, "Code: version 2.0", Basic Books, 2006, 236.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Today nearly 60% of the global population – 4,66 billion people - are online and 4,20 billion are active social media users.[2] The transformation of social media platforms into the central agora where ideas are imparted and received has indeed given an unprecedented number of people a voice in local and global affairs. Yet, in tandem with the ability to organize protests, scrutinize the actions of decision makers and make visible marginalized minorities, social media has provided a platform to extremism, terrorist content, disinformation at scale, and hate speech.

But for governments alarmed about the corrosive effects of social media, the centralized amplification of hate, harm, and hoaxes comes with a silver lining. If major platforms like Facebook, YouTube, and Twitter can be forced or persuaded into purging illegal and lawful but awful content, they can become digital chokepoints, with the visibility of illegal content dropping exponentially. Potentially, centralized platforms could even end up serving as the private enforcers of government censorship, entirely inverting the initial promise of egalitarian and unmediated free speech. The most extreme examples of this development can be seen in countries like India, Russia and Turkey[3] where intense pressure is being brought on platforms to remove speech deemed illegal or undesirable by the respective governments. A less draconian – but highly influential - version of this strategy can be seen in, *inter alia*, the pioneering German Network Enforcement Act 2017 (NetzDG) and non-binding measures such as the Christchurch Call for Action.

These initiatives combined with the sheer scale of user generated content have arguably contributed to platforms significantly expanding their efforts to police and purge hate speech. The NetzDG obligates social media platforms with a minimum of 2 million users to remove illegal content –

---

2 Datareportal: "Digital 2021: Global Overview Report", 27 January 2021. https://datareportal.com/reports/digital-2021-global-overview-report#:~:text=Internet%3A%204.66%20billion%20people%20around,now%20stands%20at%2059.5%20percent.

3 96% of the total global volume of demands originated from only five countries (including Russia, Turkey and India) Twitter removal requests. https://transparency.twitter.com/en/reports/removal-requests.html#2020-jan-jun; Karan Deep Singh & Paul Mozur, "As Outbreak Rages, India Orders Critical Social Media Posts to be Taken Down", *New York Times* 25 April 2021. https://www.nytimes.com/2021/04/25/business/india-covid19-twitter-facebook.html; Human Rights Watch,"Russia: Social Media Pressured to Censor Posts: Fines, Smear Campaigns, Potential Blocking for Non-Compliance", 5 February 2021. https://www.hrw.org/news/2021/02/05/russia-social-media-pressured-censor-posts>; Human Rights Watch, "Turkey: Social Media Law will Increase Censorship" 27 July 2020, https://www.hrw.org/news/2020/07/27/turkey-social-media-law-will-increase-censorship.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

including hate speech – within 24 hours, or risk large fines of up to 50 million euros. In the first quarter of 2018 (when the NetzDG had entered into force) Facebook removed 2,5 million pieces of content for violating its Community Standards on hate speech. This rose to 4 million in the first quarter of 2019 and 9,5 million in the first quarter of 2020. By the first quarter of 2021, Facebook purged 25.2 million pieces of 'hate speech' content.[4] This development also reflects that platforms increasingly rely on artificial intelligence to proactively identify and even remove content violating national laws and/or their terms of service. Their rate of content proactively identified by Facebook increased from 38% in the first quarter of 2018 to 96.8% in the first quarter of 2021.[5] While states impose intermediary liability to counter online harms, 'outsourcing' government mandated content regulation to private actors raises serious questions about the consequences on online freedom of expression.

The global nature of social media platforms used by people in almost all countries around the world create significant problems when it comes to determining where to draw the line on various categories of content. In the abstract, large majorities across the globe find it very important that people can speak their mind and use the Internet without censorship. However, once moving from the abstract to specific categories of speech, there are marked variations of tolerance within and between populations of countries as well as between various governments. There is, for instance, no universal agreement on whether statements offensive to minorities should be tolerated. In the Scandinavian countries and the US around 65 % of the populations believe that free speech should extend to statements offensive to minority groups while around 80%. Conversely in Kenya, Indonesia, Turkey and Tunisia, only between 18 and 27% of the populations favor tolerating such statements.[6]

One proposed remedy to bridge the gap between the conflicting attitudes and legal regimes which global social media platforms are forced to navigate is for these private actors to rely on International Human Rights Law (IHRL) when adopting their terms of service and moderating

4 Facebook Transparency Center, "Hate Speech". https://transparency.fb.com/data/community-standards-enforcement/hate-speech/facebook.

5 Facebook Transparency Center, "Proactive Rate". https://transparency.fb.com/policies/community-standards/hate-speech/.

6 Svend-Erik Skaaning & Suthan Krishnarajan, Who Cares about Free Speech? Findings from a Global Survey of Support for Free Speech" *Justitia* (May 2021). https://futurefreespeech.com/wp-content/uploads/2021/06/Report_Who-cares-about-free-speech_21052021.pdf

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

content, even if not formally bound by such legal instruments. Placing content moderation in the framework of IHRL has been discussed by scholars such as Aswad[7] and Benesch,[8] who argue that IHRL, with some modification, can be used by social media companies to moderate online content. Dvoskin takes a different approach, highlighting that adopting IHRL "might not lead to more legitimate content moderation" since this area of law "leaves many speech questions unanswered."[9] It is important to note that there are crucial differences between criminal law and private content moderation. The former involves the threat of criminal sanctions, including – ultimately - the risk of prison, whilst the latter ´merely´ results in the removal of content or, at worst, the deletion of user accounts. Moreover, when restricting freedom of expression, States must follow time consuming criminal procedures and respect legally binding human rights standards. On the other hand, private platforms are generally free to adopt terms of service and content moderation practices less protective of freedom of expression and due process than what follows under IHRL. However, when governments impose intermediary liability on private platforms through laws prescribing punishments for non-removal, platforms are essentially required to assess the legality of user content as national authorities.

In 2018, the(n) UN Special Rapporteur on the Freedom of Opinion and Expression (SRFOE), David Kaye asserted that "human rights law gives companies the tools to articulate their positions in ways that respect democratic norms and counter authoritarian demands".[10]

Given the problems with conflicting legal regimes and popular attitudes towards the limits of free speech, it is tempting to support David Kaye´s assertion that IHRL paves away ahead in the current impasse. After all, IHRL claims to be universal in nature and most states across all continents

---

7 Evelyn Mary Aswad, "The Future of Freedom of Expression Online" *Duke Law & Technology Review* 17, No.1, 2018) , 52-53.

8 Susan Benesch, "But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies" *Yale Journal on Regulation Online Bulletin 39*, No.3, 2020, 90.

9 Brenda Dvoskin, "International Human Rights Law is not Enough to Fix Content Moderation's Legitimacy Crisis", *Berkman Klein Center for Internet & Society at Harvard University*, 16 September 2020. https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd.

10 United Nations, "Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression" A/HRC/38/35. 2018. https://www.undocs.org/A/HRC/38/35.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

have ratified conventions such as the ICCPR (of course ratification does not necessarily entail compliance or genuine commitment). However, it should be acknowledged that there are serious challenges to adopting an IHRL approach to content moderation. First of all, IHRL is binding on states, not on private companies, and while the UN has developed Guiding Principles on Business and Human Rights, these are aspirational and not legally enforceable. Moreover, there are good reasons why social media platforms should be allowed to adopt and experiment with different models of terms of service and content moderation practices dependent on their size, architecture, content, focus etc. Whether content is lawful or not is a complex exercise that is heavily dependent on careful context-specific analysis. Under IHRL, restrictions of freedom of expression must comply with strict requirements of legality, proportionality, necessity and legitimacy. These requirements make the individual assessment of content difficult to reconcile with legally sanctioned obligations to process complaints in a matter of hours or days, not to mention automated content moderation. In a 2021 study Justitia found that the available data showed that on average Council of Europe member states used more than 775 days to process hate speech cases in their national criminal law system from the date of the alleged offending speech till the conclusion of the trial at first instance[11], a time frame wholly incommensurate with how fast platforms are required to remove illegal content under intermediary liability laws such as NetzDG. All these factors mean that a human rights approach to content moderation will necessarily have to be adapted to rather than copied from the current state centric model.

However, this chapter will narrowly focus on how IHRL can contribute to the definition and moderation of the controversial and contested category of "hate speech", which is at the centre of much debate and subject to increasing regulatory scrutiny by both social media platforms themselves as well as numerous states as shown above. This question is all the more relevant given the lack of any authoritative definition of hate speech and widely differing legal standards at both the state and international level. The authors argue that the interplay between ICCPR articles 19 and 20 forms the natural framework for defining and interpreting hate speech under IHRL. In recent years much effort has been spent by both the

11 Jacob Mchangama et al, "Rushing to Judgment: Are Short Mandatory Takedown Limits for Online Hate Speech Compatible with the Freedom of Expression" *Justitia,* January 2021. https://futurefreespeech.com/wp-content/uploads/2021/01/FFS_Rushing-to-Judgment-3.pdf.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Human Rights Committee (HRC), the SRFOE and member states on trying to clarify and strengthen the protection of freedom of expression under article 19, while simultaneously attempting to more clearly and narrowly define the categories of speech that qualify as impermissible hate speech under article 20(2), resulting in a number of soft law instruments as detailed below.

However, given the non-binding nature of these soft law instruments and the paucity of relevant decisions in actual hate speech cases from the HRC, the chapter will do a comparative analysis of two other sources of hate speech jurisprudence, that might be used as an interpretive guide for identifying the obligations under ICCPR articles 19 and 20, when applied in practice. First, the chapter will examine hate speech case law of the European Court of Human Rights (ECtHR) and subsequently relevant hate speech jurisprudence from the South African Constitutional Court and Supreme Court of Appeal. It will be argued that the South African model provides a more convincing, consistent and robust approach to balancing speech protected by freedom of expression against speech which falls afoul of the ban against hate speech as per the dichotomy of ICCPR articles 19 and 20. Conversely it will be argued that the jurisprudence of the ECtHR suffers from serious shortcomings that would add more confusion than clarity and weaken rather than strengthen the protection of freedom of expression if forming the basis of a human rights approach to online content moderation.

Chapter 2. International Human Rights Law: A Framework of First Reference?

1. Pros and Cons to an IHRL approach to Online Content Moderation

There are currently 173 state parties to the ICCPR, making it the most widely accepted convention regulating civil and political rights, including freedom of expression. Accordingly, ICCPR forms the natural focus point of an IHRL approach to content moderation. Article 19 (2) guarantees that "everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice". The fact that article 19 ensures the right to both receive and impart information regardless of frontiers and choice of media, is highly relevant to the Internet and social media, suggesting a positive obligation to facilitate access to information.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Article 19(3) sets out a number of permissible restrictions to freedom of expression as well as procedural and substantive safeguards that must accompany any such restrictions.

Article 19 (3) incorporates a three-part test for limiting freedom of expression. Restrictions must be "provided by law" and are "necessary" for, amongst others, "the respect of the rights or reputations of others" which for hate speech cases is the most relevant of grounds. When it comes to proportionality, the HRC notes that restrictions must be "appropriate to achieve their protective function",[12] and "must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interests to be protected."[13]

In 2011 the HRC published General Comment 34 (GC 34), which constitutes the most authoritative guidance to the obligations under article 19. According to GC 34 the ICCPR protects "even expression that may be regarded as deeply offensive."[14] This seems to entail a heightening of the threshold, which must be met before speech – including hate speech - can be restricted under article 19. For instance, in GC 34 the HRC has held that "Laws that penalize the expression of opinions about historical facts are incompatible with the obligations that the Covenant imposes on States parties in relation to the respect for freedom of opinion and expression. The Covenant does not permit general prohibition of expressions of an erroneous opinion or an incorrect interpretation of past events". [15]

This holding can be contrasted with the HRC´s decision in *Faurisson v France*, in which an academic challenged the use of gas for extermination at Nazi concentration camps. Faurisson was convicted for contesting crimes against humanity, with the HRC finding no violation of the freedom of expression as provided for by article 19. It held that "the restrictions placed on the author did not curb the core of his right to freedom of expression, nor did they in any way affect his freedom of research; they were intimately linked to the value they were meant to protect - the right to be free from incitement to racism or anti-Semitism; protecting that value could not have been achieved in the circumstances by less drastic means."

Accordingly, it would seem that post-GC 34 article 19 now prohibits so-called "memorial laws" criminalizing the denial of historical events such

12 HRC General Comment 34, para. 34.
13 HRC General Comment 34, para. 34.
14 HRC General Comment 34, para. 11.
15 HRC General Comment 34, para. 49.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

as the Holocaust, which as we shall see also marks a decisive difference between the HRC and the ECtHR.

2. *Article 20(2): An Analysis*

Article 20 (2) not only permits restrictions of freedom of expression, but states that "Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law".

As with 4 of the International Convention on the Elimination of All Forms of Racial Discrimination which prohibits, amongst others, the dissemination of racist ideas, 20 differs to other articles in the ICCPR since it imposes a positive obligation on states to prohibit certain types of speech. However, the HRC holds that "articles 19 and 20 are compatible with and complement each other. The acts that are addressed in article 20 are all subject to restriction pursuant to article 19, paragraph 3."[16] In *Ross v Canada*, the HRC underlined that "restrictions on expression which may fall within the scope of article 20 must also be permissible under article 19, paragraph 3."[17]

The 2012 report of the SRFOE underlined that "the threshold of the types of expression that would fall under the provisions of article 20(2) should be high and solid."[18] In 2011, the Office of the United Nations High Commissioner for Human Rights organised a series of expert workshops on incitement to national, racial or religious hatred, as reflected in IHRL.[19] The workshops resulted in the Rabat Plan of Action (RPA) which was launched in 2013.[20] It provides that there must be a high threshold when applying article 20 of the ICCPR.[21] To achieve this, the RPA sets

16 HRC General Comment 34: para. 50.
17 Ross v Canada Communication no 736/1997 (18 October 2000) CCPR/C/70/D/736/1997, para. 10.6.
18 Ross v Canada, para.45
19 International Justice Resource Center, "UN Launches the Rabat Plan of Action", 25 February 201. https://ijrcenter.org/2013/02/25/un-launches-the-rabat-plan-of-action/.
20 Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that constitutes Incitement to Discrimination, Hostility or Violence (launched in 2013) para. 6.
21 Rabat Plan of Action on the Prohibition of Advocacy of National, Racial or Religious Hatred that constitutes Incitement to Discrimination, Hostility or Violence (2002) para. 22.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

out a six-part threshold test to be referred to when applying article 20(2) and includes the assessment of the (i) social and political context (ii) status of the speaker, (iii) intent to incite the audience against a target group (iv) content and form of the speech (iv) extent of its dissemination and (vi) likelihood of harm, including imminence. Since its adoption, the RPA been referred to in several documents, such as in Human Rights Council Resolution 16/18 and the United Nations Strategy and Plan of Action on Hate Speech (2020).[22] The SRFOE has also referenced the RPA extensively including in the 2019 report on online hate speech.[23]

There has been relatively little case law before the HRC on article 20(2) and the degree the six-part test of the RPA has been adopted by the HRC. As such, how this test might apply to real cases cannot be readily discerned. However, in *Mohamed Rabbae, A.B.S and N.A v The Netherlands* from 2016, the HRC gave a relatively extensive overview of article 20(2). Here, the authors claimed to be victims of a violation of their rights under article 20(2) due to allegedly racist statements made by Geert Wilders, leader of the far-right Dutch Freedom Party and his subsequent acquittal by the domestic court. The HRC found that article 20(2) secures the right of persons to be free from hatred and discrimination, but that it is "crafted narrowly" to ensure the protection of freedom of expression. It recalled that this freedom may include "deeply offensive" speech and speech which is disrespectful for a religion, unless the strict threshold of article 20(2) is met.[24] The HRC found no violation of article 20(2) since the Netherlands had developed a suitable legislative framework which victims could reach out to, thereby ensuring that it took the necessary and proportionate measures to prohibit statements made in violation of article 20(2).[25] Relevant to the high threshold attached to article 20(2) is also the concurring individual opinion of Cleveland (Vice Chair of the HRC at the material time) and Politi, in which they noted, amongst others, that "hate speech and similar laws ironically are often employed to suppress the

22 United Nations "Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nations Field Presences", September 2020. https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate%20Speech_Guidance%20on%20Addressing%20in%20field.pdf.

23 Report of the Special Rapporteur on the Freedom of Opinion and Expression, "Online Hate Speech" A/74/486, 9 October 2019. https://www.undocs.org/A/74/486.

24 Mohamed Rabbae, A.B.S and N.A v The Netherlands, Communication no. 2124/2011 (14 July 2016) CCPR/C/117/D/2124/2011, para. 10(4).

25 Mohamed Rabbae, A.B.S and N.A v The Netherlands, para. 10(7).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

very minorities they purportedly are designed to protect."[26] Importantly, their concurring opinion noted the uniqueness in article 20, insofar as it requires the restriction of the "highly protected freedom of expression." This, they argue, means that article 20(2) is "narrowly circumscribed and sets the bar high for the expression that must be prohibited," [27]demonstrating just how narrowly they have construed article 20 to be. Moreover, the finding in favour of the Netherlands is also reflective of this narrow construction.

Despite the lack of binding case law and the paucity of decisions by the HRC, it is submitted that the ICCPR provides a suitable "framework of first reference" for the determination of hate speech by private social media companies, even if not formally bound by this convention.

It would also be a suitable compass for states who are seeking to impose more and more moderation duties at risk of penalties and in short time frames. The post-GC 34 cases on hate speech, the RPA and the guidance and opinions of the SRFOE can guide private companies along the path of adequately protecting the fundamental freedom of expression whilst simultaneously ensuring the safety and dignity of their users. However, the lack of a substantial body of case law applying these principles to specific instances of controversial speech, means that additional sources of hate speech jurisprudence might be needed to help interpretate the relationship between articles 19 and 20.

Chapter 3. The European Court of Human Rights: A Template to Avoid?

No other human rights court has made more decisions in general or on hate speech specifically than the ECtHR. Given that the ECtHR has jurisdiction over 47 member states ranging from Ireland to Azerbaijan and Iceland to Turkey, and that the majority of these states are democracies, it might be tempting to use ECtHR case law on hate speech as a guide to the interpretation of ICCPR article 20(2).

However, there are fundamental differences in the way the ECtHR and the HRC approaches the question of hate speech, and the amount of

---

26 Individual Opinion (concurring) of Committee Members Sarah Cleveland and Mauro Politi in Mohamed Rabbae, A.B.S and N.A v The Netherlands, Communication no. 2124/2011 (14 July 2016) CCPR/C/117/D/2124/2011, para. 8.

27 Individual Opinion (concurring) of Committee Members Sarah Cleveland and Mauro Politi in Mohamed Rabbae, A.B.S and N.A v The Netherlands, Communication no. 2124/2011 (14 July 2016) CCPR/C/117/D/2124/2011, para. 8.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

weight these two bodies attach to freedom of expression in such cases. To commence our discussion, we turn to an analysis conducted by Justitia on a total of 60 identified cases of the former European Commission of Human Rights and the ECtHR, decided upon between 1979-2020.[28] 57 of those cases were brought by speakers and 3 by the targets/victims. *61% of cases brought by the speakers resulted in the applicant's loss* through a finding of non-violation of article 10 (on freedom of opinion and expression): (21%), incompatible *ratione materiae* (9%) and manifestly ill-founded (32%). Only 39% of cases brought by the speakers on the grounds of an article 10 violation have resulted in a finding in favour of the applicant. Thus, on average, free speech restrictions have been upheld in just over one out of three hate speech cases before the ECtHR.

To demonstrate this in a qualitative manner, we will turn to some indicative (by no means exhaustive) case-law, which will cover a range of protected characteristics.

The 2009 case of *Féret v. Belgium* was brought by the leader of a nationalist Belgian party who had been ceased from office for a period of ten years for, amongst others, the preparation and dissemination of publications which included statements of the following sort: "Stop the Islamization of Belgium," "Save our people from the risk posed by Islam, the conqueror." The Court did not succumb to the Belgian government's request for an invocation of article 17 of the European Convention on Human Rights (ECHR) which prohibits the abuse of Convention rights but ruled that there was no violation of article 10. Importantly for the threshold discussion, the Court noted that:

> incitement to hatred did not necessarily call for specific acts of violence or other offences. Insults, ridicule or defamation aimed at specific population groups or incitation to discrimination, as in this case, sufficed for the authorities to give priority to fighting hate speech when confronted by the irresponsible use of freedom of expression which undermined people's dignity, or even their safety.[29]

Therefore, hate speech was deemed to include even insults and ridicules. This line of reasoning was continued in *Vejdeland v Sweden (2012)* which involved the dissemination of homophobic leaflets in school lockers. The Court found no violation of article 10 and reiterated its findings in *Féret*,

---

28 For the full database and quantitative illustrations visit: https://futurefreespeech.com/hate-speech-case-database/.

29 Féret v Belgium, Application no. 15615/07 (ECHR 16 July 2009) para. 73.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

noting that "although these statements did not directly recommend individuals to commit hateful acts, they are serious and prejudicial allegations."[30]

The low threshold was further embedded in the Court's approach in *Lilliendahl v Iceland*, which involved an applicant who wrote comments below an online news article reporting a municipal decision to strengthen education and counselling in schools for pupils identifying as lesbian, gay, bisexual or transgender. The applicant used derogatory comments such as "sexual deviation" when referring to homosexuality and said that this is "disgusting. To indoctrinate children with how sexual deviants copulate in bed." This was the first time that the Court took the question of hate speech and what it means on a more conceptual level (the term yet remains undefined by the Court). It put forth two categories of hate speech, the first being the "gravest forms of hate speech…which fall under Article 17"[31] and the second being the "less grave forms of hate speech" which include "attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population [and which] can be sufficient for allowing the authorities to favour combating prejudicial speech…"[32]

The ECtHR has also put forth conflicting positions when it has come to insult in other cases. For example, *Ibragim Ibragimov and Others v Russia* (2018) was a case which involved the banning of Muslim scholar Said Nrusi's book, as it was extremist. Here, the ECtHR found a violation of article 10, noting that:

> 'merely because a remark may be perceived as offensive or insulting by particular individuals or groups does not mean that it constitutes "hate speech." Whilst such sentiments are understandable, they alone cannot set the limits of freedom of expression. The key issue in the present case is thus whether the statements in question, when read as a whole and in their context, could be seen as promoting violence, hatred or intolerance.'[33]

30 Vejdeland and Others v Sweden, Application No. 1813/07 (ECHR 9 February 2012) para. 54.

31 Lilliendahl v Iceland (2020) Application No.29297/18 (ECHR 12 May 2020) para. 34.

32 Lilliendahl v Iceland (2020) para.36.

33 Ibragim Ibragimov and Others v Russia (Application nos. 1413/08 and 28621/11) para 115 .

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

However, in a case two years later, namely *Atamanchuk v Russia* (2020), the Court took a different approach. Here, the applicant, a journalist/politician was convicted of making statements against non-Russians, referring to them as criminals (without calling for violence). The Court found no violation of article 10, underlining that:

> 'inciting hatred does not necessarily involve an explicit call for an act of violence, or other criminal acts. Attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for the authorities to favour combating xenophobic or otherwise discriminatory speech in the face of freedom of expression exercised in an irresponsible manner.'[34]

Therefore, in the 2018 case, insult was not considered to be sufficient to allow for a restriction to article 10 whereas in the latter it was. Noteworthy is the fact that the 2020 case involved speech directed towards an ethnic group, which the Court appears to have a lower tolerance towards.

Another illustration of the inconsistency in the ECtHR's approach is the manner in which it deals with historical events. For example, the Court systematically finds negationist or revisionist speech in relation to the Holocaust[35] to constitute hate speech, sometimes ousted through the application of the so-called abuse clause in article 17. However, in a case involving the denial of the Armenian genocide,[36] it ruled that this fell within the framework of protected speech.

The treatment of totalitarian symbols is yet another indication of the contradictions found in the Court's approach to alleged hate speech. In *Fáber v Hungary (2012)*,[37] the Court found that article 10 protected an applicant who held a striped Árpád flag[38] less than 100 metres away from a demonstration against racism and hatred. In *Vajnai v Hungary* (2008),[39] during a demonstration, the applicant wore a red communist star and was convicted of the offence of using a *totalitarian symbol which the ECtHR*

---

34 Atamanchuk v Russia, Application no. 4493/11 (ECHR 11 February 2020) para.52.

35 See, inter alia, Williamson v Germany, Application No. 64496/17 (ECHR 8 January 2019), Pastörs *v. Germany, Application No. 55225/14 (ECHR 3 January 2020), Garaudy v France, Application No. 64496/17 (ECHR 7 July 2003).*

36 Perinçek v Switzerland, Application No. 27510/08 (ECHR 15 October 2015).

37 Fáber v Hungary, Application No.40721/08, ECHR 24 October 2012.

38 Used by the Hungarian Fascist Arrow Cross party, responsible for crimes against Jews during World War II.

39 Vajnai v Hungary, Application No.33629/06, ECHR 8 July 2008.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

found to be a violation of the applicant's freedom of expression. However, in the recent case of *Nix v Germany* (2018) – which a German blogger was convicted for using symbols of a banned organization after posting a picture of Heinrich Himmler wearing a swastika armband and likening him to the officers of an employment office which he alleged discriminated against his mixed-race daughter. Despite the fact that the applicant neither advocated nor defended Nazism, the Court found the conviction justified.[40]

In sum, these cases reflect that the ECtHR attaches a low threshold to freedom of expression when it comes to hate speech. This has led to an inconsistent and incoherent case law, with no proper demarcation between freedom of expression and hate speech, resulting in the permissible restriction of speech deemed merely "offensive" or "prejudicial", but with no clear nexus to any harm, speech which included no hateful intent and the selective restriction of the denial of historical events. The ECtHR case law thus fails to satisfy several of the elements of the RPA, and the higher thresholds for restricting hate speech developed by the HRC. Accordingly, using the ECtHR´s case law as a guide to interpreting ICCPR articles 19 and 20 would result in increased confusion, less clarity and a lower degree of protection of freedom of expression.

Chapter 4. South Africa: A Good Practice Template

As noted in the section on the ECtHR, social media companies may look at sources such as Court judgements for inspiration on their content moderation practices. For purposes of providing a well-rounded overview of what is out there in terms of good practices in the ambit of handling hate speech, this chapters offers an overview of key (but not exhaustive) hate speech cases that were heard before the highest courts of South Africa. We choose this country as South Africa has only relatively recently become a liberal democracy after emerging from a long period of white supremacy, which systematically denied both the equality, dignity and the freedom of expression of its non-white population. Accordingly, South Africa is perhaps uniquely suited to act as a "laboratory" when it comes to safeguarding the values of freedom, equality and dignity. Moreover, the South African Constitution is explicitly founded on the values of, inter alia, human rights, and obliges South Africa to "consider international

40 Nix v Germany, Application No. 35285/16, ECHR 13 March 2018 Para. 47.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

law" – including IHRL – when interpreting the constitution´s bill of rights. South African courts frequently rely on international precedents, including the ICCPR, when interpreting the South African constitution´s bill of rights. These factors have, we submit, contributed to South African courts developing a nuanced and substantiated approach to the treatment of hate speech, taking into consideration both the fundamental nature of free speech but also the importance of maintaining dignity and equality.

Section 16 of the South African constitution provides for the freedom of expression. Part 2 therein notes that this freedom does not extend to, *inter alia*, "the advocacy of hatred that is based on race, ethnicity, gender or religion, and that constitutes incitement to cause harm." This provision differs from article 20(2) ICCPR since it is not a positive obligation to prohibit hate speech but, instead, means that hate speech (which meets a certain threshold) is exempt from constitutional protection.

The case *Islamic Unity Convention v Independent Broadcasting Authority and Others* involved statements made on a radio show by a historian who denied the legitimacy of Israel and argued that Jews were not gassed during WWII. The South African Jewish Board of Deputies claimed that the broadcast contravened the Code of Conduct for Broadcasting Services since it was "likely to prejudice relations between sections of the population."

In its judgment, the Court pointed out that freedom of expression:

> …. lies at the heart of a democracy. It is valuable for many reasons, including its instrumental functions as a guarantor of democracy, its implicit recognition and protection of the moral agency of individuals in our society and its facilitation of the search for truth by individuals and society generally. The constitution recognizes that individuals in our society need to be able to hear, form and express opinions and views freely on a wide range of matters….[41]

The Court placed its analysis of expression within a historical context, reiterating the country's recent restrictive past and noting that restrictions would be incompatible with a "constitutionally protected culture of openness and democracy and universal human rights for South Africans of all ages, classes and colours."[42]

---

41 Islamic Unity Convention v Independent Broadcasting Authority and Others, Case CCT36/01 (11 April 2002) para. 26.

42 Islamic Unity Convention v Independent Broadcasting Authority and Others para. 25.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Court further explained the hate speech threshold and the requirement of its real life impact by noting that "not every expression of speech that is likely to prejudice relations between sections of the population would be 'propaganda for war', or 'incitement of imminent violence' or 'advocacy of hatred' which also constitutes 'incitement to cause harm'."[43] This was reiterated in subsequent case-law, such as *Qwelane* discussed below.

The Court ruled that the Code's section prohibiting the impugned speech was broader than what was permissible under the Constitution as it referred to "a section of the population" and not a specific group. It further noted that the reference to "prejudice" did not meet the harm requirement needed for satisfying section 16 of the constitution. Comparatively, two points can be made. Firstly, that, by protecting prejudicial speech, the Court's decision is in line with the high threshold set out by article 20(2) of the ICCPR and further assessed by the RPA as well as HRC case law (see for example *Rabbae* and the extension of the freedom of expression to 'deeply offensive' speech). In addition, the test developed by the Constitutional Court in the Islamic Unity Convention case is more speech protective than the ECtHR which has permitted the restriction of prejudicial speech (see, for example, *Vejdeland*). Moreover, the decision sides with GC 34 over the case law of the ECtHR when it comes to the controversial question of whether to protect even the denial of historical crimes such as the Holocaust.

In relation to incitement, the Constitutional Court recently held that a law criminalizing incitement to "any offence" was "unquestionably overbroad and its inhibition of free expression is markedly disproportionate to its conceivable benefit to society."[44] The case revolved around statements made by the president of the political party "Economic Freedom Fighters," who called his supporters to illegally occupy land. In his majority decision Chief Justice Mogoeng Mogoeng noted that freedom of expression is the 'lifeblood of constitutional democracy' and that '[w]hen citizens are very angry or frustrated, it serves as the virtual exhaust pipe through which even the most venomous of toxicities within may be let out to help them calm down, heal, focus and move on.'

---

43 Islamic Unity Convention v Independent Broadcasting Authority and Others para. 34.

44 Economic Freedom Fighters, Julios Selo Malema v Minister of Justice and Correctional Services, National Director of Public Prosecutions, Case CCT 201/19, Para. 61.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Court's position was, once again, informed by the country's apartheid history. The judgement referred to the fact that the right to freedom of expression was violated during the "highly intolerant and suppressive past"[45] and, it "thus has to be treasured, celebrated, promoted and even restrained with a deeper sense of purpose and appreciation of what it represents." Although the Court also emphasized that freedom of expression is not absolute, nor more important than other rights, it stressed that limitations can only occur in specific circumstances, such as when national interest, dignity, physical integrity or democracy is threatened. The Court noted that this complied with the country's international obligations in respect of limitations to free expression making a specific reference to article 19 ICCPR.[46]The Supreme Court´s view of free speech as a vital democratic exhaust pipe and the country´s history of white supremacy as a caution *against* censorship, marks a stark difference to the ECtHR. The Strasbourg court tends to stress the (supposed) capability of controversial speech to cause harm and danger – even absent any direct incitement to harm - and sees European history as offering a compelling argument *in favour* of restricting extreme speech.

The Supreme Court of Appeal (SCA) has also developed a high threshold in relation to the restriction of hate speech, as witnessed in the case of *Qwelane v South African Human Rights Commission*. This case involved a 2008 publication by Jon Qwelane, a well-known anti-apartheid activist and journalist in the Sunday Sun. The article was titled "Call me names but gay is not okay…" and used homophobic language and was accompanied by a cartoon comparing homosexuality to bestiality. The article stated, *inter alia*, that:

> The real problem, as I see it, is the rapid degradation of values and traditions by the so-called liberal influences of nowadays; you regularly see men kissing other men in public, walking holding hands and shamelessly flaunting what are misleadingly termed their 'lifestyle' and 'sexual preferences…. At this rate how soon before some idiot demands to 'marry' an animal and argues that this constitution 'allows it'?

45 Economic Freedom Fighters, Para. 2.
46 It does so in footnote 51.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

In 2017, the Johannesburg High Court decided that certain statements were "hurtful, harmful, incite[d] harm and propagate[d] hatred"[47] thereby violating Section 10(1) of the Equality Act. Qwelane appealed the case to the SCA on the grounds that the Equality Act's definition of hate speech was unconstitutional since it prohibited more speech than provided for in section 16(2) of the Constitution. The SCA referred to the freedom of expression as the "lifeblood of a democratic society." It noted that section 10 of the Equality Act did, in fact, go beyond what was constitutionally permissible under section 16(2) and warned that "one must be careful not to stifle the views of those who speak out of genuine conviction."[48] It placed its assessment within a historical framework, holding that "given our history...freedom of expression must also be prized."[49] As such, it found section 10 of the Equality Act to be unconstitutional and gave Parliament 18 months (as of November 2019) to remedy the current content of the said section. The high threshold adopted in *Qwelane* by the SCA particularly was based on two cases which were heard together in 2018, namely *Moyo v Minister of Justice and Constitutional Development* and *Sonti v Minister of Justice and Correctional Services and Others* (2017). The SCA noted that, for restrictions to speech to be legitimate, there must be a nexus between the speech and actual harm (not merely perceived harm) and, as such, "no one is entitled to be insulated from opinions and ideas that they do not like even if those ideas are expressed in ways that place them in fear…."[50]

The SCA maintained the high threshold to hate speech after *Qwelane.* In December 2018, it ruled on *Masuku and Another v South African Human Rights Commission obo South African Jewish Board of Deputies* (2018). The cases involved statements made by Masuku, the secretary of the International Relations arm of the Congress of South African Trade Unions. In the framework of the Israel-Palestine conflict, Masuku made statements such as:

> "Let us bombard the COSATU offices with phone calls to let them know our anger. It is hard[er] to ignore phone calls than email.

47 Qwelane v South African Human Rights Commission and Another, Case 686/2018, 29 November 2019, para. 10.
48 Qwelane, para. 70.
49 Qwelane para. 84.
50 Moyo v Minister of Justice and Constitutional Development and Others; Sonti v Minister of Justice and Correctional Services and Others, Cases 287/2017; 286/2017, para. 31.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

> Maybe we should start a policy that Israel-loyal Jews refuse to employ COSATU members in retaliation to COSATU's evil actions."

Again, the Court highlighted that speech may be "hurtful of people's feelings or wounding, distasteful, politically inflammatory or downright offensive [but this] does not exclude it from protection."[51]

The above approach to the free speech – hate speech debate marks a stark contrast to the ECtHR´s position on homophobic speech as set out in *Vejdeland and Others v Sweden*, where merely prejudicial allegations were sufficient to constitute hate speech that a State could prohibit without violating article 10. This position was also adopted in a 2020 ECtHR case, *Lilliendahl v Iceland*, which involved homophobic and transphobic speech. Here, the Court reiterated its position in *Vejdeland*, nothing that speech which is "prejudicial" can also constitute hate speech.[52] As such, the Court "[saw] no reason to disagree with the Supreme Court's assessment that the applicant's comments were 'serious, severely hurtful and prejudicial.'"[53]

It must be noted that the South African Human Rights Commission appealed the SCA's judgement at the Constitutional Court. In July 2021, the Constitutional Court[54] found that only the inclusion of the term "hurtful" was unconstitutional, but that the elements of hate and harm were constitutional. As such, it ruled that Qwelane's article constituted hate speech in line with the other elements of Section 10(1) of the Equality Act (hateful and harmful speech). Nevertheless, it did underline that "hate speech travels beyond mere offensive expression and can be understood as extreme detestation and vilification which risks provoking discriminatory activities against that group", accordingly while the Constitutional Court appears to have modified the very speech protective direction of South African courts vis-a-vis the prohibition of hate speech, the current threshold is still significantly more speech protective than what follows under the EctHR, and arguably more in line with what follow under ICCPR Articles 19 and 20(2).

---

51 Masuku and Another v South African Human Rights Commission obo South African Jewish Board of Deputies, Case 1062/2017, 4 December 2018 , para. 31.

52 Lilliendahl v Iceland, 2020, para. 36.

53 Lilliendahl, para. 39.

54 Qwelane v South African Human Rights Commission and Another (CCT 13/20) [2021] ZACC 22 (31 July 2021

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Conclusion

The author argues that the South African approach, which emanates from recent experience with systemic speech repression is nuanced and substantiated, providing a more rigorous and convincing balancing between freedom of expression and hate speech. The highest courts of South Africa have found that speech which is merely "offensive" or "prejudicial" is protected, whereas such speech, has often fallen afoul of the ECtHR. The same is true of statements denying historical crimes such as the Holocaust. Whilst the protection of expression was impacted by the final decision in this case, the authors argue that South Africa continues to constitute a good example of a substantiated approach to hate speech. This jurisdiction marks a significant contrast to the ECtHR´s approach to hate speech, which is steeped in the doctrine of "militant democracy" according to which statements that allegedly undermine (essentially undefined) democratic values are undeserving of protection.

On global social media platforms with users from all continents and cultures with widely diverging and clashing conceptions of where free speech ends and hate speech begins, a robust, narrow and harm-based definition of hate speech is more likely to be operational than one which includes deeply subjective notions of "offense" and "prejudice". Accordingly, stakeholders such as private companies, states and international organizations could look at the judgments of the highest courts of South Africa as a guide to re-considering current approaches to the treatment of online hate speech. Moreover, this case-law is an effective ambit through which stakeholders can align content moderation requirements with the thresholds set out by IHRL and particularly article 20(2) with the HRC deciding cases such as *Rabbae* and holding that speech extends even to 'deeply offensive' speech. In brief, South African jurisprudence (from its highest courts) provides for a substantiated approach to hate speech, preventing over-restriction (for example allowing prejudicial speech) whilst placing analysis in the realm of real-life experience (its own apartheid).

The current digital era is marked by increasing pressure on social media platforms to quickly remove content such as "hate speech." The obligation to remove such a contested and poorly defined area of speech within short time spans on global platforms is ill suited to offer the necessary safeguards for freedom of expression. This approach is more likely to initiate a global censorship race to the bottom, than act as a bulwark of liberty, and indeed such a development already seems to be well under way. While no quick fix is likely to resolve this situation, IHRL, provides the best, or least bad, "framework of first reference" for both states and major platforms when

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

it comes to determining the relationship between protected expression and impermissible hate speech. In particular, articles 19 and 20 of the ICCPR offers a promising way ahead, which offers a more robust and speech protective way forward than the incoherent case law of the ECtHR. Yet, given the paucity of legally binding cases relating to the ICCPR, South African case law on the relationship between freedom of expression and hate speech offers a compelling interpretational aide when further defining the relationship between article 19 and 20.

Bibliography

Aswad, Evelyn Mary. "The Future of Freedom of Expression Online." *Duke Law & Technology Review* 17, no.1 (2018): 26-70.

Benesch, Susan. "But Facebook's Not a Country: How to Interpret Human Rights Law for Social Media Companies." *Yale Journal on Regulation Online Bulletin 39*, no.3 (2020): 86-111.

Dvoskin, Brenda. "International Human Rights Law is not Enough to Fix Content Moderation's Legitimacy Crisis." *Berkman Klein Center for Internet & Society at Harvard University*, September 16, 2020. https://medium.com/berkman-klein-center/international-human-rights-law-is-not-enough-to-fix-content-moderations-legitimacy-crisis-a80e3ed9abbd.

Lessig, Lawrence. *Code: version 2.0.* New York: Basic Books, 2006.

Mchangama, Jacob et al. "Rushing to Judgment: Are Short Mandatory Takedown Limits for Online Hate Speech Compatible with the Freedom of Expression." *Justitia*, January 2021.

Singh, Karan Deep and Mozur, Paul. "As Outbreak Rages, India Orders Critical Social Media Posts to be Taken Down." *New York Times*, April 25, 2021. https://www.nytimes.com/2021/04/25/business/india-covid19-twitter-facebook.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Online Shaming - a New Challenge for Criminal Justice

Kristiina Koivukari, Päivi Korpisaari

Abstract: Online shaming is a harmful phenomenon that violates the psychological and sometimes even physical wellbeing of the target (or victim) of the action. Shaming and other forms of online hate speech also affect the use of freedom of expression in society by reducing the amount – or at least the range – of opinions expressed and information available in public. The aim of this article is to discuss whether initiating a shaming action or participating in it could or even should be criminalized. As a conclusion, we argue that an offence comprehensively covering acts of online shaming would be difficult or even impossible to formulate without violating the requirements of freedom of expression and certain fundamental principles of criminal law.

Keywords: human rights, freedom of expression, right to private life, social media, shaming, criminalizing, principles of criminal law

1. Freedom of expression and social media

Freedom of expression is an essential value in every democratic society and a fundamental right in European law. It is guaranteed in the United Nations' Universal Declaration of Human Rights and in the European Convention on Human Rights. It includes the freedom to hold opinions as well as to receive and impart information and ideas. According to the European Court of Human Rights (ECtHR) freedom of expression constitutes 'one of the essential foundations of a democratic society and one of the basic conditions for its progress'.[1] Freedom of expression is also

1 For example, *ECtHR, Barthold v. Germany*, App. no. 8734/79, 25 March 1985; *ECtHR, Handyside v. the United Kingdom*, App. no. 5493/72, 7 December 1976; *ECtHR, Zana v. Turkey*, App. no. 18954/91, 25 November 1997 (GC); *ECtHR, Von Hannover v. Germany (No. 2)*, App. nos. 40660/08 and 60641/08, 7 February 2012 (GC); *ECtHR, Axel Springer AG v. Germany*, App. no. 39954/08, 7 February 2012 (GC); *ECtHR, Gillberg v. Sweden*, App. no. 41723/06, 3 April 2012.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a social good, promoting truth, democracy and participation. It can also be regarded as an end in itself, promoting individual self-fulfilment and as an individual good.[2] As well as protecting positive and insignificant (or otherwise neutral) expressions, freedom of expression also applies to expressions that offend, shock or disturb.[3]

Another important right is the right to private life. This covers the physical, psychological and moral integrity of a person, as well as the right to establish and develop relationships with other human beings. The right to private life gives protection against public dissemination of a person's private information or photos in situations where individuals can legitimately expect that that kind of information is not published without their prior consent.[4] The right to private life guarantees dignity and autonomy since revealing private matters without consent takes away an individual's control and can deprive them of their dignity or reputation in the eyes of

2 See for more detail Jan Oster, *Media Freedom as a Fundamental Right* (Cambridge: Cambridge University Press, 2015), 13-20. Already more than fifty years ago Thomas I. Emerson, "Toward a General Theory of the First Amendment," *The Yale Law Journal* 72, no. 5 (1963): 878-879 considered freedom of expression important for four reasons: '(1) as assuring individual self-fulfilment, (2) as a means of attaining the truth, (3) as a method of securing participation by the members of the society in social, including political, decision-making, and (4) as maintaining the balance between stability and change in the society.'

3 For example, *ECtHR, Hertel v. Switzerland*, App. no. 25181/94, 25 August 1998, § 46; *ECtHR, Stoll v. Switzerland*, App. no. 69698/01, 10 December 2007 (GC), § 101; *ECtHR, Steel and Morris v. the United Kingdom*, App. no. 68416/01, 15 February 2005, § 87; *ECtHR, Mouvement raëlien suisse v. Switzerland*, App. no. 16354/06, 13 July 2012 (GC), § 48; *ECtHR, Handyside v. the United Kingdom*, App. no. 5493/72, 7 December 1976, § 49; *ECtHR, Observer and Guardian v. the United Kingdom*, App. no. 13585/88, 26 November 1991, § 59.

4 Regarding photos, *ECtHR, Von Hannover v. Germany (No. 2)*, App. nos. 40660/08 and 60641/08, 7 February 2012, (GC), § 96. See also *ECtHR, Lillo-Stenberg and Sæther v. Norway*, App. no. 13258/09, 16 January 2014, § 26. Regarding other kinds of private information see *ECtHR, von Hannover v. Germany*, App. no. 59320/00, 24 June 2004, §§ 50-53; *ECtHR, Sciacca v. Italy*, App. no. 50774/99, 11 January 2005, § 29; *ECtHR, Flinkkilä and others v. Finland*, App. no. 25576/04, 6 April 2010, § 75; *ECtHR, Saaristo and others v. Finland*, App. no. 184/06, 12 October 2010, § 61; *ECtHR, Von Hannover v. Germany (No. 2)*, App. nos. 40660/08 and 60641/08, 7 February 2012, (GC), § 95. See also *ECtHR, Petrina v. Romania*, App. no. 78060/01, 14 October 2008, § 27, and *ECtHR, Rothe v. Austria*, App. no. 6490/07, 4 December 2012.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

others.[5] The right to private life may also cover protecting reputation[6] and honour.[7]

The Internet makes it possible to express oneself without the restrictions imposed by traditional media. As the ECtHR stated in *Delfi*, "user-generated expressive activity on the Internet provides an unprecedented platform for the exercise of freedom of expression".[8] Possibilities to express oneself anonymously encourage free speech, expression of various ideas and revealing grievances and abuses. In addition to the right to expression, the Internet and search engines also play a major role in obtaining information and ideas.

As a space for open communication, social media and the Internet enable formation of online cultures (and countercultures) where individuals express their ideas and opinions quickly and world-wide to large groups of people. Unfortunately, the potential created by social media is not always used for the common good, with the result that social networking sites have become platforms for both information and disinformation. In addition, the speed of communication in social media and the ability to express oneself anonymously has increased the number of obscene insults towards individuals and ethnic, religious or other groups of people. As the ECtHR put it in *Delfi*, "Defamatory and other types of clearly unlawful speech, including hate speech and speech inciting violence, can be disseminated like never before, worldwide, in a matter of seconds, and sometimes remain persistently available online."[9]

---

5 Päivi Korpisaari, "Balancing freedom of expression and the right of private life in the European Court of Human Rights - application and interpretation of the key criteria," *Communications Law* 22, no. 2 (2017): 39.

6 *ECtHR, Chauvy and Others v. France*, App. no. 64915/01, 29 June 2004, § 70; *ECtHR, Abeberry v. France*, App. no. 58729/00, 21 September 2004 (dec.); *ECtHR, Leempoel & S.A. ED. Ciné Revue v. Belgium*, App. no. 64772/01, 9 November 2006, § 67; *ECtHR, White v. Sweden*, App. no. 42435/02, 19 September 2006, § 26; *ECtHR, Pfeifer v. Austria*, App. no. 10802/84, 25 February 1992, § 35; *ECtHR, Fürst-Pfeifer v. Austria*, App. nos. 33677/10 and 52340/10, 17 May 2016, § 35.

7 *ECtHR, Radio France and others v. France*, App. no. 53984/00, 30 March 2004; *ECtHR, Cumpănă and Mazăre v. Romania*, App. no. 33348/96, 17 December 2004 (GC), *ECtHR, Sanchez Cardenas v. Norway*, App. no. 12148/03, 4 October 2007; *ECtHR, A v. Norway*, App. no. 28070/06, 9 April 2009, § 64.

8 *ECtHR, Delfi AS v. Estonia*, App. no. 64569/09, 16 June 2015, § 110.

9 *ECtHR, Delfi AS v. Estonia*, App. no. 64569/09, 16 June 2015, § 110.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

2. *Shaming as harmful action online*

One form of harmful use of social media is online shaming, which has been a topical issue over the last few years.[10] There is no universal or commonly accepted definition of shaming, and indeed, the use of the term varies slightly. The kind of online shaming treated in this paper could also be discussed as cyber-bullying or -harassment.[11]

We understand the concept of online shaming as a certain kind of *organised* shaming or harassing action online. Shaming consists of *numerous* harassing expressions and is understood as an action in which someone *intentionally initiates* online vilification or a hate campaign against another person, usually on social media or another online platform. Online shaming refers to a systematic activity aiming at silencing people or harassing them, for example, by threatening them or disseminating their private (or untrue) information on the Internet. In the long run, only a fear of online shaming can affect the willingness of some people to participate in public discussions, which further restricts the range of topics that are discussed in public and the ways in which certain (heated or delicate) issues are discussed.

Separate acts of online shaming resemble or might constitute a criminal offence (depending also on what acts are criminalized and how in the state in question). However, acts of shaming differ from offences such as defamation or dissemination of information violating personal privacy in that the action of shaming and the harm it causes as a whole is always a sum of several, even tens or hundreds of separate acts.[12] Moreover, the complex of shaming differs from the offence of, for instance, stalking, by always being committed by several people. On the other hand, as hate speech is usually understood as expressions of hate against a person or a group based on group characteristics (such as race, sex, sexual orientation,

---

10 Online hate speech has an impact on victims' wellbeing, social trust, self-image and social relations. See Teo Keipi et al., *Online Hate and Harmful Content: Cross-National Perspectives* (London: Routledge, 2017).

11 The concepts of cyber-bullying and -harassment are used perhaps more often specifically in the context of studying the behaviour of youth and adolescents in the online environment as well as sexual or gender based harassment online. See e.g. Peter Coe, "The Social Media Paradox: An Intersection with Freedom of Expression and the Criminal Law," *Information & Communications Technology Law* 24, no. 1 (2015): 27-29.

12 Extensively on shaming and the different ways shaming can violate the right to privacy, Emily B. Laidlaw, "Online Shaming and the Right to Privacy," *Laws* 6, no. 1 (2017).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

religion, and so on), in shaming, the trigger for expressions of hate or disrespect can be basically anything, for instance, the target's opinions or ideas that they have shared, their work or position of trust.

In practice, a campaign can happen or be initiated in numerous different ways and in varying environments, and for many different reasons. One censorious, derogatory or mocking remark made by someone who reaches a wide audience, for instance on social media, can generate a flood of hateful and disrespectful comments by other people against the individual chosen as the target. The comments can be anything from mocking to threatening or they can include, for instance, dissemination of information that violates personal privacy. In addition, acts of shaming can also lead to different types of harassment beyond the online environment: "physical threat", malicious accusations, groundless complaints, stalking, calls to the family members or employer of the target, and the like.[13] Despite having some effects outside the online environment, social media is essential for shaming actions in providing a platform to initiate and perform such campaigns.

The kind of shaming described above is close to or partly overlaps with *doxing*, *trolling*, *virtual mobbing* or *flaming*, and even mere *gossiping*. In practice, these phenomena might be difficult to differentiate from each other, as it is impossible to define online shaming exhaustively. One shaming action can include a mixture of different types of harassment and different conduct can have different motives. In addition to the confusing terminology, there are some differences as to approaching the phenomenon; hence the questions that follow.[14] In the type of shaming discussed here,

---

13 See also Guy Aitchison and Saladin Meckled-Garcia, "Against Online Public Shaming: Ethical Problems with Mass Social Media," *Social Theory and Practice* 47, no. 1 (2021): 7.

14 For instance, online public shaming can be approached as a means of social control and moral condemnation or even an informal reputational punishment to be used when someone has (allegedly) transgressed moral norms. Seen as such, the phenomenon does not necessarily have to be considered solely and in every case harmful and wrong, but in some cases also a desirable way of collectively expressing opinions, moral commitments and condemning the morally reprehensible. Discussion from this perspective, Behnam Taebi and Azar Safari, "On Effectiveness and Legitimacy of 'Shaming' as a Strategy for Combatting Climate Change," *Science & Engineering Ethics* 23 no. 5 (2017); Paul Billingham and Tom Parr, "Online Public Shaming: Virtues and Vices," *Journal of Social Philosophy* 51, no. 3 (2020). As for recent Finnish discussion on shaming, the phenomenon is seen only as undesirable harassment, and it has been discussed mainly as an occupational issue. In other words, shaming is understood as a way to harass and silence e.g. journalists, state employees, and researchers, without any other reason

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

someone actively and intentionally initiates the process in which others participate by (mostly publicly, but not necessarily) commenting on the target, sending messages, and reposting others' comments and "liking" them. The aim of the process might be, for instance, to shame the target, express disapproval of them or silence them,[15] but once the process has started, it easily gets out of hand, beyond the control of the initiator or anyone participating in the shaming action. Therefore, the effects that online shaming may have on the target or public discussion in general do not necessarily equate with the original intentions of the initiator, let alone other participants.

Since the actions of online shaming are severely disturbing to the target themselves, even a risk of winding up as a target might affect the willingness of some people to participate in public discussions in general or in certain subject areas, or it might affect the way they are willing to discuss anything in public.[16] In the most alarming cases the threat of shaming affects the issues that, for instance, journalists are willing to bring up or researchers are willing to study. Thus, along with causing serious mental pain or even psychological illness to the target, online shaming or a risk of it potentially affects and distorts public discussion and the public's right to obtain information. So, from the victim's perspective, and in order to enable and encourage public discussion even on delicate or controversial issues, something definitely should be done.

There are of course some variations in the criminal codes of different states as to which offences might apply to shaming and under what conditions. However, in many cases the range of potentially applicable offences is rather wide, and in addition, the rules on complicity might apply too. Similarly, because of the complicated nature of shaming and the rather scattered criminal law provisions applicable, in practice, cases of shaming are usually difficult to get hold of and investigate or prosecute. Another

---

or motivation for shaming action than the target's having a different opinion to that of the perpetrator. Seen from this perspective, shaming must be wrong and reprehensible. At the same time, however, this perspective neglects the problem of how to differentiate harmful shaming from justifiable criticism of the target.

15 Aitchison and Meckled-Garcia emphasise the importance of noticing that in online public shaming people are harassed "because of a characterisation of who or what they are", not what they say or express. Aitchison and Meckled-Garcia, "Against Online Public Shaming," 6.

16 On this kind of chilling effect and the reasons behind it, David Bromwich and George Kateb, eds., *On Liberty: Rethinking the Western Tradition* (New Haven: Yale University Press, 2003), 76. See also e.g. Aitchison and Meckled-Garcia, "Against Online Public Shaming," 6.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

problem is that even if some acts participating in online shaming constitute offences, the crimes committed do not appear very serious, seen separately and not as a whole. This means that the police might not consider the crimes worth investigating, also taking into account that crimes taking place online and anonymously are difficult to investigate,[17] and no effective investigative methods are available for minor offences. As Keipi and others point out, "angry and hateful online users may easily disturb the online activity of dozens or even hundreds of other users without having to face any consequences for their actions".[18] Moreover, the ECtHR has interpreted the right to freedom of expression rather permissively,[19] and exercising caution in applying restrictions on the right to free speech can be justified on the basis that broader restrictions might lead to "a floodgate of trivial cases" and create a chilling effect.[20] As the current criminal law does not function very well in protecting targets of online shaming, would it be possible and reasonable to criminalize shaming actions in a distinct criminal law provision designed to tackle the problems described above?

3. *Criminalizing online shaming?*

The aim of this paper is to discuss questions relating to shaming as common European issues rather than including any comparison of different approaches to shaming or any online harassment and hate speech in different European countries. This causes certain shortcomings in terms of the

17 On the problems that anonymity causes in the context of crimes committed online, see e.g. Kathryn Chick, "Harmful Comments on Social Media," *York Law Review* 1, (2020): 102-104.

18 Keipi et al., *Online Hate and Harmful Content,* 71.

19 See *ECtHR, Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v. Hungary,* App. No. 22947/13, 2 February 2016, where the ECtHR stated that the comments were vulgar but not clearly illegal. One of the comments was "People like this should go and shit a hedgehog and spend all their money on their mothers' tombs until they drop dead." However, compare to *Delfi*, where the comments were mainly hate speech and speech inciting violence towards the director of the ferry company.

20 Chick, "Harmful Comments on Social Media," 88 and the CPS Guidelines referred to by her, CPS/Director of Public Prosecutions, "Guidelines on Prosecuting Cases Involving Communications Sent Via Social Media," June 20, 2013, accessed April 4 2021, http://data.parliament.uk/DepositedPapers/Files/DEP2013-1025/social_media_guidelines.pdf. On difficulties of having internet and social media crimes prosecuted, also e.g. Coe, "The Social Media Paradox".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

accuracy of our claims: the ideas and claims on criminal law presented in this paper remain at a general level and might not apply to any jurisdiction as such. As Fletcher puts it: 'the languages of criminal law, with their rich moral overtones, are deeply embedded in particularistic cultures of guilt and blaming. There is no serious possibility of developing a value-free, quasi-scientific language of criminal law that could claim universal understanding.'[21] However, certain fundamental principles and values are shared by all European criminal justice systems.[22] Hence, discussing the question of criminalizing shaming at a general level is reasonable, since the essential issues of potential criminalization indeed conflict with the central values of European criminal justice.

The possibility or even need to criminalize shaming actions is examined through restrictions that the principles of legality and individual autonomy along with the requirements of freedom of expression, provide when criminalizing conduct initiating shaming or participating in shaming. The starting point must be freedom of expression as protected by Article 10 of the ECHR, and the fact that according to the ECtHR, this protects expressions broadly, including even expressions "that offend, shock or disturb the State or any section of the population."[23] Although restricting freedom of expression is possible, restrictions are limited *inter alia* to those necessary in a democratic society. As for the idea or principle of legality, this is a fundamental part of any democratic *Rechtstaat*, and is written, for example, into the European Convention on Human Rights (Article 7) as well as national constitutions and criminal laws.[24] Some of the essential requirements of the principle of legality are the requirements of written law and maximum certainty: "the condition [of written law] is met in the case where the individual concerned is in a position, on the basis of the wording of the relevant provision and with the help of the interpretative

21 George P. Fletcher, *The Grammar of Criminal Law: American, Comparative, and International. Volume One: Foundations* (Oxford: Oxford University Press, 2007), 118.

22 E.g. Alan Norrie, *Crime, Reason and History, A Critical Introduction to Criminal Law*, 2nd edn. (London: Butterworths, 2001); Kristiina Koivukari, "The crumbling narrative of modern European criminal justice" (Dissertation, University of Helsinki, 2020).

23 E.g. *Handyside v. the United Kingdom*, App. no. 5493/72, 7 December 1976, § 49.

24 See also Alexandros Kargopoulos, "Fundamental rights, national identity and EU criminal law," in *Research Handbook On EU Criminal Law,* eds. Valsamis Mitsilegas, Maria Bergström and Theodore Konstadinides (Cheltenham, UK: Edward Elgar Publishing, 2016), 126-128; Christina Peristeridou, *The Principle of Legality in European Criminal Law* (Cambridge: Intersentia, 2015).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

assistance given by the courts, to know which acts or omissions will make him criminally liable".[25] Moreover, the principle of individual autonomy presumes that "each individual should be treated as responsible for his or her own behaviour", which in turn must be respected in criminalizing any conduct.[26] This means, in practice, that individuals should not be punished for accidents or when they have not "recognised the harmful aspect of their conduct or its consequences".[27]

3.1 Conduct initiating shaming action

As the principles of individual autonomy and legality *inter alia* suggest, not any conduct in any manner can be made criminal. As a starting point, the Latin terms of *actus reus* and *mens rea* established in English criminal law well illustrate the questions addressed here. To put it simply, *actus reus* refers to the guilty act whereas *mens rea* refers to the guilty mind. In order to punish someone, the offender must have committed an offence defined as a criminal act in the criminal code, and must have done so intentionally (or possibly recklessly or negligently).[28] Further, as the principle of maxi-

---

25 *ECtHR, Kokkinakis v. Greece*, App. no. 14307/88, 25 May 1993, § 52. See also Case C-303/05 *Advocaten voor de Wereld VZW v. Leden van de Ministerraad* EU:C:2007:261, § 50, referring to ECtHR judgement in *Coëme and Others v. Belgium,* App. Nos. 32492/96, 32547/96, 32548/96, 33209/96 and 33210/96, 22 June 2000. Similarly in Case C-308/06 *The Queen, on the application of International Association of Independent Tanker Owners (Intertanko) and Others v. Secretary of State for Transport* EU:C:2008:312, § 71 and in Case C-42/17 *Criminal proceedings against M.A.S. and M.B*. EU:C:2017:936.

26 Andrew Ashworth, *Principles of criminal law,* 6th edn. (Oxford: Oxford University Press, 2009), 23.

27 Norrie, *Crime, Reason and History*, 35-36.

28 E.g. Ashworth, *Principles of criminal law*. An in-depth and critical analysis of these concepts, see Norrie, *Crime, Reason and History*. In this paper, the discussion on the possibility of criminalising shaming is limited to considering it as an intentional crime for the sake of clarity. It is, however, worth noting that intention as an element of the offence of shaming might not be the first choice in all jurisdictions and regarding all participants. Moreover, there are of course major differences in the ways intention or liability in general is understood in different jurisdictions. E.g. Jeroen Blomsma, "Mens Rea and Defences in European Criminal Law" (Cambridge: Intersentia, 2012). Furthermore, discussing intention (e.g. different notions on degrees of intention) or questions of liability in detail are beyond the scope and reach of this paper; hence, where these issues are touched on, the analysis is rather cursory.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

mum certainty provides that an offence must be clearly defined in law, people should be given fair warning, and it should not be too difficult to draw a line between acts that constitute a punishable offence and those that do not. These requirements must also be reflected when deciding whether and how to criminalize different acts. So, from the perspective of criminalizing online shaming, we should first be able to clearly define acts that are punishable, but this should be done so that it is possible to evaluate later in every case whether (and prove that) the offender acted intentionally.

The starting point in criminalizing online shaming must be to define conduct that initiates a hate campaign. Online shaming should be defined as direct or indirect incitement of other people to somehow disturb the target. However, as incitement could also be indirect, in practice, it would be difficult to know the actual intentions and motivations of the initiator. A critical or mocking remark made with no purpose of initiating a shaming action would easily look like the offence of shaming if other people were nevertheless provoked to post disturbing comments concerning the person criticised. Yet, according to the principle of individual autonomy, no one should be punished purely based on the consequences of an otherwise legitimate act. Moreover, the distress experienced by the target cannot form a benchmark for criminal activity, since justified and legitimate criticism may also cause different kinds of negative feelings in the one criticized. It is almost impossible to objectively evaluate the state of mind of the potential offender and to prove that they committed the crime intentionally and of their own free will, at least if at the same time we do not want to criminalize most critical remarks referring to someone personally and reaching a large audience.[29]

As we cannot know the actual motivation of the initiator, the only possibility is to try to define the circumstances indicating that the provocation and harm caused was intentional (assuming intention would be one of the essential elements of the offence). Let us assume that one criterion would be the size of the audience in the media or the site where the criticism was published, and another would be the previous activity of the initiator.

29 Describing actions somewhat similar to shaming as informational and reputational cascades, Cass Sunstein illustrates well the dynamics of a rumour spreading in social media. In his examples, it is obvious that we cannot know the intentions and motivations of the participants as the participants might not even recognise the reasons for their actions themselves. Cass Sunstein, "Believing False Rumors," in *The Offensive Internet: Speech, Privacy, and Reputation*, eds. Saul Levmore and Martha C. Nussbaum (Cambridge, Mass: Harvard University Press, 2010), 92-96.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

So, for instance, if a journalist or a researcher criticised someone's opinion on a politically sensitive or heated issue, and this provoked disturbing reactions towards the subject of criticism, the initial critical comment could be evaluated as the offence of initiating a shaming action. If we also suppose that the critical remark was made on social media, it was not the first time the journalist or the researcher in question criticised the same person, and they had many followers, the circumstances could surely be interpreted as intentional shaming even if the person in question only intended to criticise (and not to shame) someone or something. Hence, it would be difficult to distinguish (in a clearly defined and objective legal norm) between illegal shaming actions and justified criticism drawing attention to worthwhile political causes.

3.2 Conduct participating in shaming action

The example above illustrates the difficulty of criminalizing the *initiation* of online shaming so that harassing behaviour could be comprehensively criminalised while leaving out criticism that should be allowed as everyone's right to freedom of expression. It is, however, similarly difficult or even more difficult to criminalize conduct that involves *participating* in shaming action. Yet, as Aitchison and Meckled-Garcia argue, those who join in the shaming action are "still guilty of participating in a shaming action, however imperceptible their contribution".[30] The participants are indeed a crucial factor in the process, since without their contribution there would be no shaming in the first place, but perhaps only some random negative comments of less significance to the target. Therefore, if shaming were to be criminalized, it would not be enough if the offence covered only the deeds of the initiator. However, from the perspective of the principle of maximum certainty and individual autonomy, defining the liability of the participants extra carefully and clearly would be important, since everyone should be able to exercise their freedom of expression and comment on delicate issues without the fear of accidentally committing a criminal offence.

Thus, the legislator should first be able to define punishable shaming action in a way that leaves room for similar legitimate activity, such as drawing attention to important social, moral and political issues and commenting on them publicly. After this, the provision should be able to

30 Aitchison and Meckled-Garcia, "Against Online Public Shaming," 7.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

define participation in shaming by differentiating conduct that contributes to punishable action (hence, should be punishable in itself) from a justifiable comment or remark that is made while a shaming action is running.

Assuming that shaming would be possible to define properly and would be criminalized as such, what kind of conduct could and should be punishable as *participating* in such action? Let us think about a situation in which someone has initiated a shaming action against another person, and as a result, several people were provoked to post public comments and send private messages to and about the target on different social media and online platforms. In assessing the liability of the participants, we would face problems of defining clearly enough the line between exercising one's freedom of expression and participation in shaming at three different stages. Firstly, how would every social media user be able to know about all or most of the comments and messages the target has received, particularly taking into account the rather rapid reaction expected in social media conversations? Secondly, how would they know when this complex of comments and messages constitutes a punishable shaming action (and not a similar action that is, however, considered justifiable criticism) hence indicating a risk of being prosecuted as a participant in case they decide to comment on the issue? Thirdly, how could they know if their own comments are considered a part of shaming action and not merely comments or criticism on a topical issue while a shaming action is still ongoing? In other words, if someone wanted to comment on the issue linked to the target and the ongoing shaming action, yet, without the intention of participating in shaming and harming the target, how could they do that without the risk of being prosecuted as a participant in shaming?

According to Article 10(2) of the ECHR, restrictions on freedom of expression must be prescribed by law, they must be necessary in a democratic society, and they must protect the interests mentioned in the Article. Criminalizing shaming does not necessarily violate any of these conditions *per se*. However, as mentioned earlier, the requirement of written law means that everyone should know "on the basis of the wording of the relevant provision … which acts or omissions will make him criminally liable".[31] Similarly to the culpability of the initiator, the above mentioned questions and many others on participation are impossible to define clearly and objectively in a law. The criteria and evaluation of those criteria would easily fall short of the requirements of the principles of legality

31 *ECtHR*, *Kokkinakis v. Greece*, App. no. 14307/88, 25 May 1993.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

and individual autonomy as well as the requirements of objectivity and non-arbitrariness. Moreover, problems regarding the clarity of the offence would have a "chilling effect" on the permissible exercise of freedom of expression.[32]

4. Conclusions

As argued in this paper, it is difficult to investigate and prosecute offences restricting freedom of expression in terms of crimes committed in the online environment. The difference between criminalised and legitimate expressions is often equivocal even when only one person and their expression(s) are under scrutiny. In a case of online shaming, where a vast amount of expressions and different conduct are committed by numerous people, evaluating whether some of these expressions constitute some offences separately and / or together, or whether the expressions are or should be allowed or restricted according to the standards on freedom of expression is difficult.

Several reasons account for why rules on restricting freedom of expression must be clear. It is important to know what kind of criticism is allowed, but also to avoid a chilling effect on public discussion; everyone should be able to make critical remarks without fear of other people being provoked into sending hateful messages, and this leading to an accusation of online shaming. Therefore, as it is difficult to draw a line between an expression intentionally initiating a shaming action and merely making a critical remark on something or someone, it is not possible to criminalize *initiating* online shaming. Likewise, everyone should be able to make critical remarks without fear of being prosecuted as a *participant* in a shaming action without even being aware of such an action having been running.

On the other hand, everyone should be able to discuss publicly without fear of ending up as a target of online shaming. Despite being severely

---

32 On the "chilling effect" in the praxis of the ECtHR, see e.g. *Yaşar Kaplan v. Turkey*, App. no. 56566/00, 24 January 2006, § 35; *Aslı Güneş v. Turkey*, App. no. 53916/00, 27 September 2005 (dec.); *Nikula v. Finland*, App. no. 31611/96, 21 March 2002, § 54; *The Magyar Jeti Zrt v. Hungary*, App. no. 11257/16, 4 December 2018, §§ 83-84; *Eon v. France*, App. no. 26118/10, 14 March 2013, §§ 34-36; *Margulev v. Russia*, App. no. 15449/09, 8 October 2019, § 42; *Sylka v. Poland*, App. no. 19219/07, 3 June 2014 (dec.); *Guja v. Moldova*, App. no. 14277/04, 12 February 2008 (GC), § 78; *Fuentes Bobo v. Spain*, App. no. 39293/98, 29 February 2000, § 49 and *Heinisch v. Germany*, App. no. 28274/08, 21 July 2011, § 91.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

disturbing to the target and harmful for public discussion, it is difficult or even impossible to comprehensively criminalize acts of harmful shaming without excessively restricting freedom of expression. In addition, criminalizing shaming as some kind of joint action would mean a deviation from the principles or ideas of legality and individual justice that form the cornerstones of European criminal justice systems. At least as long as criminal law concentrates on individuals and their specific and individual acts and as long as it presupposes rationality of both the law and people governed by law, criminal law is ill-equipped to deal with multifaceted social problems. However, it remains to be seen whether the EU's proposal on the Digital Services Act[33] can provide a safer online environment to citizens and afford better protection to their fundamental and human rights. While the possibilities of using criminal law to prevent hate speech are limited, other means might be used to prevent hate speech and its harmful effects.[34]

Bibliography

Aitchison, Guy and Meckled-Garcia, Saladin. "Against Online Public Shaming: Ethical Problems with Mass Social Media." *Social Theory and Practice* 47, no.1 (2021), 1-31.

Ashworth, Andrew. *Principles of criminal law,* 6th edn. Oxford: Oxford University Press, 2009.

Billingham, Paul and Parr, Tom. "Online Public Shaming: Virtues and Vices." *Journal of Social Philosophy* 51, no. 3 (2020), 371-390.

Blomsma, Jeroen. "Mens Rea and Defences in European Criminal Law." Cambridge: Intersentia, 2012.

Bromwich, David and Kateb, George, eds. *On Liberty: Rethinking the Western Tradition*. New Haven: Yale University Press, 2003.

Chick, Kathryn. "Harmful Comments on Social Media." *York Law Review* 1, (2020): 83-110.

Coe, Peter. "The Social Media Paradox: An Intersection with Freedom of Expression and the Criminal Law." *Information & Communications Technology Law* 24, no. 1 (2015): 16-40.

33 Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM(2020) 825 final.

34 For example, the Digital Services Act proposal includes rules for online intermediary services with notice and action procedures, complain and redress mechanisms and measures against abusive notices and counter-notices.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

CPS/Director of Public Prosecutions. "Guidelines on Prosecuting Cases Involving Communications Sent Via Social Media." June 20, 2013. Accessed April 4 2021. http://data.parliament.uk/DepositedPapers/Files/DEP2013-1025/social_media_guidelines.pdf.

Emerson, Thomas I. "Toward a General Theory of the First Amendment." *The Yale Law Journal* 72, no. 5 (1963): 877-956.

Fletcher, George P. *The Grammar of Criminal Law: American, Comparative, and International. Volume One: Foundations*. Oxford: Oxford University Press, 2007.

Kargopoulos, Alexandros. "Fundamental rights, national identity and EU criminal law." In *Research Handbook On EU Criminal Law,* edited by Valsamis Mitsilegas, Maria Bergström and Theodore Konstadinides, 125-147. Cheltenham, UK: Edward Elgar Publishing, 2016).

Keipi, Teo, Näsi, Matti Johannes, Oksanen, Atte, and Räsänen Pekka. *Online Hate and Harmful Content: Cross-National Perspectives.* London: Routledge, 2017.

Koivukari, Kristiina. "The crumbling narrative of modern European criminal justice." Dissertation, University of Helsinki, 2020.

Korpisaari, Päivi. "Balancing freedom of expression and the right of private life in the European Court of Human Rights - application and interpretation of the key criteria." *Communications Law* 22, no. 2 (2017): 39-50.

Laidlaw, Emily B. "Online Shaming and the Right to Privacy." *Laws* 6, no. 1 (2017): 1-26.

Norrie, Alan. *Crime, Reason and History, A Critical Introduction to Criminal Law,* 2nd edn. London: Butterworths, 2001.

Oster, Jan. *Media Freedom as a Fundamental Right*. Cambridge: Cambridge University Press, 2015.

Peristeridou, Christina. *The Principle of Legality in European Criminal Law*. Cambridge: Intersentia, 2015.

Sunstein, Cass. "Believing False Rumors." In *The Offensive Internet: Speech, Privacy, and Reputation*, edited by Saul Levmore and Martha C. Nussbaum, 91-106. Cambridge, Mass: Harvard University Press, 2010.

Taebi, Behnam and Safari, Azar. "On Effectiveness and Legitimacy of 'Shaming' as a Strategy for Combatting Climate Change." *Science & Engineering Ethics* 23 no. 5 (2017): 1289-1306.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Role of Occupational Safety and Health Legislation in Hate Speech Regulation.
Employers' responsibility to prevent and respond to the risk of hate speech at work – the Finnish perspective

Enni Ala-Mikkula

Abstract: Occupational safety and health legislation plays an important role in regulating hate speech when the target is an employee. In the case of hate speech at work, there is a need for the employer to implement both preventive and responsive safety measures. The required safety measures can be categorized either as a procedural instruction or as a measure of support. In addition, there might be a need for working arrangements. The decision on which measures are needed is for employers to make since no specific requirements are set in the Finnish Occupational Safety and Health Act (738/2002, OSHA). The Finnish OSHA has its basis in EU law and is built on similar objectives and requirements to EU directives.

Keywords: hate speech, online hate, cyber hate, social media, occupational safety and health, work, risk, employer, responsibility, safety measure

Chapter 1. Introduction

Hate speech constitutes a new challenge not only for criminal law but for other fields of law as well. One of these fields facing new challenges is occupational safety and health legislation. In addition, hate speech is a challenge for each employer's occupational safety and health activities. Both regulation and workplace activities have historically been centred round the physical safety of employees.[1] However, the scope of occupational safety and health regulation is not limited only to the physical aspects of

1 See Tapio Kuikko, *Työturvallisuus ja sen valvonta*, 4th ed. (Helsinki: Talentum, 2006), 17; Kari Eklund and Asko Suikkanen, *Työväensuojelusta työsuojeluun: Työsuojelun ja työolojen kehitys Suomessa 1970-luvulla* (Helsinki: Tammi, 1982), 10.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

safety and health: it also involves psychological and social dimensions.[2] The employer is first and foremost responsible for the safety and health of employees at work.[3]

Employees also require protection when it comes to the risk of hate speech at work. Hate speech can be encountered at work, for example in a situation where the employee is encouraged to be active in social media by the employer or the employee's expert status is otherwise such that he or she is in the public eye due to their work.[4] Someone encountering hate speech can be caused mental load which can, while continuing and in excessive quantities, result in serious consequences and can eventually compromise the employee's health.[5] The problem of increased online hate in the workplace should thus be taken seriously. The phenomenon is, however, quite new to workplaces and has not really, or at least not sufficiently, been addressed in public discussions concerning occupational safety and health.[6] As a result, some ambiguities and lack of awareness still seem to remain in terms of the employer's responsibilities in the context of hate speech encountered at work.[7]

The aim of this study[8] is to determine the actions required from the employer in case its employees are targeted with hate speech at work.

---

2 According to the Finnish Occupational Safety and Health Act (738/2002, OSHA) section 1, the term 'health' as used in the Act covers both physical and mental health of employees. See also Government of Finland, *HE 59/2002 vp: Hallituksen esitys eduskunnalle työturvallisuuslaiksi ja eräiksi siihen liittyviksi laeiksi* (Helsinki: Government of Finland, 2002), 16, 23.

3 See for example Berta Valdés de la Vega, "Occupational Health and Safety: An EU Law Perspective," in *Health and Safety at Work: European and Comparative Perspective*, ed. Edoardo Ales (The Netherlands: Kluwer Law International BV, 2013), 16.

4 See Päivi Rauramo et al., "Sosiaalisen median työkäyttö: Työsuojelunäkökulma," Työturvallisuuskeskus, last modified August 18, 2014, https://tyoturvallisuuskeskus.mobiezine.fi/zine/8/cover.

5 See Rauramo et al., "Sosiaalisen median työkäyttö."

6 See, however, "Häiritsevä palaute: Apua vihapuheeseen," Häiritsevä palaute, accessed October 16, 2020, https://www.xn--hiritsevpalaute-0kbh.fi.

7 Kari Mäkinen, *Sanat ovat tekoja: Vihapuheen ja nettikiusaamisen vastaisten toimien tehostaminen* (Helsinki: Sisäministeriö, 2019), 69, https://julkaisut.valtioneuvosto.fi/handle/10024/161613; Oikeusministeriö, *Against hate -hankkeen suosituksia viharikosten ja vihapuheen vastaiseen työhön* (Helsinki: Oikeusministeriö, 2019), 22, https://yhdenvertaisuus.fi/documents/5232670/13949561/Against+Hate+-hankkeen+suositukset+-+FI/58f4e479-001c-daed-0e8d-a60375886602/Against+Hate+-hankkeen+suositukset+-+FI.pdf.

8 The study has been conducted at the University of Helsinki, Faculty of Law as a part of the Hate and public sphere -research project funded by the Helsingin Sanomat Foundation.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The perspective of the study is mainly that of Finland, and the study will concentrate on the requirements imposed on employers by the Finnish Occupational Safety and Health Act (738/2002, OSHA). The study is based on the provisions of the Finnish OSHA which are applicable in a situation where a risk of hate speech is present at work. The aim is to provide an overview of employer responsibilities, the safety measures required, and the role of occupational safety and health legislation as one part of hate speech regulation. Hence, the subject of study is defined in a way which does not enable profound, detailed, or exhaustive analysis of the responsibilities or their requirements.

The Finnish OSHA was reformed at the beginning of the 2000s on the basis of the requirements of EU law. Through the influence of the Directive on the introduction of measures to encourage improvements in the safety and health of workers at work (89/391/EEC, framework directive) the Finnish OSHA also contains a more proactive and dynamic standpoint on occupational safety and health thinking and activities.[9]

According to the introductory phrases of the framework directive, its provisions apply to all kinds of risks in the working environment. As a directive covering all risks connected with safety and health in the workplace, the framework directive's aim has, according to its name, been to serve as a basis for more specific directives. At the same time, the scope of the directive can be interpreted in a way which is not limited only to the physical aspects of work, but in a wider understanding of the term 'working environment' and safety within that environment.[10] The framework directive is based on the idea of prevention and improvement. Employers should, when acting accordingly, also keep themselves informed about the latest advances in technology and scientific findings concerning design of the workplace and the risks that work entails.[11] The approach is flexible and the responsibilities and requirements set for the employer concerning the safety and health of employees also apply when the world of work is rapidly changing.[12]

---

9 See for example Government of Finland, *HE 59/2002 vp*, 1.

10 See David Walters, "The Framework Directive," in *Regulating Health and Safety Management in the European Union: A study of the Dynamics of Change*, ed. David Walters (Brussels: P.I.E.-Peter Lang, 2002), 43.

11 See general obligations on employers under article 6 of the framework directive.

12 Walters, "The Framework Directive," 46.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 2. The case of the Finnish Occupational Safety and Health Act

2.1. Employers' general obligations and the aim of preventing the risk of hate speech at work

Like the framework directive, the Finnish OSHA is a kind of a framework law which is general in nature. That is also well-founded because of the Act's broad scope of application.[13] This also means that there is room for consideration when employers make decisions on required safety and health measures or activities on different occasions in diverse workplaces.[14] The Act contains two kinds of responsibilities set for the employer. The general responsibilities should be observed in every workplace regardless of the work done there or the risks that the work entails. In addition, the Act contains a few more specific responsibilities. These risk-specific responsibilities are, however, general in nature, as well.[15]

Since the Finnish OSHA was reformed in 2002, the risks caused by digitalization and the use of social media at work were not taken into consideration when drafting the law. Later some changes have been in the Act, but the newest kinds of risks caused by digitalization have not yet been specially considered. These are risks which are not regulated in inferior statutes either. In terms of the risk of hate speech at work, this means that the requirements set for the employer and its occupational safety and health activities are solely based on the Finnish OSHA and its more or less general provisions.[16] Hate speech at work is, however, a risk to be taken into account, since the Finnish OSHA is based on similar thoughts and requirements concerning occupational safety and health in the workplace as the framework directive.[17]

As noted earlier, both general and risk-specific obligations are set for the employer in the Finnish OSHA. The starting point for the employer's occupational safety and health obligations is its general duty to exercise care.[18] According to section 8 employers must take care of the safety and health of employees while at work by taking necessary measures. This is

13 See Government of Finland, *HE 59/2002 vp*, 29.
14 Pertti Siiki, *Työturvallisuuslainsäädäntö: Työnantajan ja työntekijän velvollisuudet ja oikeudet* (Helsinki: Edita Publishing Oy, 2002), 12.
15 Siiki, *Työturvallisuuslainsäädäntö*, 52.
16 See Siiki, *Työturvallisuuslainsäädäntö*, 52.
17 See Government of Finland, *HE 59/2002 vp*, 22.
18 See Government of Finland, *HE 59/2002 vp*, 22.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

also the main objective of the obligations set for the employer: to protect employees' safety and health at work.[19]

Section 8 also contains other objectives set for the employer's occupational safety and health activities: the objective of continuous improvement and the objective of comprehensive safety management.[20] These objectives and the principles of risk prevention mentioned under section 8 are factors that raise the requirement level set for occupational safety and health activities in the workplace. However, there are also factors which lower the level of requirements. These include, for example, unforeseeability[21] and the necessity requirement.[22]

According to section 8, the employer is, in addition, responsible for continuously monitoring the working environment. Through continuous monitoring the employer can detect risks in the working environment, the state of the working community, and the safety of working practices. Another important obligation at the risk observation phase is analysis and assessment of risks at work.[23] This is covered by section 10, according to which the employer should systematically and adequately analyse and identify the hazards and risks caused by work. If risks detected cannot be eliminated, the employer should assess their consequences for employees' safety and health. Risks identified should primarily be eliminated, but if that is not possible they should at least be minimized. While reacting to detected risks, the employer should obey the principles of risk prevention mentioned in section 8 and, for example, adopt safety measures with a general impact before adopting any individual measures.

The obligation to provide instruction and guidance for employees, section 14, is one of the individual measures. According to section 14, employers should give their employees necessary information on workplace risks. Employers should also ensure that their employees are given instruction and guidance in order to eliminate risks of their work and to

---

19 See Enni Ala-Mikkula, *Työnantajan työsuojeluvastuu: Tutkimus työnantajan keskeisistä työsuojeluvelvollisuuksista sekä niissä työnantajan työsuojelutoiminnalle asetetusta vaatimustasosta* (Helsinki: Alma Talent, 2020), 75.

20 See Government of Finland, *HE 59/2002 vp*, 22; Jorma Saloheimo, *Työturvallisuus: Perusteet, vastuu ja oikeusturva*, 3rd ed. (Helsinki: Talentum Pro, 2016), 77-78, 80-81.

21 See OSHA section 8.2: "Such unusual and unforeseeable circumstances which are beyond the employer's control, and such exceptional events the consequences of which could not have been avoided despite the exercise of all due care, are taken into consideration as factors restricting the scope of the duty to exercise care."

22 See Ala-Mikkula, *Työnantajan työsuojeluvastuu*, 199-200, 203.

23 Ala-Mikkula, *Työnantajan työsuojeluvastuu*, 69-70.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

avoid any risk from their work jeopardizing safety and health. As section 14 is part of an employer's general occupational safety and health responsibilities, it is an obligation of each employer to provide the necessary instruction and guidance for employees. As an individual safety measure, instruction and guidance provided supplements those measures which are more general by their impact, such as structural, technical, or organisational measures.[24]

All in all, the employers' general obligations aim to prevent the hazards and risk factors of the working environment in advance. To be able to prevent creation of risk factors or to eliminate or minimize them, the employer must detect and recognize risks which concern the work and workplace in question. One of these might be the risk of hate speech at work. If this is the case, the risk of hate speech must also be assessed and proper measures taken in order to react and respond to the risk. The general obligations which employers should take into account in order to be appropriately prepared for the risk of hate speech at work are presented in table 1.

24 See Ala-Mikkula, *Työnantajan työsuojeluvastuu*, 159.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Table 1. The nature and requirements of employers' general obligations (OSHA chapter 2) of key importance when aiming at preventing the risk of hate speech at work.

	Occupational Safety and Health Act / Employers' general obligations (chapter 2)			
Provision	**Section 8**	**Section 10**	**Section 13**	**Section 14**
Subject	Employers' general duty to exercise care	Analysis and assessment of risks at work	Work design	Instruction and guidance to be provided for employees
The nature of the responsibility	Proactive, the objectives and limits of employers' occupational safety and health activities and safety measures which aim at detecting risks or which guide the process of choosing the right response measures	Proactive, aim at detecting risks	Proactive, aim at detecting risks especially during the process of design and change	Proactive, individual measure of response
Required measures	Preventing creation of risks, eliminating or minimizing risks, prioritizing measures which have a general impact, adjustment to technological developments, continuous monitoring	Systematic analysis and identification of risks, assessment if they cannot be eliminated	Designing and planning work according to the physical and mental capacities of employees	Necessary information on risks for employees, adequate orientation to working methods, instruction and guidance in order to avoid risks, completion of instruction and guidance
Objective of required measures	Protection of employees, continuous improvement, comprehensive safety management; necessity requirement and unforeseeability as limits	Understanding of risks and assessing their probability and gravity; eliminating risks or measures needed for minimizing risks	Avoiding or reducing risks to the safety and health of employees	Employees are also capable of eliminating and avoiding risks to safety and health
Factors to consider / source of potential risks	Work, working conditions and working environment, employees' personal capacities	Especially the risks mentioned in chapter 5, accidents and occupational diseases, employees' personal capacities, factors related to workload, potential risks to reproductive health	Workload factors	Work, working conditions, working methods and working equipment, changes in working tasks, adjustment, cleaning and repair work, disturbances and exceptional situations

Risks must also be assessed when designing and planning work. According to section 13, the physical and mental capacities of employees must be taken into account when designing and planning their work, in order to avoid or reduce risks from workload factors to the safety and health of employees. However, there is no need for individual assessment that considers the differences between each individual in stamina and tolerance to stress. Instead, the work should be designed in a way which enables an average person to perform it without being excessively loaded.[25] Overall, when assessing risks, the employer must, according to section 10, consider factors related to workload and also the employees' age, gender, occupa-

25 Government of Finland, *HE 59/2002 vp*, 35.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tional skills and other personal features and abilities. Special attention must be paid to risks of work that are the basis for employers' risk-specific responsibilities in the Finnish OSHA.

2.2. Employers' risk-specific responsibilities and responses to the risk of hate speech at work

The Finnish OSHA does not include any risk-specific responsibilities set for the employer that would specifically address the risk of hate speech at work. However, there are responsibilities which concern not only the physical but also the mental and social aspects of employees' safety and health. Those risk-specific responsibilities which may also apply to a situation where there is a need to respond to the risk of hate speech at work are presented in table 2.

Table 2. The nature and requirements of employers' risk-specific obligations (OSHA chapter 5) to be considered when responding to the risk of hate speech at work.

	Occupational Safety and Health Act / Employers' risk-specific obligations (chapter 5		
Provision	**Section 25**	**Section 27**	**Section 28**
Subject	Avoiding and reducing workloads	Threat of violence	Harassment
Nature of responsibility	Reactive, prerequisites: 1) noticed exposure, 2) compromise in health, 3) awareness	Proactive, prerequisites: 1) assessed, evident threat of violence	Reactive, prerequisites: 1) occurrence of harassment, 2) compromise in health, 3) awareness
Required measures	Available means	Appropriate safety arrangements and equipment, opportunity to summon help, procedural instructions	Available means
Objective of required measures	Analysis of workload factors, avoiding or reducing risk	Preventing the threat of violence and incidents of violence, controlling or restricting the effects of violent incidents	Remedying the situation
Source of potential risks	Work, insufficient control, working arrangements etc.	Especially clients, other outsiders	Internal conflicts, clients, other outsiders

One of these responsibilities is avoiding and reducing workloads, section 25, which has a more personal and reactive approach to workload than

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

that of section 13.[26] According to section 25, the employer is obliged to take available measures to analyse workload factors and to avoid or reduce the risk caused by those factors if it is noticed that an employee is exposed to workloads while at work in a manner that endangers their health. The obligation is realized when the employer has become aware of the matter.

The situation is similar in terms of section 28 and employer's obligations concerning harassment. If harassment or other inappropriate treatment of an employee at work occurs and this causes hazards or risks to the employee's health, the employer must, after becoming aware of the matter, take available measures to remedy the situation. The harassment provision is mainly targeted at situations where harassment occurs inside the organisation, but someone who treats employees inappropriately may also be a client or other outsider.[27] Harassment and inappropriate treatment can take different forms. One possible form is online harassment. This kind of harassment can also lead to workload which endangers an employee's health.

Harassment online can also mean threatening someone's physical safety and health. If this is the case, section 27 can also be applied. Unlike sections 25 and 28, section 27 requires preventive measures from the employer in case there is an evident threat of violence in the workplace. Work and working conditions must be arranged so that the threat of violence and incidents of violence are prevented as far as possible. This means appropriate safety arrangements and equipment, and procedural instructions. In the instructions, controlling threatening situations should be considered in advance. In addition, practices for controlling or restricting the effects that violent incidents can have on employee safety should be presented. This is the most individualized and concrete safety measure required from the employer by the Finnish OSHA that also applies in situations where a risk of hate speech is present at work.

2.3. Concrete safety measures based on responsibilities

All in all, safety measures which can be expected from the employer in the case of hate speech at work aim either at preventing the risk of hate speech or at responding to the risk of hate speech and its consequences.

---

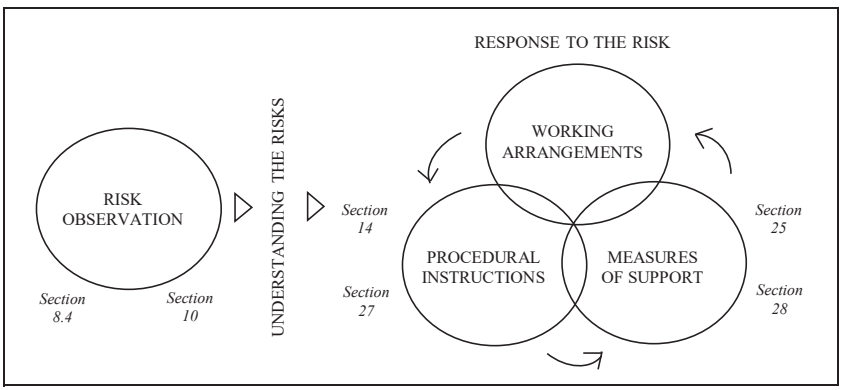
26 See Jenny Rintala, "Työn psykososiaaliset kuormitustekijät työturvallisuuslaissa," in *Työturvallisuusoikeus*, Johanna Havula et al. (Helsinki: Edita, 2018), 158.

27 Government of Finland, *HE 59/2002 vp*, 43.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Safety measures can be categorized either as a procedural instruction or as a measure of support.[28] In addition, some measures concerning working arrangements are available, and the need for those measures and arrangements may be based either on assessment of risks and a high probability or gravity of the risk of hate speech or on discussions that the employer and employee have had after the employee has faced hate speech at work and responsive measures are being considered. These safety measures and the role of different responsibilities in preventing and responding to the risk of hate speech at work are presented in figure 1.

Figure 1. Safety measures required and aimed either at preventing the risk of hate speech at work or employed as a response to risk.



Procedural instructions can contain instructions for supervisors and employees on how to act and proceed in a case of online harassment or in the case of an evident threat of violence in the workplace. There may also be a need for instructions concerning information security, the use of social media at work in general, the moderation policy of the organization – just to name a few.[29] The rules of discussion and the principles of modera-

28 Rauramo et al., "Sosiaalisen median työkäyttö."

29 See "Edilex Uutiset: Työpaikoilla tulisi olla yhteisesti sovitut periaatteet ja käytännöt somessa havaittuun tai koettuun epäasialliseen kohteluun puuttumiseksi," Edilex, accessed October 20, 2020, https://www.edilex.fi/uutiset/41223; J. Nathan Matias, "Posting Rules in Science Discussions Prevents Problems & Increases Participation," CivilServant, last modified April 29, 2019, https://civilservant.io/moderation_experiment_r_science_rule_posting.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tion should be defined by the administrator of each platform.[30] If it is a question of the employer's own websites and social media channels, this should be attended to by the employer concerned. Employers should for their part aim at promoting a correct and respectful discussion culture on the platforms they administer.[31]

The need for different kinds of instructions is based on either section 14 in general or on section 27 in terms of an evident threat of violence. In addition, the employer may be obliged to offer special training for employees. However, these measures are always secondary in comparison to those measures which are general in impact.[32] That is, the employer should try to eliminate the risk if possible, instead of just settling for the possibility of minimizing the risk by instructing employees. However, the risk of hate speech when it appears anonymously and in social media is of such a nature that it is hardly ever completely eliminable.

Measures of support aim at a situation where an employee who has encountered hate speech at work receives the amount of support needed in order to deal with what has happened.[33] Occupational health services should be in place and the employee should be provided with other discussion possibilities, too, with the supervisor and peer support if needed.[34] The employer should also provide the employee with help and support if it comes to possible proceedings involving the authorities.

If the risk of hate speech at work is considered high, stronger safety measures should be considered, too. These measures are aimed mainly at working arrangements or working methods and are such by nature that they normally strongly affect the nature of the work itself and the essence of expert work. That is why these measures, like de-identification of work outputs and changes in working assignments, may not be particularly desirable or even possible options. Of course, some technical measures,

30 Päivi Korpisaari, "Sananvapaus verkossa – yksilöön kohdistuva vihapuhe ja verkkoalustan ylläpitäjän vastuu," *Lakimies* 7-8 (December 2019): 951.

31 Mäkinen, *Sanat ovat tekoja*, 18.

32 See Kuikko, *Työturvallisuus ja sen valvonta*, 42; Saloheimo, *Työturvallisuus*, 91.

33 See Mika Illman, *Järjestelmällinen häirintä ja maalittaminen: Lainsäädännön arviointia* (Helsinki: Valtioneuvosto, 2020), 112, https://julkaisut.valtioneuvosto.fi/handle/10024/162579.

34 Rauramo et al., "Sosiaalisen median työkäyttö"; Saloheimo, *Työturvallisuus*, 112; Illman, *Järjestelmällinen häirintä*, 112.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

too, such as automatic comment moderation and discussion facilitation, may – indeed should – be used, too, if possible.[35]

These are all measures for which a need can be based on the obligations set for the employer by the Finnish OSHA. The decision on what measures are needed is for employers to make since no specific requirements are set in law.[36] Because the risk of hate speech at work or other risks caused by digitalization have not been mentioned or specially identified in the Finnish OSHA, the risk of hate speech at work does not seem to receive the attention it should in practice.[37] The need for assessing these kinds of risks should be emphasized, at least in the instructions given by the authorities to workplaces in the future.[38] Additionally, the reactive nature of obligations concerning harassment and workload and strong emphasis on the physical aspect of an evident threat of violence should be re-considered.[39] The employer should be explicitly obliged also to take preventive measures in terms of dealing with harassment and workload at work, and regulation should better embody the fact that employees can also end up in threatening situations in other circumstances than during face-to-face encounters. Digitalization has brought with it new risks with effects on the safety of the working environment, a situation that should be taken into account when considering amendments to the law.

35 See "Häiritsevä palaute"; "Online Harassment Field Manual: Best Practices for Employers," PEN America, accessed October 20, 2020, https://onlineharassmentfieldmanual.pen.org/best-practices-for-employers/.

36 See Siiki, *Työturvallisuuslainsäädäntö*, 52.

37 See Aleksi Knuutila et al., *Viha vallassa: Vihapuheen vaikutukset yhteiskunnalliseen päätöksentekoon* (Helsinki: Valtioneuvosto, 2019), 10, https://julkaisut.valtioneuvosto.fi/handle/10024/161812.

38 See Suomen Lakimiesliitto, "Lausunto maalittamista koskevaan selvitykseen" (Report, Helsinki, August 8, 2020), 2; Päivi Rauramo, *Työsuojelu ja työhyvinvointi asiantuntija- ja toimistotyössä* (Helsinki: Työturvallisuuskeskus, 2020), 76, 78. Cf. Sosiaali- ja terveysministeriö, *Riskien arviointi työpaikalla -työkirja* (Helsinki: Sosiaali- ja terveysministeriö, 2015), https://ttk.fi/tyoturvallisuus_ja_tyosuojelu/tyosuojelu_tyopaikalla/vastuut_ja_velvoitteet/tyon_vaarojen_selvittaminen_ja_arviointi.

39 See Tieteentekijöiden liitto, "Lausuma koskien Valtioneuvoston kanslian ns. maalittamista koskevaa selvityspyyntöä" (Report, Helsinki, August 31, 2020), 4.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 3. Conclusion

As online hate is a phenomenon related to social media and can be practised through anonymous comments, it is typically beyond the individual employer's sphere of influence. When the perpetrator is not a part of the employer's organisation but a client or other outsider, the employer's means of preventing or responding to online harassment targeting its employees are limited. Outsiders do not operate under the employer's direction[40] and the employer lacks supervisory measures. Therefore, online hate constitutes a work-related risk which cannot be totally prevented in advance. The risk of hate speech at work should, however, be recognized and understood by both employers and their employees,[41] and guidelines and instructions should be prepared in case the risk later materialises. This is a requirement which should be fulfilled in order to limit the effects that facing hate speech at work can have on an employee's health.[42]

Since the employer's possibilities to prevent hate speech targeting employees are limited, some other kind of legislation and regulation aiming at restraining hate speech should also be in place. This is a question of the combined effect which different laws can have together.[43] Hate speech is a complex problem and there is no simple solution. Instead, there is a need for broad legislative measures concerning, for example, criminal and procedural law, and other activities regarding, for example, occupational safety and health, too.[44] In short, a combination of diverse measures should be utilized when trying to control the increase in open expressions of hate in the context of social media.[45]

---

40 See the Finnish Employment Contracts Act (55/2001) chapter 1, section 1: "This Act applies to contracts (employment contracts) entered into by an employee, or jointly by several employees as a team, agreeing personally to perform work for an employer under the employer's direction and supervision in return for pay or some other remuneration."

41 See for example Rauramo, *Työsuojelu ja työhyvinvointi*, 76, 78.

42 Illman, *Järjestelmällinen häirintä*, 111.

43 Aluehallintovirasto, "Lausunto maalittamista ja vihapuhetta koskevaan selvitystyöhön" (Report, Helsinki, August 31, 2020), 4.

44 Poliisihallitus, "Lausunto maalittamista koskevaan selvitykseen" (Report, Helsinki, July 24, 2020), 6-7.

45 See Teo Keipi et al., *Online Hate and Harmful Content: CrossNational perspectives* (London and New York: Routledge, 2017), 1-2, OAPEN Free.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Finnish Government conducted a review on legislation concerning targeting,[46] with different interest groups, such as employer and employee organisations, stating in their reports that measures should be multiple when dealing with a multidimensional problem such as targeting. Some criminal legislation should be in place in order to tackle the problem through regulatory effect. However, criminal legislation alone is not an answer as criminal processes are often slow and heavy. In addition, the question of freedom of expression arises. This is a fundamental right which should not be restrained excessively by criminal legislation. Therefore, the crime threshold in terms of hate speech and targeting should be set quite high. Since hate speech and targeting can be considered as challenges for criminal law, there is a need for other measures to counteract the consequences of hate speech, too, as the consequences may prove to be seriously damaging and harmful.[47] One aspect to be considered is the occupational safety and health viewpoint and employers' responsibilities which have been under scrutiny in this study. The safety measures required from the employer have an important role to play when the target of hate speech is an employee and the employee's work duties or position is the reason behind hate speech.[48]

The right to work in peace and safety at work is a fundamental right, as is freedom of expression; indeed, it should be a guarantee for each employee. On the one hand society and on the other hand individual employers are obliged to ensure that employees are free to do their job in a safe and sound environment.[49] Each employer has a general duty of care set by the Finnish OSHA in terms of the safety and health of its employees at work. Employer responsibilities in the context of occupational safety and health are based on the employer's general duty of care throughout the EU,[50] and in general it covers all kinds of different risks and hazards caused by

46 Targeting can be understood as "a complex of hateful expressions in which someone sparks off a hate campaign against another, usually on social media". Päivi Korpisaari and Kristiina Koivukari, "Hate speech and targeting individuals online – a new challenge for criminal law" (Concept note for Hate Speech & Platform Regulation, international online-workshop, October 17, 2020), 1.

47 Suomen Journalistiliitto, "Lausunto maalittamista koskevaan selvitykseen" (Report, Helsinki, August 24, 2020), 1.

48 Illman, *Järjestelmällinen häirintä*, 105.

49 Suomen Syyttäjäyhdistys ry, "Lausuma maalittamista koskevassa asiassa" (Report, Helsinki, August 6, 2020), 5.

50 See article 6.1, framework directive: "Within the context of his responsibilities, the employer shall take the measures necessary for the safety and health protection of workers –". See also Walters, "The Framework Directive," 46.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

work, the working environment, or working conditions. The general duty of care requires protective measures from the employer, but the choice of safety measures necessary rests ultimately with each employer and is dependent on the work involved and on the risks it entails. At the same time the employer's occupational safety and health responsibilities are an example of a legislative structure through which the responsibilities of social media enterprises on discussions and expressions of hate presented via their platforms could also be regulated.[51]

As for criminal sanctions, an amendment to the Criminal Code of Finland (39/1889) is under consideration. The proposed amendment would enable a public prosecutor to bring charges for menace[52] not only when the injured party reports the offence for bringing charges, but also in other circumstances when a person has been threatened due to their working duties and the offender does not belong to the personnel of the workplace.[53] If these kinds of hate crimes were under public prosecution, the employer's possibilities to take care of its responsibility to protect its employees' health and safety at work could also be enhanced and improved.[54] In the case of actionable offences, legal proceedings may seem more personified and the risk of increased threats and continued harassment due to legal proceedings may decrease employee willingness to report offences.[55]

---

51 Lorna Woods and William Perrin, *Online harm reduction: a statutory duty of care and regulator* (Dunfermline: Carnegie UK Trust, 2019), 5, https://www.carnegieuktrust.org.uk/publications/online-harm-reduction-a-statutory-duty-of-care-and-regulator/ in which they propose a similar kind of statutory duty of care to regulate social media enterprises when it comes to reducing harm in social media.

52 See the Criminal Code of Finland chapter 25, section 7: "A person who raises a weapon at another or otherwise threatens another with an offence under such circumstances that the person so threatened has justified reason to believe that his or her personal safety or property or that of someone else is in serious danger shall, unless a more severe penalty has been provided elsewhere in law for the act, be sentenced for menace to a fine or to imprisonment for at most two years."

53 See Government of Finland, *HE 226/2020 vp: Hallituksen esitys eduskunnalle laiksi rikoslain 25 luvun 9 §:n muuttamisesta* (Helsinki: Government of Finland, 2020), 1.

54 See Jani Hannonen, *Luonnos hallituksen esitykseksi laiksi rikoslain 25 luvun 9 §:n muuttamisesta: Lausuntotiivistelmä* (Helsinki: Oikeusministeriö, 2020), 13-15, https://julkaisut.valtioneuvosto.fi/handle/10024/162439; Suomen Lakimiesliitto, "Lausunto maalittamista koskevaan selvitykseen," 1-2.

55 See Suomen tuomariliitto, "Lausuma maalittamista koskevassa asiassa" (Report, Tampere, August 14, 2020), 3-4; Yhdenvertaisuusvaltuutettu, "Lausunto maalittamista koskevaan selitykseen" (Report, Helsinki, September 2, 2020), 2. – This is something that has recently been pointed out by the European Commission, as well. See "February infringements package: key decisions," European Commis-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Criminal sanctions should, however, be utilized as a last resort to restrict hate speech.[56] This is also in line with the objectives of occupational safety and health legislation which aims first and foremost at preventing risks, including the risk of hate speech, at work in advance, instead of settling for responsive measures after employees have already been targeted with hate speech and there is a need for remedies. It is ultimately quite rare that perceived online hate would result in criminal responsibility or liability for damages which would provide legal protection for the victim. Hence, the protection and support provided by the employer and its occupational safety and health activities is of great importance in cases where hate speech is work-related.[57] Employer support and occupational safety and health measures are always needed in the case of work-related harassment which may potentially harm employees' health – in those cases where harassment experienced does not constitute an offence as well.[58]

Bibliography

Ala-Mikkula, Enni. *Työnantajan työsuojeluvastuu: Tutkimus työnantajan keskeisistä työsuojeluvelvollisuuksista sekä niissä työnantajan työsuojelutoiminnalle asetetusta vaatimustasosta*. Helsinki: Alma Talent, 2020.

Aluehallintovirasto. "Lausunto maalittamista ja vihapuhetta koskevaan selvitystyöhön." Report, Helsinki, August 31, 2020.

Bayer, Judit, and Petra Bárd. *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*. Brussels: European Union, 2020. https://doi.org/10.2861/28047.

Edilex. "Edilex Uutiset: Työpaikoilla tulisi olla yhteisesti sovitut periaatteet ja käytännöt somessa havaittuun tai koettuun epäasialliseen kohteluun puuttumiseksi." Accessed October 20, 2020. https://www.edilex.fi/uutiset/41223.

Eklund, Kari, and Asko Suikkanen. *Työväensuojelusta työsuojeluun: Työsuojelun ja työolojen kehitys Suomessa 1970-luvulla*. Helsinki: Tammi, 1982.

---

sion, last modified February 18, 2021, https://ec.europa.eu/commission/presscorner/detail/en/INF_21_441.

56 See Judit Bayer and Petra Bárd, *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches* (Brussels: European Union, 2020), 14, https://doi.org/10.2861/28047.

57 Atte Oksanen and Päivi Korpisaari, "Viha ja julkisuus: Päätöksentekoon vaikuttamaan pyrkivä vihapuhe yhteiskunnallisesta ja oikeudellisesta näkökulmasta" (Research plan, October 2019), 5.

58 Illman, *Järjestelmällinen häirintä*, 139.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

European Commission. ”February infringements package: key decisions.” Last modified February 18, 2021. https://ec.europa.eu/commission/presscorner/detail/en/INF_21_441.

Government of Finland. *HE 226/2020 vp: Hallituksen esitys eduskunnalle laiksi rikoslain 25 luvun 9 §:n muuttamisesta*. Helsinki: Government of Finland, 2020.

Government of Finland. *HE 59/2002 vp: Hallituksen esitys eduskunnalle työturvallisuuslaiksi ja eräiksi siihen liittyviksi laeiksi*. Helsinki: Government of Finland, 2002.

Hannonen, Jani. *Luonnos hallituksen esitykseksi laiksi rikoslain 25 luvun 9 §:n muuttamisesta: Lausuntotiivistelmä*. Helsinki: Oikeusministeriö, 2020. https://julkaisut.valtioneuvosto.fi/handle/10024/162439.

Häiritsevä palaute. “Häiritsevä palaute: Apua vihapuheeseen.” Accessed October 16, 2020. https://www.xn--hiritsevpalaute-0kbh.fi.

Illman, Mika. *Järjestelmällinen häirintä ja maalittaminen: Lainsäädännön arviointia*. Helsinki: Valtioneuvosto, 2020. https://julkaisut.valtioneuvosto.fi/handle/10024/162579.

Keipi, Teo, Matti Näsi, Atte Oksanen, and Pekka Räsänen. *Online Hate and Harmful Content: CrossNational perspectives*. London and New York: Routledge, 2017. OAPEN Free.

Knuutila, Aleksi, Heidi Kosonen, Tuija Saresma, Paula Haara, and Reeta Pöyhtäri. *Viha vallassa. Vihapuheen vaikutukset yhteiskunnalliseen päätöksentekoon*. Helsinki: Valtioneuvosto, 2019. https://julkaisut.valtioneuvosto.fi/handle/10024/161812.

Korpisaari, Päivi, and Kristiina Koivukari. “Hate speech and targeting individuals online – a new challenge for criminal law.” Concept note for Hate Speech & Platform Regulation, international online-workshop, October 17, 2020.

Korpisaari, Päivi. "Sananvapaus verkossa – yksilöön kohdistuva vihapuhe ja verkkoalustan ylläpitäjän vastuu." *Lakimies* 7-8 (December 2019): 928-952.

Kuikko, Tapio. *Työturvallisuus ja sen valvonta*. 4th ed. Helsinki: Talentum, 2006.

Matias, J. Nathan. ”Posting Rules in Science Discussions Prevents Problems & Increases Participation.” CivilServant. Last modified April 29, 2019. https://civilservant.io/moderation_experiment_r_science_rule_posting.html

Mäkinen, Kari. *Sanat ovat tekoja: Vihapuheen ja nettikiusaamisen vastaisten toimien tehostaminen*. Helsinki: Sisäministeriö, 2019. https://julkaisut.valtioneuvosto.fi/handle/10024/161613.

Oikeusministeriö. *Against hate -hankkeen suosituksia viharikosten ja vihapuheen vastaiseen työhön*. Helsinki: Oikeusministeriö, 2019. https://yhdenvertaisuus.fi/documents/5232670/13949561/Against+Hate+-hankkeen+suositukset+-+FI/58f4e479-001c-daed-0e8d-a60375886602/Against+Hate+-hankkeen+suositukset+-+FI.pdf.

Oksanen, Atte, and Päivi Korpisaari. “Viha ja julkisuus: Päätöksentekoon vaikuttamaan pyrkivä vihapuhe yhteiskunnallisesta ja oikeudellisesta näkökulmasta.” Research plan, October 2019.

PEN America. “Online Harassment Field Manual: Best Practices for Employers.” Accessed October 20, 2020. https://onlineharassmentfieldmanual.pen.org/best-practices-for-employers/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Poliisihallitus. ”Lausunto maalittamista koskevaan selvitykseen.” Report, Helsinki, July 24, 2020.

Rauramo, Päivi, Janne Kiiskinen, Tanja Lehtoranta, Kerttuli Harjanne, and Heidi Schrooten. ”Sosiaalisen median työkäyttö: Työsuojelunäkökulma.” Työturvallisuuskeskus. Last modified August 18, 2014. https://tyoturvallisuuskeskus.mobiezine.fi/zine/8/cover.

Rauramo, Päivi. *Työsuojelu ja työhyvinvointi asiantuntija- ja toimistotyössä*. Helsinki: Työturvallisuuskeskus, 2020.

Rintala, Jenny. ”Työn psykososiaaliset kuormitustekijät työturvallisuuslaissa.” In *Työturvallisuusoikeus*, by Johanna Havula, Timo Jarmas, Seppo Koskinen, Anu-Tuija Lehto, Nina Meincke, Jaana Paanetoja, Timo Pehrman, Jenny Rintala, Jan Schugk, Toni Sortti, Heli Tikkanen, Vesa Ullakonoja, and Anne Vänskä, 140-169. Helsinki: Edita, 2018.

Saloheimo, Jorma. *Työturvallisuus: Perusteet, vastuu ja oikeusturva*. 3rd ed. Helsinki: Talentum Pro, 2016.

Siiki, Pertti. *Työturvallisuuslainsäädäntö: Työnantajan ja työntekijän velvollisuudet ja oikeudet*. Helsinki: Edita Publishing Oy, 2002.

Sosiaali- ja terveysministeriö. *Riskien arviointi työpaikalla -työkirja*. Helsinki: Sosiaali- ja terveysministeriö, 2015. https://ttk.fi/tyoturvallisuus_ja_tyosuojelu/tyosuojelu_tyopaikalla/vastuut_ja_velvoitteet/tyon_vaarojen_selvittaminen_ja_arvioinnti.

Suomen Journalistiliitto. ”Lausunto maalittamista koskevaan selvitykseen.” Report, Helsinki, August 24, 2020.

Suomen Lakimiesliitto. ”Lausunto maalittamista koskevaan selvitykseen.” Report, Helsinki, August 8, 2020.

Suomen Syyttäjäyhdistys ry. ”Lausuma maalittamista koskevassa asiassa.” Report, Helsinki, August 6, 2020.

Suomen tuomariliitto. ”Lausuma maalittamista koskevassa asiassa.” Report, Tampere, August 14, 2020.

Tieteentekijöiden liitto. ”Lausuma koskien Valtioneuvoston kanslian ns. maalittamista koskevaa selvityspyyntöä.” Report, Helsinki, August 31, 2020.

Valdés de la Vega, Berta. “Occupational Health and Safety: An EU Law Perspective.” In *Health and Safety at Work: European and Comparative Perspective*, edited by Edoardo Ales, 1-27. The Netherlands: Kluwer Law International BV, 2013.

Walters, David. “The Framework Directive.” In *Regulating Health and Safety Management in the European Union: A study of the Dynamics of Change*, edited by David Walters, 39-57. Brussels: P.I.E.-Peter Lang, 2002.

Woods, Lorna, and William Perrin. *Online harm reduction: a statutory duty of care and regulator*. Dunfermline: Carnegie UK Trust, 2019. https://www.carnegieuktrust.org.uk/publications/online-harm-reduction-a-statutory-duty-of-care-and-regulator/.

Yhdenvertaisuusvaltuutettu. ”Lausunto maalittamista koskevaan selitykseen.” Report, Helsinki, September 2, 2020.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Combating Disinformation

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Platform (un)accountability. Reviewing Platform Responses to the Global Disinfodemic One Year Onward

Trisha Meyer, Alexandre Alaphilippe

Abstract: This chapter compares Facebook, Google, TikTok and Twitter's responses to COVID-19 and US elections-related disinformation in 2020, furthering our understanding of often opaque moderation practices. Most prominently, online platforms heavily emphasized amplification of credible information, including through provision of free advertising space. They also rapidly and regularly expanded their policies in order to ban, remove, demote or label disinformation as harmful but not illegal. In 2020, the editorial role of online platforms became visible as never before. Their ability to react quickly is both encouraging and worrying, if not accompanied by a known hierarchy of principles and stringent transparency and review measures.

Keywords: content moderation; platform power; COVID-19; US 2020 Elections; disinformation; Facebook; Google; TikTok; Twitter

1. Introduction

One year ago, the COVID-19 virus brought economies and societies to a screeching halt. A global health pandemic ensued. One year later, as vaccinations roll out, we hope for return to a 'new normal', with a renewed appreciation of the need for social connection in our lives. In our isolation, community has proven more important than ever.[1]

Parallel to the spread of the virus has been the spread of disinformation, which Posetti, Bontcheva et.al. describe as a 'disinfodemic' in their ITU/UNESCO study on balancing responses to disinformation with freedom

1 Jonathan Sacks, *Morality: Restoring the Common Good in Divided Times* (New York: Basic Books, 2020).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of expression, media and information literacy, and critical independent journalism.[2]

During this health and information pandemic, online platforms are under intense scrutiny to tackle the disinfodemic rampant on their services. Their terms of service, community guidelines, as well as national legislation, seek to dissuade users from posting illegal content – and increasingly, too, more broadly and vaguely, harmful content.[3] Attention for platforms' powerful intermediary role in online speech precedes 2020, but the pressure on them to 'clean up' their services is unprecedented.

In this chapter[4], we take a close look at how online platforms have responded to health and political disinformation in 2020. We publish detailed comparative timelines of Facebook, Google, TikTok and Twitter's responses to COVID-19 and US general election-related disinformation, in an effort to further our understanding of their content moderation practices. We start by providing a brief sketch of the policy and theoretical context in which these platform responses take place. The editorial role of platforms has become undeniable but is currently largely unregulated. We also explain our methodology and provide details on the dataset we are making publicly available, followed by our comparative analysis of responses by four platforms to the global disinfodemic in 2020.

We conclude that online platforms heavily emphasized amplification of credible COVID-19 related information of the World Health Organization (WHO) and other public health authorities, including through provision of free advertising space. Platforms even launched their own content initiatives, most prominently visible through information panels, but Facebook also livestreamed interviews with leading health professionals and TikTok co-produced media and information literacy videos.

---

2 Julie Posetti and Kalina Bontcheva, 'Disinfodemic: Deciphering COVID-19 Disinformation. Policy Brief 1' (Paris: United Nations Educational, Scientific and Cultural Organization, 2020); Kalina Bontcheva et al., 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression' (Geneva and Paris: International Telecommunication Union and United Nations Educational, Scientific and Cultural Organization, 2020).

3 David Kaye, *Speech Police: The Global Struggle to Govern the Internet* (New York: Columbia Global Reports, 2019).

4 A short version of this chapter appeared on the EU Disinfolab blog in February 2021, Trisha Meyer and Alexandre Alaphilippe, 'One Year Onward: Platform Responses to COVID-19 and US Elections Disinformation in Review', 24 February 2021, https://www.disinfo.eu/publications/one-year-onward-platform-responses-to-covid-19-and-us-elections-disinformation-in-review/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Online platforms also rapidly and regularly expanded their policies, especially on Misleading and Harmful Content, in order to ban, remove, demote or label disinformation harmful but not illegal. Facebook and Google in particular use their advertising policy to aggressively pre-screen paid content on their platforms, and Twitter experimented with additional 'friction',[5] slowing down users' reactions through prompts when users sought to share misleading content during the US elections.

In 2020, the role of online platforms in content moderation became visible as never before. We argue that their ability to react quickly is both encouraging and worrying, if not accompanied by a known hierarchy of principles and stringent transparency and review measures.

2. *Policy and theoretical context*

The legal underpinnings to the current approach to platform regulation find their origins in internet legislation of the late 1990s and early 2000s. Section 230 of the Communication Decency Act (and other sectoral legislation, such as the Digital Millennium Copyright Act) in the United States and the E-Commerce Directive in the European Union were among the first laws granting internet intermediaries limited liability when content generated by their users infringed local intellectual property, speech or security laws.[6] Online platforms did not exist in the same shape or on the same scale as they do now, but the general recognition was that the internet would not be able to grow and flourish if internet intermediaries, whatever shape they took, had to worry all that much about how users were using their services.[7] Only when illegal content and behaviour come to the attention of intermediaries does action need to be taken to remove

5 Ezra Klein, 'The Case for Slowing Everything Down A Bit', *Vox*, 19 November 2018, https://www.vox.com/technology/2018/11/19/18101274/google-alphabet-facebook-twitter-addiction-speed.

6 European Parliament and Council of the European Union, 'Directive 2000/31/EC of the European Parliament and of the Council of the European Union on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market' (2000); US Copyright Office, 'Digital Millennium Copyright Act. Pub. L. No. 105-304, 112 Stat. 2860' (1998); US Congress, 'Communication Decency Act (Title V of the Telecommunications Act)' (1996).

7 Ian Brown and Christopher T. Marsden, *Regulating Code: Good Governance and Better Regulation in the Information Age* (Cambridge, Mass: MIT Press, 2013); Roger Brownsword, *Rights, Regulation, and the Technological Revolution* (Oxford and New York: Oxford University Press, 2008).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

egregious content. However, it is clear that this provides a strong incentive to be content-agnostic and play the ignorance card.[8]

This early approach of limited liability fit into a context of a tech-optimism that the internet would bring positive and empowering societal change. It is also a regulatory mirror of the internet's main architectural end-to-end principle to keep the core as efficient and flexible as possible.[9] However, by 2013, the mood towards internet intermediaries started to shift, most notably with Snowden's revelations of mass government surveillance facilitated by telecoms companies.[10] Then, in 2016, fear of undue (foreign) influence in the US general elections and the UK Brexit referendum fanned the flames further. In 2018, the turn towards tech-pessimism was complete when the Facebook-Cambridge Analytica scandal finally came to light. The temptation to harvest user data and getting in front of the influence curve, whether for political or economic gain, proved greater than the early internet rules could curtail.[11]

The result has been a flurry of regulatory inquiries and proposals to curb the excesses of platform power.[12] Some focus on platforms' economic power and consider vigorous application of competition law or new tax rules the way forward; others target the political power gained through micro-targeting and advertising; some still recognize the need to support

8 Heidi Tworek, 'Social Media Platforms and the Upside of Ignorance', *Centre for International Governance Innovation*, 9 September 2019, https://www.cigionline.org/articles/social-media-platforms-and-upside-ignorance.

9 Janet Abbate, *Inventing the Internet* (London & Cambridge, MA: MIT Press, 1999); Brian Carpenter, 'RFC1958. Architectural Principles of the Internet' (Online: Internet Architecture Board, June 1996), https://tools.ietf.org/html/rfc1958.

10 Zygmunt Bauman et al., 'After Snowden: Rethinking the Impact of Surveillance', *International Political Sociology* 8, no. 2 (1 June 2014): 121–44, https://doi.org/10.1111/ips.12048; Julia Pohle and Leo Van Audenhove, 'Post-Snowden Internet Policy: Between Public Outrage, Resistance and Policy Change', *Media and Communication* 5, no. 1 (2017): 1–6, http://dx.doi.org/10.17645/mac.v5i1.932.

11 Robin Mansell, *Imagining the Internet: Communication, Innovation and Governance* (Oxford: Oxford University Press, 2012); Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019).

12 On the rise of platform governance, see Robert Gorwa, 'What Is Platform Governance?', *Information, Communication & Society* 22, no. 6 (12 May 2019): 854–71, https://doi.org/10.1080/1369118X.2019.1573914. For a comparative overview of currently proposed platform regulation, see Freddy Mayhew, 'Regulating Facebook and Google: The Growing Global Big Tech Backlash', *Press Gazette*, 18 February 2021, sec. News, https://www.pressgazette.co.uk/regulating-facebook-google/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

journalism and media and information literacy. Worrying from our perspective is the desire of some regulators to do away with the mostly content-agnostic approach of internet intermediaries. Although the intention to protect vulnerable groups is well-grounded, so was the discomfort felt at the banning of US President Trump by Twitter and Facebook.[13] Private companies decided on their own terms where acceptable speech ends. To be clear, they do this all the time.[14]

Despite this chapter's focus on online platforms, we would like to broaden our horizon momentarily. With the advent of each new technology, we both herald and cower at its invention, but tend to overplay its impact and downplay our agency to chart its course.[15] Importantly, in this chapter we should avoid confusing cause and effect. Recommender systems and algorithms on online platforms exasperate but are not the cause of digital disinformation. In addition, technology ('code' as Lawrence Lessig[16] calls it) is only one of several means of regulating such societal problems. It is powerful but should be considered alongside other approaches (which Lessig divides into law, market, and norms). Platforms are not off the hook, but a comprehensive approach is needed.

Indeed, there is reason for concern at outsourcing speech control to online platforms. This should not be in hands of private corporations, especially when they are largely unaccountable and the explainability of their decision-making leaves much to be desired.[17] The emphasis should therefore not be on expanding content moderation from illegal to harmful content, but rather on creating transparency in the process of content

13 Alex Hern, 'Opinion Divided over Trump's Ban from Social Media', *The Guardian*, 11 January 2021, https://www.theguardian.com/us-news/2021/jan/11/opinion-divided-over-trump-being-banned-from-social-media.

14 Judit Bayer, 'Between Anarchy and Censorship. Public Discourse and the Duties of Social Media', CEPS Paper in Liberty and Security in Europe No. 2019-03 (Online: CEPS, May 2019), https://www.ceps.eu/ceps-publications/between-anarchy-and-censorship/.

15 Andrew Feenberg, *Questioning Technology* (London & New York: Routledge, 1999); Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (New Haven: Yale University Press, 2018).

16 Lawrence Lessig, *Code: And Other Laws of Cyberspace, Version 2.0* (New York: Basic Books, 2006).

17 Kaye, *Speech Police*. For an academic and civil society discussion on minimum standards for content moderation, see ACLU Foundation of Northern California et al., 'Santa Clara Principles on Transparency and Accountability in Content Moderation', Santa Clara Principles, accessed 15 March 2021, https://santaclaraprinciples.org/images/scp-og.png.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

moderation. Their editorial role is evident, but how decisions are made currently is not. We will return to these thoughts at the end of the chapter.

3. Methodology and dataset

For this chapter, we reconstructed a timeline of responses of Facebook, Google, TikTok and Twitter, on the basis of reports submitted to the European Commission as part of the Fighting COVID-19 Disinformation Monitoring Programme, as well as updates posted on their company blogs.[18] It is important to note that we analysed what platforms announced and reported, not whether these measures were implemented.

We mapped their responses by month and against the platform disinformation response typology we developed as part of our contribution to the UNESCO/ITU Balancing Act study mentioned in the introduction.[19] In particular, we divide platform responses into four types of 'content' mod-

18 We used the sources below to map the platform responses on a month-by-month basis. This was not always a straightforward exercise, and we would be very happy to rectify any error you may spot!
We did not include company updates related to support for health workers, small businesses, non-profits, children, social movements, communities, mental health, emotional well-being or diversity, as these were not specific to combating disinformation on the platforms.
All: monthly platform reports from August 2020 for the European Commission Fighting COVID-19 Disinformation Monitoring Programme, https://ec.europa.eu/digital-single-market/en/news/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme
Facebook: Facebook Coronavirus Newsroom updates, https://about.fb.com/news/2020/12/coronavirus/; Facebook US 2020 Elections report, https://about.fb.com/actions/preparing-for-elections-on-facebook/; Facebook Key Elections Investments and Improvements timeline, https://about.fb.com/wp-content/uploads/2020/12/Elections-Investments-and-Improvements.pdf
Google: Google Keyword COVID-19 updates, https://blog.google/inside-google/covid-19/; Elections Google updates, https://elections.google/ - updates
TikTok: TikTok Safety Center – COVID-19, https://www.tiktok.com/safety/resources/covid-19?lang=en&appLaunch=; TikTok Safety updates, https://newsroom.tiktok.com/en-us/safety; TikTok Integrity for the US Elections, https://www.tiktok.com/safety/resources/2020-us-elections
Twitter: Twitter Blog, https://blog.twitter.com/; Twitter Coronavirus updates, https://blog.twitter.com/en_us/topics/company/2020/covid-19.html; Twitter Blog Elections tag, https://blog.twitter.com/en_us/tags.blog–elections.html

19 Bontcheva et al., 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression'.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

eration: flagging/labelling, blocking/removing, limiting/demoting and prioritizing/amplifying; and four types of 'other' moderation: specific to accounts, advertising, users, and research/review.

The aim of this breakdown into different types of 'moderation responses' is to map online platforms' change in emphasis over time in a granular fashion. Each is a different manifestation of the editorial role that platforms play in moderating online speech.

Table 1 below shows our mapping for Facebook's responses to COVID-19 and US election disinformation in 2020 as an illustration (this is only 1/8th of the dataset). We cordialy invite you to consult the complete dataset online[20]. In this online resource, we publish two timelines, with the data organized by platform and by response type.

20 Trisha Meyer, 'Comparative Timeline of Platform Responses to COVID-19 and US Elections Disinformation, Organised by Platform and by Response (Updated Regularly)', Google Sheets, 18 February 2021, https://bit.ly/3ySbJXc.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Table 1. Facebook responses to COVID-19 and US elections related disinformation in 2020 (own compilation)

FACEBOOK	Continuous	Jan	Feb	March	April	May	June	July	Aug	Sep	Oct	Nov	Dec
Flagging / labelling content	*gradually expands fact-checking collaborations; increases use of AI to detect misinformation and deep fakes*			commits to invest $2M additional funding for fact-checking, news and research on misinformation; content moderators are sent home	provides educational pop-ups for users who reacted to harmful COVID-19 misinformation; content moderators partially return to office		starts labelling of state-controlled media	starts adding labels to posts on voting, including federal officials and candidates		launches second wave of funding for fact-checking and news; announces stronger labels on posts from candidates and campaigns that prematurely claim victory or attempts to delegitimize the elections	promotes awareness campaigns of public health agencies on vaccines		

https://doi.org/10.5771/9783748929799 https://www.inlibra.com/de/agb#agb - Open Access - (cc) BY

Blocking / removing content	announces policy on removal of misleading manipulated media	removes COVID-19 disinformation "with immiminent physical harm" on Facebook and Instagram	removes COVID-19 information from recommendations "unless posted by a credible health organization" (Instagram)		announces stricter policy on content that interferes with or suppresses voting		removes health groups from recommendations; announces stronger voter restriction policies	bans calls for poll watching when combined with militarized language		removes debunked false claims on vaccines
Limiting / demoting content			demotes and issues warnings on debunked COVID-19 disinformation that falls within community guidelines		demotes and issues warnings on debunked voting disinformation that falls within community guidelines		sets forwarding limit on Messenger		redirects searches for terms related to Qanon on Facebook and Instagram to credible resources of GNET	
Prioritizing / am-		displays educational	launches COVID-19 Edu-	expands educational	launches Voting Informa-	global reminder to wear				

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

		pop-ups of WHO and public health officials in search on Facebook and Instagram	cational Centre, in collaboration with WHO, on news feed (Facebook); shows information of WHO and public health agencies at top of feed and adds stickers to promote accurate information (Instagram); launches information hub and WHO Health Alert	pop-ups to Groups, prompts group admins to share live broadcasts from public health officials; partners with DCD to develop COVID-19 curriculum for group admins	tion Centre; starts prioritizing original news reporting	face coverings, adds facts about COVID-19 to information centre	

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

		(WhatsApp); launches Messenger Coronavirus Community Hub (Messenger)		starts verifying identity of people with high-reach posts on Facebook in the USA			expands Dangerous Individuals and Organizations Policy to combat groups tied to violence (such as QAnon)	reinforces User Recidivism Policy		updates enforcement against QAnon, removing any Facebook pages, groups and Instagram accounts representing Qanon, even if they contain no violent content	
Account-specific responses	*continues action and collaboration against Coordinated Inauthentic Activity*											

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Ad-specific responses	*continues use of Ads Library*	expands required authentication of advertisers for political/social issue ads	bans ads promoting COVID-19 cure	provides free advertising space to WHO and public health agencies for awareness raising on COVID-19; temporarily bans advertisements and commerce listings for COVID-19 related products			adds US House and Senate ad tracker to Ad Library; announces stricter policy on ads that interfere with or suppress voting; allows promotion and trade in non-medical masks	allows promotion and sale of hand sanitizers and surface disinfection wipes	announces restrictions on new political/social issue ads in final week of campaign; announces ban on ads that prematurely claim victory or attempt to delegitimize the elections	bans ads discouraging vaccines; announces temporary suspension of social issue, electoral and political ads after polls close	blocks creation of new ads about social issues, elections or politics immediately before election day, temporarily suspends ads about social issues, elections or politics after election day	updates ad policy for COVID-19 vaccines, e.g. allowing ads on safe access to vaccine, but prohibiting sale or expedited access
User-specific responses				launches digital literacy programme (Get Digital); live inter-	live interview between Mr Zuckerberg, Ms Chan		adds new control that allows people to see fewer political/	live interview between Mr Zuckerberg and Dr Fauci			live interview between Mr Zuckerberg and Dr Fauci	

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

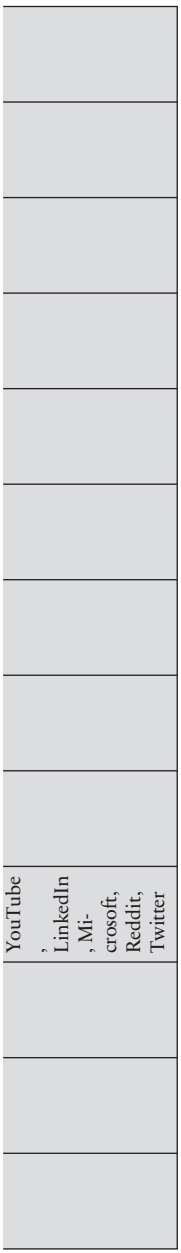
		view between Mr Zuckerberg and Dr Fauci	and Dr Frieden		social issue ads on Facebook and Instagram; launches media literacy campaign (Stamp Out False News)		launches Presidential Election Research Initiative			
Review / research-specific responses			connects health organisations and developer partners to scale response to COVID-19; supports COVID-19 developer hackathon; content	announces new Data for Good tools to help health researchers track and combat COVID-19; content moderators partially re-	releases new visualisations, data sets and new survey to combat COVID-19					

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

			moderators are sent home, increases reliance on automated technology	turn to office							
		launches Election Operations Centre for primary elections in the USA	commits to invest $100M in news industry; $1M through local news covering COVID-19 in USA and Canada; joint announcement on COVID-19 coordination from Facebook, Google,								
increases use of AI to detect misinformation and deep fakes											

Other

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA



https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

4. Platform-specific responses

In this section, we highlight Facebook, Google, TikTok and Twitter's main responses to COVID-19 and US general election-related disinformation in 2020 as a basis for our comparison.

Facebook (Facebook, Messenger, Instagram, Whatsapp)

As our timeline shows, Facebook was busy, with a frenzy of activity in February and March 2020 as the COVID-19 pandemic broke out; a steep ramping up of US election-related activities as of June; and a gradual response in preparation for the COVID-19 vaccine rollout as of September. Their COVID-19 response emphasizes the *prioritization* of authoritative content, *free advertising* for public health agencies and *demotion* of debunked information. They also remove COVID-19 related disinformation with 'imminent physical harm'. Meanwhile Facebook's US election response focused on policies related to *political and social issues ads* and policies that allow for *removal* of content that interfered with or suppressed voting. They also added *warning messages* to debunked content.

Google (Search, YouTube, AdSense)

Google's COVID-19 response was gradual, starting with *prioritization* and amplification of accurate COVID-19 related content and free *advertising* credits for public health authorities. Notably they also published a COVID-19 Medical Misinformation policy and expanded their Harmful Health Claims policy to *remove* content that contradicts authoritative and scientific consensus on the health crisis. Google's US election response focused on *security* and *amplification* of trusted news. It is important to note that *advertisement*-related policies are a powerful tool for Facebook and Google to wield. Both Facebook and Google temporarily paused US election ads after the polls closed.

TikTok

TikTok's COVID-19 response started earlier than other platforms and was concentrated in time (Jan-March). A similar approach was followed for

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

vaccines in December. It stresses information *prioritization* and amplification through in-app notices, stickers, and brand takeovers. In October TikTok launched Project Halo, a science *communication* effort, to raise awareness and confidence in vaccine. TikTok's US election response was similarly concentrated in time (Aug-Oct) and focused on an in-app guide and *public service announcements*. During the month of October until the end of Election Day, they provided daily updates on their election response. TikTok does not allow political ads. Similar to Facebook and Google, it donated *ad space* to public health authorities. In February 2021, TikTok announced that they will add *friction* to their disinformation response arsenal. When they identify a video with unsubstantiated claims, Tiktok will show a banner *warning* and include several warning prompts before viewers share a flagged video.

Twitter

In 2020 Twitter played an extensive editorial role on its platform, through use of *labels, warnings, removal, reducing visibility, adding friction, promoting authoritative content*. As part of its COVID-19 response, Twitter broadened its policy definition of harm to include content that contradicts COVID-19 public health guidance. In February and May, they also issued guidance on their staged approach to manipulated and synthetic media and potentially misleading content. A frenzy of activity occurred in the lead up to and aftermath of US elections on *content and account* level. In December, Twitter reported that their more extensive version of friction (Quote Tweet rather than Retweet; removing 'liked by' and 'followed by' recommendations, only surfacing 'additional context' trends) did not bear expected results.

5. Comparison and key take-aways

Table 2. Comparison of platform responses to COVID-19 and US election-related disinformation (own compilation) [main responses related to C = COVID-19; E = US elections]

Main response type per platform	Facebook	Google	TikTok	Twitter
Flagging/labelling content	E		C	C / E
Blocking/removing content	C/ E	C	C / E	C / E

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Limiting/demoting content	C / E		E	E
Prioritizing/amplifying content	C	C / E	C / E	C / E
Account-specific	E			E
Advertising-specific	C / E	C / E	C	C
User-specific			C	
Review/disinformation research-specific				

The timing of online platform responses to COVID-19 corresponds with the arrival of the virus in Europe and North America in March 2020 – despite having users globally. TikTok, the only non-Western (Chinese) social media company in our sample, is an exception. Its COVID-19 response ramps up in February 2020. Similarly striking is the platforms' response to US election disinformation. Platforms have come a long way since the 2016 US general elections and UK Brexit referendum. They have been prompted by governments to keep records of political ad spending, to mitigate foreign interference, to ensure fair and free elections. Yet the actions taken by platforms in the 2020 US elections were unprecedented. In particular, in a span of only a few months, labelling and removing political speech and figures became normalized.

As the pandemic hit, we saw online platforms heavily *prioritize authoritative content* provided by public health officials through in-app notices, educational pop-ups and prompts, launching dedicated hashtags and educational centres, and surfacing credible public health information at the top of feeds and in COVID-19 related searches. Six months later, similar action to emphasize authoritative content was taken in preparation for the US elections, and towards the end of 2020 to counter vaccine disinformation. One relatively novel development were the grants of *free advertising credits* to the World Health Organization (WHO) and public health authorities. Google, Facebook and Twitter also provided large grants for journalism and fact-checking.

In parallel – and prominently used during the run-up to and in the aftermath of the US general elections – platforms regularly updated their *policies* related to Misleading and Harmful Content, Sensitive Events, Civic Integrity to *ban, remove or demote content and ads* that contradicted public COVID-19 health guidance and undermined confidence in the elections. Efforts to counter QAnon led Facebook to expand its Dangerous Individuals and Organizations Policy to include organizations tied to violence in August 2020. Much later, in January 2021, Twitter updated its Coordi-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

nated Harmful Activity Policy. Infamously, both platforms permanently suspended President Trump's accounts in January 2021 for inciting the violence at the Capitol Hill riots.

In 2020, platforms also extended their use of warning messages and stickers to *label and flag* potentially misleading content, caution to share further and point to credible information. In February 2020, Twitter started taking action against synthetic and manipulated media, and in May against potentially misleading COVID-19, election- (and vaccine-) related content (see Figure 1 below). Slightly later, in June 2020, Facebook started labelling state-controlled media, and in July the accounts of political candidates and federal officials.

Figure 1: Twitter's approach to misleading content (May 2020)[21]

Misleading Information	Label	Removal
Disputed Claim	Label	Warning
Unverified Claim	No action	No action*
	Moderate	Severe
Propensity for Harm		

6. *Conclusion*

In response to game changer events such as the global health pandemic or the use of disinformation by US President Trump, online platforms took unprecedented measures to minimize harm by improving their content moderation efforts. Some policy updates were clearly planned, such as Twitter's graduated response to synthetic and manipulated media, while others were kneejerk responses to ongoing events. This rapid expansion of

21 Yoel Roth and Nick Pickles, 'Updating Our Approach to Misleading Information', *Twitter Blog* (blog), 11 May 2020, https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

disinformation policies culminated in bans of President Trump and Parler on multiple platforms in January 2021.

This ability to react quickly is both encouraging and worrying. Encouraging, because it demonstrates that platforms, under societal pressure, behave as a public interest utility in specific cases. Yet at the same time, it is worrying as the majority of measures taken fail to address the root causes of the architecture of information distribution. Without this, discussions around censorship and its abuses will prevail over the work needed to build a more inclusive information ecosystem.

The emphasis of current regulatory discussions on platforms needs to be on accountability. Our analysis was based on information we were able to gather from company blogs and reports submitted to the European Commission. While their reporting is a step forward, this information only became comparable data with significant additional effort but it does not offer insights into the implementation or consequences of action taken.

We need detailed metrics on online content distribution. Crucially this should include transparency in terms of content promotion/demotion in addition to removal of content, as an initial means of auditing algorithms. The online advertising ecosystem is also deserving of reinforced scrutiny to gain a better understanding of the impact of changes in ad policies. In light of platforms' emphasis on granting advertising credit, it seems appropriate to establish a register of beneficiaries of ad-credits detailing amounts granted and spent. Civil society (academics, researchers, journalists, civil society organizations) and independent regulators should also be empowered in their role of enforcing accountability of online platforms. The stick behind the door might need to be available to sanction bad faith actors, especially when there are repeated efforts to escape transparency and accountability.

Finally, to return to the opening paragraph of the chapter, regulating platforms will be in vain if we do not tackle the causes of the disinfodemic at the same time. This requires rebuilding trust by listening to others, celebrating our differences, and committing to common objectives. If 2020 can teach us anything, we hope it is that our social bonds and communities are more resilient than we perhaps thought yet also require continual collective and individual commitment.

Bibliography

Abbate, Janet. *Inventing the Internet*. London & Cambridge, MA: MIT Press, 1999.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ACLU Foundation of Northern California, Center for Democracy & Technology, Electronic Frontier Foundation, New America's Open Technology Institute, Irina Raicu, Nicolas Tuzor, Sarah Myers West, and Sarah T. Roberts. 'Santa Clara Principles on Transparency and Accountability in Content Moderation'. Santa Clara Principles. Accessed 15 March 2021. https://santaclaraprinciples.org/images/scp-og.png.

Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon, and R. B. J. Walker. 'After Snowden: Rethinking the Impact of Surveillance'. *International Political Sociology* 8, no. 2 (1 June 2014): 121–44. https://doi.org/10.1111/ips.12048.

Bayer, Judit. 'Between Anarchy and Censorship. Public Discourse and the Duties of Social Media'. CEPS Paper in Liberty and Security in Europe No. 2019-03. Online: CEPS, May 2019. https://www.ceps.eu/ceps-publications/between-anarchy-and-censorship/.

Bontcheva, Kalina, Julie Posetti, Denis Teyssou, Trisha Meyer, Sam Gregory, Clara Hanot, and Diana Maynard. 'Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression'. Geneva and Paris: International Telecommunication Union and United Nations Educational, Scientific and Cultural Organization, 2020.

Brown, Ian, and Christopher T. Marsden. *Regulating Code: Good Governance and Better Regulation in the Information Age*. Cambridge, Mass: MIT Press, 2013.

Brownsword, Roger. *Rights, Regulation, and the Technological Revolution*. Oxford and New York: Oxford University Press, 2008.

Carpenter, Brian. 'RFC1958. Architectural Principles of the Internet'. Online: Internet Architecture Board, June 1996. https://tools.ietf.org/html/rfc1958.

European Parliament and Council of the European Union. Directive 2000/31/EC of the European Parliament and of the Council of the European Union on Certain Legal Aspects of Information Society Services, in particular Electronic Commerce, in the Internal Market (2000).

Feenberg, Andrew. *Questioning Technology*. London & New York: Routledge, 1999.

Gillespie, Tarleton. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven: Yale University Press, 2018.

Gorwa, Robert. 'What Is Platform Governance?' *Information, Communication & Society* 22, no. 6 (12 May 2019): 854–71. https://doi.org/10.1080/1369118X.2019.1573914.

Hern, Alex. 'Opinion Divided over Trump's Ban from Social Media'. *The Guardian*, 11 January 2021. https://www.theguardian.com/us-news/2021/jan/11/opinion-divided-over-trump-being-banned-from-social-media.

Kaye, David. *Speech Police: The Global Struggle to Govern the Internet*. New York: Columbia Global Reports, 2019.

Klein, Ezra. 'The Case for Slowing Everything Down A Bit'. *Vox*, 19 November 2018. https://www.vox.com/technology/2018/11/19/18101274/google-alphabet-facebook-twitter-addiction-speed.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Lessig, Lawrence. *Code: And Other Laws of Cyberspace, Version 2.0*. New York: Basic Books, 2006.

Mansell, Robin. *Imagining the Internet: Communication, Innovation and Governance*. Oxford: Oxford University Press, 2012.

Mayhew, Freddy. 'Regulating Facebook and Google: The Growing Global Big Tech Backlash'. *Press Gazette*, 18 February 2021, sec. News. https://www.pressgazette.co.uk/regulating-facebook-google/.

Meyer, Trisha. 'Comparative Timeline of Platform Responses to COVID-19 and US Elections Disinformation, Organised by Platform and by Response (Updated Regularly)'. Google Sheets, 18 February 2021. https://docs.google.com/spreadsheets/d/1KR-YECATOyEHy_jd1pWXsujuHeL8sE6WWZpk7Ib8LSM/edit?usp=sharing&usp=embed_facebook.

Meyer, Trisha, and Alexandre Alaphilippe. 'One Year Onward: Platform Responses to COVID-19 and US Elections Disinformation in Review', 24 February 2021. https://www.disinfo.eu/publications/one-year-onward-platform-responses-to-covid-19-and-us-elections-disinformation-in-review/.

Pohle, Julia, and Leo Van Audenhove. 'Post-Snowden Internet Policy: Between Public Outrage, Resistance and Policy Change'. *Media and Communication* 5, no. 1 (2017): 1–6. http://dx.doi.org/10.17645/mac.v5i1.932.

Posetti, Julie, and Kalina Bontcheva. 'Disinfodemic: Deciphering COVID-19 Disinformation. Policy Brief 1'. Paris: United Nations Educational, Scientific and Cultural Organization, 2020.

Roth, Yoel, and Nick Pickles. 'Updating Our Approach to Misleading Information'. *Twitter Blog* (blog), 11 May 2020. https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html.

Sacks, Jonathan. *Morality: Restoring the Common Good in Divided Times*. New York: Basic Books, 2020.

Tworek, Heidi. 'Social Media Platforms and the Upside of Ignorance'. *Centre for International Governance Innovation*, 9 September 2019. https://www.cigionline.org/articles/social-media-platforms-and-upside-ignorance.

US Congress. Communication Decency Act (Title V of the Telecommunications Act) (1996).

US Copyright Office. Digital Millennium Copyright Act. Pub. L. No. 105-304, 112 Stat. 2860 (1998).

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Disinformation in the Perspective of Media Pluralism in Europe – the role of platforms[1]

Elda Brogi, Konrad Bleyer-Simon

Abstract: Media freedom and media pluralism are recognised as pillars of contemporary democracies. Technological advancements have not only created new opportunities to boost media freedom and media plurality, but prompted new sources of risks. One of them is the scale and the impact of disinformation on public opinion. Although not illegal, it may pose a growing threat, for instance, to the integrity of elections, including manipulation, as well as to efforts to respond to the COVID-19 pandemic. This paper aims to provide an overview of problems and opportunities, based on experiences from projects the authors are involved in. It provides a brief overview of multi-country data coming from country experts of the EU-wide Media Pluralism Monitor data collection, and describes the challenges and opportunities of European measures to fight disinformation, based on the work of the European Digital Media Observatory.

Keywords: disinformation, misinformation, Code of Practice on Disinformation, European Union, EU, platform regulation, EDMO, Media Pluralism Monitor

Chapter 1. Introduction

In the past years, disinformation has become another challenging issue in content moderation online, next to hate speech; although it is often not illegal, it can cause public harm. This led to the rethinking and re-interpretation of the rationale of the liability exemption for online platforms, as well as the governance of the digital environment.

The policy discussion is ongoing at both the EU and the member state level. In Germany, the Bundestag passed the so-called Network Enforcement Act (NetzDG, also referred to as the Facebook Act) in 2017, which

1 The opinions and views expressed in this chapter are those of the Authors.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

required social media providers to proactively remove certain types of criminal content. This requirement was criticised early on by civil society for possibly damaging freedom of expression and freedom of the press. Many other EU member states decided to also react to the problem in one way or another in their national regulatory systems. In the EU, a key policy tool is the Code of Practice on Disinformation (the Code), which brought together online platforms, such as Google, Facebook, Twitter and TikTok, as well as a number of other stakeholders in an initial effort to fight disinformation in the context of a self-regulatory framework, applying in the framework of existing laws, including the e-Commerce Directive 2000/31/EC, with specific reference to articles 12 to 15 on exemption of liability.[2]

While both the EU members and the Commission took interesting steps in defining a policy against disinformation, for now, there has been limited concrete progress in defining an effective governance strategy. This is in part due to a lack of understanding of the criteria used by online platforms in their content moderation and the design of their recommendation systems. In addition, considering the trends in EU regulation, there is no defined methodology to assess how and with which consequences the platforms act in order to limit the spread of disinformation.

In this paper, the Authors reflect on the measures taken in order to tackle disinformation at EU level, starting from the results of the Media Pluralism Monitor project, and following with the early research carried out under the newly established European Digital Media Observatory (EDMO). The latter multi-disciplinary project brings together fact-checkers, media literacy experts and academic researchers with the aim of understanding and analysing the disinformation phenomenon. The Authors' contribution in EDMO focuses on research and policy analysis on disinformation, including suggestions on a methodology to assess to what extent the Code's implementation impacts the overall disinformation phenomenon. The Authors describe the current shortcomings of the Code, which have a lot to do with the text's lack of detailed practical guidance for its signatories. To help overcome this deficit, the Commission has announced its intent to transform the Code into a co-regulatory instrument. As already visible in the Commission's guidelines for strengthening the

2 European Commission, "Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market" Directive 2000/31/EC", COM(2020) 825 final, December 15, 2020. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Code, compliance with the commitments would then be assessed with the help of well-defined key performance indicators (KPIs) at service-level, while the overall impact of the Code would be assessed with a set of structural KPIs. Finally, we focus on the issue of "trustworthiness" as a key attribute of online contents and content publishers, which can help guide online platforms in their efforts to improve the health and the plurality of online media landscapes. But here again, current measures are lagging behind, while current initiatives that work on providing trustworthiness indicators for platforms may risk creating unintended side-effects to media pluralism.

Chapter 2. Disinformation and the threat to media pluralism

In light of the challenges posed to European democracy by the spread of disinformation, the European Commission expressed the need for a pan-European response, and in January 2018 established the High Level Expert Group on Fake News.[3] This group was made up of industry representatives, civil society, policy makers and scholars, aiming to provide advice on policy initiatives to tackle the problems of online disinformation on the European level. It produced a report in March of the same year, which recommended a multidimensional approach to increase the transparency of online news, the promotion of media literacy, the development of tools to empower users, to safeguard the diversity and sustainability of the news ecosystem in Europe, as well as to promote research on the issue of disinformation.[4] Ahead of the 2019 European elections, the EU followed up by sponsoring a "European approach"[5] to tackle disinformation. This led to the signing of the Code of Practice on Disinformation (the Code),[6] the first major initiative developed at EU level to fight disinformation.

3 It was later renamed to High Level Expert Group on Fake News and Online Disinformation.

4 High Level Expert Group on Fake News and Online Disinformation, "A multi-dimensional approach to disinformation - Report of the independent High level Group on fake news and online disinformation", European Commission, 2018, https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation.

5 European Commission, "Tackling online disinformation: a European Approach" COM/2018/236 final, 2018. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236.

6 European Commission, "Code of Practice on Disinformation", 2018. https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

It followed the Expert Group's recommendations and encouraged online platforms to self-regulate, ensure the transparency of political advertising and restrict the automated spread of disinformation in the European Economic Area.

In parallel with these developments, the Media Pluralism Monitor[7], an EU-wide data collection administered by the Centre for Media Pluralism and Media Freedom (CMPF) at the European University Institute (EUI), has tried to measure the risks for media pluralism stemming from disinformation. Overall, the findings of these assessments show that disinformation is seen as a serious threat all over the EU, but the debate around disinformation and its regulation is still in its early phases; there is a need to find a common language and common policies, compliant with the rule of law.

Looking at the years 2018-2019, the CMPF has introduced elements in its questionnaire that investigate the transparency of online political advertising. The variables aim to define the role and the limits of online platforms' activity, as well as the procedures for their accountability when dealing with political content online. Opacity in political advertising is seen as a key enabler of the rapid spread of disinformation. Results of the MPM2020 sub-indicator "rules on political advertising online" show, as expected, considering the novelty of this debate, that in 25 countries, parties and candidates were not fully transparent about the spending and methods they used in their social media campaigns. In 18 countries some issues were noted in relation to the implementation of the Code, with regard to clear labelling and registering of political and issue-based advertising, and in terms of indicating who paid for it. Overall, there is very little regulation of political advertising online, largely due to a lack of understanding of the criteria used by online platforms in content moderation and the design of recommendation systems.[8] Not a lot has changed in the year 2020, when the CMPF asked again its country teams about

7 The Media Pluralism Monitor (MPM) is a tool that was developed by the CMPF to assess the risks for media pluralism in a given country. It is based on the prototype of the MPM that was designed by the 2009 Independent Study on Indicators for Media Pluralism in the Member States – Towards a Risk-Based Approach carried out by KU Leuven, JIBS, CEU, Ernst & Young, and a team of national experts, https://ec.europa.eu/information_society/media_taskforce/doc/pluralism/pfr_report.pdf.

8 Elda Brogi, Roberta Carlini, Iva Nenadic, Pier Luigi Parcu and Mario Viola de Azevedo Cunha, *Monitoring media pluralism in the digital era: Report 2020* (Florence: European University Institute, 2020), https://cadmus.eui.eu/bitstream/handle/1814/67828/MPM2020-PolicyReport.pdf?sequence=5&isAllowed=y, 77-81.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the assessment of the situation. Still 17 EU countries saw problems related to the implementation of the Code and its effectiveness in the national context.[9]

Chapter 3. The Code of Practice on Disinformation

The Code of Practice on Disinformation was the first major initiative developed at EU level to define a policy on disinformation online, within the current EU legislative framework. In its text, disinformation is understood as verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm (in line with the definition provided by the Commission Communication on tackling on-line disinformation, 2018).[10] The text adds that deceptive content is disseminated either for economic gain (monetisation) or with the intent of deceiving the public. It also emphasises the component of "public harm" as disinformation is a threat "to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security".[11] The Commission's guidelines for a new, strengthened version of the Code, which were published on 26. May 2021, just a few days before submitting our revised manuscript, extend this definition to include some forms of misinformation[12] as well, meaning harmful content that is spread unintentionally.[13]

---

9 Elda Brogi et al. *Monitoring media pluralism in the digital era: Report 2021* (Florence: European University Institute, forthcoming 2021).

10 European Commission, *Code of Practice*; European Commission, *Tackling online disinformation*.

11 European Commission, *Code of Practice.*

12 Claire Wardle and Hossein Derakhshan differentiate between three key forms of information disorders: misinformation (when the information is not true, but it is not created and shared with the intent of doing harm), disinformation (when the untrue content was created and shared with the intent of doing harm) and malinformation (when the information is factually true, but it is shared in a way that it can cause harm). See: Claire Wardle and Hossein Derakhshan, "Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information," in *Journalism, 'Fake News' & Disinformation. Handbook for Journalism Education and Training*, eds. Cherilyn Ireton and Julie Posetti, 44-56. Paris: UNESCO, 2018. https://unesdoc.unesco.org/ark:/48223/pf0000265552/PDF/265552eng.pdf.multi

13 "European Commission Guidance on Strengthening the Code of Practice on

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Code provides commitments to the online platforms that sign up to it in five main areas (which are referred to as the five pillars):

A. "Scrutiny of ad placements" is about preventing providers of disinformation from monetising their content ("reduce revenues"), mainly by urging online platforms to exclude them from their advertising services.
B. "Political advertising and issue-based advertising" aims to make sure that political advertisement can be clearly identified by users. For this aim, users should also have an understanding of why they were targeted by a given advertisement.
C. "Integrity of services" refers to two linked issues, fake accounts and automated bots, that play a key role in the spread of disinformation. The text calls for effective efforts to close fake accounts and to publicly issue policies on what "constitutes impermissible use of automated systems".
D. "Empowering consumers" aims at making access to "trustworthy" sources of information easier, by nudging consumers to access sources that are less likely to spread disinformation, as well as by providing them with tools to reliably assess the credibility of sources and content, and easily report those that spread disinformation.
E. "Empowering the research community" affirms signatories' willingness to allow research on disinformation and political advertisement on their platforms, and support efforts to track disinformation by giving some access to "privacy protected datasets" and supporting joint projects.

The Code was first signed by advertisers, the software developer Mozilla, as well as the online platforms Facebook, Google and Twitter in October 2018. Microsoft joined in May 2019, while TikTok signed the Code in June 2020. After agreeing on the Code, signatories have pledged to report on the actions taken in order to further the goals that were identified.[14]

Disinformation" European Commission, 4-5, https://ec.europa.eu/newsroom/dae/redirection/document/76495

14 See: "Annual self-assessment reports of signatories to the Code of Practice on Disinformation 2019," European Commission, https://digital-strategy.ec.europa.eu/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 4. Limited impact

The Code of Practice is an interesting step in defining a policy against disinformation, as its signatories have adhered to self-commitments that currently are not required from them by law, while its implementation and assessment can be read as a pilot test of the Digital Services Act, a 2020 proposal to update the EU's legal framework with a safer digital space in mind.[15] The impact of the Code is nevertheless limited, for the time being. Problems can be traced back to three groups of issues: lack of guidance, limited compliance and the small number of signatories.

First, the Code does not provide detailed practical guidance for its signatories, just a set of vaguely defined commitments that the platforms are expected to achieve with whatever means they see fit, defining, in the end, some potentially good practices. Terms used in the commitments can be either misinterpreted, or they provide grounds for platforms to selectively comply with their obligations. Many of them lack proper definitions. For example, the commitments are aimed at "Relevant Signatories", but who the relevant signatories are is not specified, and platforms thus have the freedom to determine for themselves what commitments they will comply with. In addition, signatories are expected to use "commercially reasonable efforts" – but this criterion is not detailed either. As some of the platforms were financially profiting from the activities of purveyors of disinformation, in their case it is not necessarily "commercially reasonable" to give up revenues. It is against this background that the Digital Services Act (DSA) proposal announced the establishment of a powerful framework for transparency and clear accountability, which enables democratic oversight over online platforms.[16] The Commission's guidance would turn the Code of Practice into a "Code of Conduct" and would only allow signatories to opt out of commitments in case they provide relevant and public justification.

---

15 Elda Brogi and Iva Nenadic, European plan to increase transparency and accountability of the gatekeeper online platforms to protect democracy: EDMO`s role in the Commission's digital policy approach http://www.medialaws.eu/european-plan-to-increase-transparency-and-accountability-of-the-gatekeeper-online-platforms-to-protect-democracy-edmos-role-in-the-commissions-digital-policy-approach/

16 European Commission, "The Digital Services Act package", 2020. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Secondly, the European Regulators Group for Audiovisual Media Services (ERGA)[17] has pointed to problems related to compliance and transparency, as the Code relies on self-reporting, and thus statements of platforms cannot be verified. There is an absence of standards for its evaluation and for reporting, lack of oversight on compliance, lack of sanctions for non-compliance, and lack of data against which to check the statements and reports created by platforms themselves.[18] In fact, ERGA has found in its cooperation with national regulators that the reported achievements of platforms are not as successful as the platforms themselves make them sound.[19] The Commission itself highlighted the most serious deficiencies in the attempts to demonetise purveyors of disinformation. Thus, the new Guidelines ask for a common reporting template and a set of key performance indicators (KPIs) to more effectively measure signatories' compliance.

Thirdly, the number of signatories is small. Although the initial group of signatories was extended, among others by TikTok, there are still many online platforms missing. A debate is open on whether messaging platforms should sign the Code[20], as experience from the past years shows that messaging services such as Messenger, Telegram or WhatsApp are among the amplifiers of the spread of disinformation content.[21] The new guidelines would address this by introducing different reporting requirements for small and large signatories, depending on their market share in Europe; and encouraging private messaging services as well as representatives of the advertising sector to join.

17 ERGA, "ERGA Report on Disinformation: Assessment of the Implementation of the Code of Practice", 2020. https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf.

18 Iva Nenadić, "Unpacking the 'European approach' to tackling challenges of disinformation and political manipulation," *Internet Policy Review* 8, No. 4 (2019): 1-22.

19 ERGA, "ERGA Report on Disinformation", 17-18.

20 Messaging apps are aimed for personal communication, as they entail a sender and a recipient for a private message. Nonetheless, they are currently used to convey content to big groups of recipients and offer easy functions to share messages from one group to another, thus reaching a mass audience.

21 Samuel Woolley, "Encrypted messaging apps are the future of propaganda," *Brookings*, May 1, 2020. https://www.brookings.edu/techstream/encrypted-messaging-apps-are-the-future-of-propaganda/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 5. Some suggestions to address the shortcomings

Continuous monitoring of platforms' compliance and the independent assessment of their activities to limit disinformation are key to the success of Code of Practice and are part of a new regulatory toolbox that is somehow already sketched by the proposal of the DSA and the guideline for the Code 2.0. It is important, therefore that a particular focus is devoted to the way in which an independent assessment of the Code's implementation can be done, looking at what standards to use when evaluating the compliance of the platforms with the Code's obligations, and the effects of the Code implementation, and what kind of governance to foresee in order to create an oversight mechanism that is effective and respectful of the rule of law. This is the focus of the research carried out (amongst many other activities) under the EDMO project. It aims at contributing to the definition of an assessment methodology that includes standards for platforms' reporting that enable the verification of platforms' compliance with the measures taken when implementing the Code. This methodology will be complemented with the definition of indicators that enable assessing the Code's impact in limiting the spread of disinformation ans on the health of the digital information environment.

To enable a comprehensive evaluation, the Code's overall methodology designed encompasses (a) a service-level and (b) a structural assessment. The first assessment looks at platforms' compliance while the second one is interested in the Code's wider impact. What we consider reasonable is to develop and test a methodology that is: inclusive (considering current and potential future signatories of the Code); feasible (capable of being implemented on a regular basis under different forms of regulatory regime); mixed-methods based (combining quantitative and qualitative indicators); and data informed (relying on an increased transparency of platforms and functional data access).

At the time of writing, we cannot provide a complete list of service-level and structural KPIs. However, we find it important to emphasise that indicators and KPIs should be phrased in a way that prevents platforms to arbitrarily (re)interpret the questions. For this reason, we propose that each indicator be framed as a question and be complemented by a set of clearly defined guidelines. While answering the questionnaire, signatories should provide exact numbers related to the measures they have taken. This includes, among others, reporting on content sources removed or suspended due to being identified as untrustworthy by platforms or by fact-checkers (with detailed information on removals, suspensions, the length of suspensions, the number of reinstated accounts, number of

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

relapsing accounts), as well as information on the number of accounts reported by users and fact-checkers, the number of cases acted on, and number of complaints found justified (with a breakdown of reasons and grounds of intervention). This degree of detail is important to have a clear understanding of the extent of the problem and the exact nature of efforts taken by signatories. In the case of structural indicators, we suggest the use of audience samples to have a clear understanding of users' consumption of untrustworthy sources of information, while the methodology also should rely on the input of civil society members, fact checkers and other stakeholders who should provide additional qualitative and quantitative data in line with their expertise. The assessment of the impact of the Code implementation on the disinformation phenomenon should also rely on an analysis of the legal, economic, political and social context in which the Code has an effect. This allows us to consider all the agents that potentially could affect the spread of disinformation in a specific national media environment.

Chapter 6. Trustworthiness as a feature of the online information environment?

When signing the Code, Google, Facebook, TikTok and other signatories have committed to make changes to their algorithms based on so-called "trustworthiness indicators" which would reduce the risk that users get misled by shifty content. Thus, a first and key focus in our work under EDMO was the analysis of what could constitute the possible indicators that would allow online service providers to prioritise content that is informative and not likely to mislead or deceive users. The Code assigns great importance to the term "trustworthiness" when it comes to signatories' commitments. Pillar A (scrutiny of ad placements) highlights the importance of indicators of trustworthiness when identifying the sites where advertisement can be placed without (unintentionally) monetising purveyors of disinformation; and Pillar D (empowering consumers) mentions indicators of trustworthiness as the basis of content prioritisation and media literacy measures.

In the Code, the term "trustworthiness" refers first and foremost to content sources, and is often mentioned in connection with ownership transparency and the "verified identity" of content creators.[22] Indicators of trustworthiness are expected to provide the basis for platforms for im-

22 European Commission, "Code of Practice".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

proving the findability of trustworthy content sources and "diluting" visibility (downranking) of their non-trustworthy counterparts.[23] "Ranking", "prioritising" or "pushing up" trustworthy content is often mentioned in related documents and assessments as the method that makes the best use of the indicators.[24]

As such, in the current context, we define "trustworthiness" as a term that refers to the source or publisher of a piece of information. A publisher of information can be regarded as trustworthy (or credible) when the users' chance of being exposed to false or misleading content (dis- or misinformation) by that source is relatively low. Moreover, it is expected that a trustworthy publisher has a procedure in place to make sufficient and timely corrections, for the case that it publishes false or misleading content. A trustworthy source of information is, generally, transparent in its ownership, authorship and sourcing of information, in addition, it holds procedures in place to clearly label advertisement and paid content, as well as separating fact from opinion.

These considerations can be relevant when platforms have to make decisions that aim to contribute to a trustworthy online ecosystem. The European Commission points out that online platforms have supported the development of projects by independent third parties to design trustworthiness and credibility indicators, such as the Trust Project, the Cred-

23 In parallel with the discussion of trustworthiness of online content sources, it must be acknowledged that the EU audiovisual policy is facing the challenges of defining standards for the online environment and is proposing, since its most recent revision in 2018, not only "prominence" of European works as an obligation for all on-demand AVMS (Article 13(1), Recital 35 AVMSD), but also that "Member States may take measures to ensure the appropriate prominence of audiovisual media services of general interest" (Article 7(a), Recital 25 AVMSD). Member States are still in the process of adopting national prominence frameworks and approaches significantly vary from country to country. Some built on long standing traditions regarding PSM, others consider the use of 'quality labels'. See Eleonora Maria Mazzoli and Damian Tambini, "Prioritisation uncovered. The discoverability of public interest content online", Council of Europe, 2020. https://rm.coe.int/publication-content-prioritisation-report/1680a07a57.

24 European Commission VVA, "Study on the assessment of the Code of Practice against Disinformation SMART 2019/0041", 2020. https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation; ERGA, "ERGA Report in Disinformation"; European Commission, "Staff Working Document (SWD (2020)180 Final Assessment of the Code of Practice on Disinformation" 2020. https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ibility Coalition or the Journalism Trust Initiative (JTI).[25] However, an evaluation by VVA highlighted that there is no detailed information available on the integration of these indicators in platforms' search services and recommender systems.[26] There is also no mention in the documents of the criteria the platforms use to determine one source's trustworthiness, aside from recommendation by fact checkers, and Microsoft's partnership with large, "vetted" sources.[27]

In order to assist the operationalisation of trustworthiness indicators, the CMPF has looked at the three above-mentioned initiatives – the Credibility Coalition, the JTI and the Trust Project – as well as the Newsguard[28] browser extension. Their indicators focused, among others, on the following key areas:

a. Past conduct of publisher: looking at whether or not it was found repeatedly publishing verifiably false information;
b. The sourcing of articles: focusing on the diversity of sources used in published items, the transparent sourcing of articles (existence and quality of references, hyperlinks, quotes from identified sources), the openness of methods used to acquire information, reliance on reader feedback and the logical soundness of content published
c. Correction and labelling: looking at whether errors and inaccuracies were corrected or clarified on time, advertising and sponsored content was clearly labelled, fact was separated from opinion.
d. Transparency of funders and content creators: the emphasis is on the disclosure of ownership and financing of the media organisation, as well as the disclosure of authors, including their contact details.

In its Staff Working Document, the Commission indicates a preference for *ex ante* measures, e.g. when recommending the following option: "Ex ante approval by ad-placement service providers of websites selling advertisement space, possibly based on trustworthiness indicators agreed with advertisers (a 'white list' approach)".[29] This *ex ante* approach and white list is in line with the Code's attempts to classify content producers or content sources as trustworthy and untrustworthy, and the effort can be

25 "Journalism Trust Initiative". https://www.journalismtrustinitiative.org/, "The Trust Project," https://thetrustproject.org/, "Credibility Coalition". https://credibilitycoalition.org/.
26 VVA, "Study on the assessment of the Code of Practice".
27 European Commission, "Staff Working Document," 6.
28 "NewsGuard. https://www.newsguardtech.com/.
29 European Commission, "Staff Working Document," 8.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

supported by the above listed indicators as well. Some of these indicators can be checked automatically (e.g. existence of a masthead, owner information, as well as additional indicators, such as being registered with the country's media authority, or checking the average number of outside links, corrections, etc.) or provide the basis of self-reporting (such as the machine-readable, detailed questionnaire of JTI). Others need the active work of users and fact checkers (such as reporting suspicious contents by users and then flagged by fact-checkers).

However, this approach may raise some concerns. Even if the Commission's guidance emphasises that users can decide for themselves whether they want the services provided to them to be curated by trustworthiness indicators or not,[30] tools that rely solely on these indicators when determining trustworthiness of content sources may create a media environment in which established players gain further competitive advantage, while new players will face unprecedented barriers to entry. This can lead to serious problems for media pluralism and can distort the media market in a way that news players will find their access to the advertising market or other revenue sources further limited. The overreliance on these indicators can also silence diverging or non-mainstream voices.

At discussions among stakeholders, representatives of publishers have also signalled that reporting about one's trustworthiness (or even auditing this reporting) based on indicators like the ones developed by JTI or the Trust Project cannot be made mandatory. Thus, they argue, media outlets should not be labelled untrustworthy simply for not being party to such a project or initiative. Not to mention that the Code itself highlights that measures should be consistent with Article 8 of the European Convention on Human Rights (right to respect of private and family life), the fundamental right of anonymity and pseudonymity, and the proportionality principle – these could all be violated by too stringent reporting requirements on, among others, ownership or authorship. In addition, the Code also highlights Article 10 of the European Convention on Human Rights (freedom of expression), as decisions on prioritisation might limit users' access to relevant ideas and information.

In light of the previously highlighted concerns, we recommend an approach that is built on carrots, but without evident sticks. This approach would mean that content sources with a large enough audience would be asked to provide sufficient information about their compliance with indicators. Although non-compliance would not be punished with downgrad-

30 European Commission, "Guidance", 16.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ing, compliance should be rewarded with upgrading (prioritising) one's content. In practice this would mean that non-compliant publishers could go on using the services according to the current terms, while compliant sites would receive a boost by being shown more frequently to users, and by being included on a list of trusted partners for advertisers.

In parallel, fact-checkers would monitor content, or react to reporting by users. Those content creators who are caught repeatedly publishing misinformation or disinformation would be downgraded in rankings. As the detailed assessment of trustworthiness is not feasible in the case of small or new players, they should get a chance to use the organic (not paid) services of social media to reach audiences without constraints, as long as there is no sign of malicious use of the content-sharing platforms. Social media platforms themselves have already introduced some checks and requirements for users or accounts that come into play once they aim to monetise their content or boost their messages;[31] these requirements can also be used for quick trustworthiness checks to filter out which providers have to be subject to increased scrutiny.

Chapter 7. Conclusion

Disinformation is a challenging issue in content moderation, as it refers to content that is often not illegal, but can cause harm. As such, it reshapes the ways in which we think of the liability of online platforms and the governance of the digital environment. The findings of the EU-wide Media Pluralism Monitor data collection show that disinformation is increasingly seen as a risk to both media pluralism and democratic processes in EU member states, and the policy responses are often regarded as unsuitable to address the problem. This has been reiterated by our assessment of the Code of Practice. Even if signatories make a pledge for cooperation, the measures taken by online platforms under the commitments of the Code of Practice often fall short of what they have committed to. Problems can be traced back to three groups of issues: lack of guidance, limited compliance and the small number of signatories. Policymakers, on the

31 See Google.com, "YouTube Channel Monetization Policies," https://support.google.com/youtube/answer/1311392?hl=en#zippy=%2Cfollow-adsense-program-policies. Facebook.com, "Facebook Community Standards. https://www.facebook.com/business/help/185404538833362?id=2520940424820218&recommended_by=321041698514182.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

other hand, might leave possible side-effects, such as limits to freedom of expression or a decrease of media pluralism, unattended.

There are some important attempts to strengthen the Code. The Commission came up with guidelines for a new Code 2.0 that functions as a co-regulatory framework, KPIs and indicators are developed to better assess platforms' compliance and the Code's impact, while independent third parties are working on indicators of trustworthiness to provide internet users with the necessary tools for informed online navigation. These efforts are still just taking shape, thus stakeholders, such as representatives of academia, civil society, fact-checking organisations, the media, the advertising sector and regulatory authorities, need to be ready to continue the deliberation and come up with proposals that address current shortcomings.

Bibliography

Brogi, Elda, Roberta Carlini, Iva Nenadic, Pier Luigi Parcu and Mario Viola de Azevedo Cunha. *Monitoring media pluralism in the digital era: Report 2020*. Florence, 2020. https://cadmus.eui.eu/bitstream/handle/1814/67828/MPM2020-PolicyReport.pdf?sequence=5&isAllowed=y.

Brogi, Elda and Nenadic, Iva. "European plan to increase transparency and accountability of the gatekeeper online platforms to protect democracy: EDMO`s role in the Commission's digital policy approach", 2021. http://www.medialaws.eu/european-plan-to-increase-transparency-and-accountability-of-the-gatekeeper-online-platforms-to-protect-democracy-edmos-role-in-the-commissions-digital-policy-approach/.

Brogi, Elda et al. *Monitoring media pluralism in the digital era: Report 2021*. Florence, forthcoming 2021.

Credibility Coalition. "Credibility Coalition: Our goal: to understand the veracity, quality and credibility of online information." Accessed July 14, 2021. https://credibilitycoalition.org/.

ERGA. "ERGA Report on Disinformation: Assessment of the Implementation of the Code of Practice", 2020. https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf.

European Commission. *Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market*. Brussels: European Commission, 2000. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32000L0031.

European Commission. T*ackling online disinformation: a European Approach. COM/2018/236 final*, Brussels: European Commission, 2018. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

European Commission. *Action Plan against disinformation.* Brussels: European Commission, 2018. https://digital-strategy.ec.europa.eu/en/library/action-plan-against-disinformation.

European Commission. *Code of Practice on Disinformation.* Brussels: European Commission, 2018. https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation.

European Commission. *European Democracy Action Plan: making EU democracies stronger.* Brussels: European Commission, 2020. https://ec.europa.eu/commission/presscorner/detail/en/IP_20_2250.

European Commission VVA. *Study on the assessment of the Code of Practice against Disinformation SMART 2019/0041*. Brussels: European Commission, 2020. https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation.

European Commission. *The Digital Services Act package.* Brussels: European Commission, 2020. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.

European Commission. *Staff Working Document (SWD (2020)180 Final - Assessment of the Code of Practice on Disinformation.* Brussels: European Commission, 2020. https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement.

European Commission. *Annual self-assessment reports of signatories to the Code of Practice on Disinformation*. Brussels: European Commission, 2019. https://digital-strategy.ec.europa.eu/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019.

European Commission. *European Commission Guidance on Strengthening the Code of Practice on Disinformation European Commission.* Brussels: European Commission, 2019. https://ec.europa.eu/newsroom/dae/redirection/document/76495.

High Level Expert Group on Fake News and Online Disinformation. *A multi-dimensional approach to disinformation - Report of the independent High level Group on fake news and online disinformation*, Brussels: European Commission, 2018. https://digital-strategy.ec.europa.eu/en/library/final-report-high-level-expert-group-fake-news-and-online-disinformation.

Google. "YouTube Channel Monetization Policies" Accessed July 14, 2021. https://support.google.com/youtube/answer/1311392?hl=en#zippy=%2Cfollow-adsense-program-policies. Facebook.com, "Facebook Community Standards. https://www.facebook.com/business/help/185404538833362?id=2520940424820218&recommended_by=321041698514182.

Journalism Trust Initiative. "Journalism Trust Initiative. Rewarding trustworthy Journalsm" Accessed July 14, 2021. https://www.journalismtrustinitiative.org/.

Mazzoli, Eleonora Maria and Damian Tambini. *Prioritisation uncovered. The discoverability of public interest content online*. Strasbourg: Council of Europe, 2020. https://rm.coe.int/publication-content-prioritisation-report/1680a07a57.

Nenadić, Iva. "Unpacking the 'European approach' to tackling challenges of disinformation and political manipulation," *Internet Policy Review* 8, No. 4 (2019): 1-22.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

NewsGuard. "NewsGuard: Das Vertrauens-Tool furs Netz," Accessed July 14, 2021. https://www.newsguardtech.com/.

The Trust Project. "The Trust Project: How do you know which news stories you can trust?" Accessed July 14, 2021. https://thetrustproject.org/.

Wardle, Claire and Hossein Derakhshan. "Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information," in *Journalism, 'Fake News' & Disinformation. Handbook for Journalism Education and Training*, eds. Cherilyn Ireton and Julie Posetti, 44-56. Paris: UNESCO, 2018. https://unesdoc.unesco.org/ark:/48223/pf0000265552/PDF/265552eng.pdf.multi.

Woolley, Samuel. "Encrypted messaging apps are the future of propaganda," *Brookings*, May 1, 2020, https://www.brookings.edu/techstream/encrypted-messaging-apps-are-the-future-of-propaganda/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Regulation of Online Disinformation in Singapore

Peng Hwa Ang, Gerard Goggin

Abstract: IT-savvy Singapore is typically seen by many governments as a model for governance, especially in the technology space. Understandably, when Singapore passed its Protection Against Online Falsehoods and Misinformation Act in 2019 to address misleading information, many took notice. This article discusses the process of the passage and early uses of the law. The distinctive features of the law are that it can only be used by the government (not citizens), against falsehoods online (not offline) and the truthfulness of a statement is determined by a government minister, with an appeal to the court in the event the truthfulness of the statement is disputed. Statements are allowed to stay online on condition that a correction statement by the Singapore Government is inserted on the same page. While platforms and online media have all complied when given such "correction directions", one overseas Singaporean user has resisted and has had his Facebook page geo-blocked in Singapore. The Singaporean approach suggests that while it can work, there are limitations to the law.

Keywords: social media, Singapore, fake news, anti-fake news law, misinformation, disinformation, Facebook, freedom of expression, censorship

Chapter 1. Introduction

When the Singapore Government presented the Protection Against Online Falsehoods and Misinformation Bill in 2019, it was immediately criticised by several parties including the International Commission of Jurists, which said the law made the government "the sole arbiter of what information is permissible online and what is not".[1]

1 ICJ, "Singapore: Parliament must reject internet regulation bill that threatens freedom of expression." April 4, 2019, accessed July 6, 2021, https://www.icj.org/singapore-parliament-must-reject-internet-regulation-bill-that-threatens-freedom-of-expression/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Notwithstanding the criticism, and perhaps because of the legend of Singapore being more information-technology savvy, the law has been studied in several countries. It was viewed favourably when compared with the Malaysian anti-misinformation law that was eventually scrapped.[2] Nigeria's anti-social media bill, which failed to pass, was titled "Protection from Internet Falsehood and Manipulation and for Other Related Matters Bill". The similarity in title and content led to criticism of plagiarism.[3] Sri Lanka studied the Singapore law before passing its own anti-fake news.[4]

Many governments have been reluctant to move in early to regulate platforms. First, this is partly due to the international dominance of philosophies that favour experimentation and innovation in new markets. Second, because of the challenges in regulating many areas of digital platforms that have accumulated over 20 or more years of efforts in regulation and governance of Internet, social and mobile media and apps. Ironically, in some respects digital platforms can be easy to identify for regulation purposes – because of their 'platform' characteristics, and especially because many of the large ones have conspicuous owners or custodians. However, the platforms are not often subject to licensing regimes, nor is the nature of their offending services, public concerns, or 'market failures' easy to address – as the case of content moderation shows (where automation and algorithms can only go so far).

Singapore has typically wished to cultivate the business, economic, productivity, and social connectivity of platforms, with the kind of open and facilitative approach it has used for ICTs over many years. This can be seen in the 'sandbox' approach to fintech apps, where the Singapore Government, like a number of others, has sought to fashion a new 'light-touch' regulatory approach to a clearly highly lucrative emerging area of digital platforms. However, the major exception for Singapore has been the regulation of particular kinds of Internet content that are not consistent with its norms and expectations on appropriate types of speech, or do not respect social cohesion (especially in relation to racial and intercultural

2 Kok, 2021.

3 Sunday Aborisade, "Anti-social media bill: Senator defends alleged plagiarism of Singapore statute," Punch Newspapers, accessed June 23, 2021, https://punchng.com/anti-social-media-bill-senator-defends-alleged-plagiarism-of-singapore-statute/.

4 Shreetesh Angwalkar and Roxanne Powell, "Culture Matters: Sri Lanka Implements Singapore Style Law to Control Fake News," Spherex, accessed June 23, 2021, https://spherex.com/regulation/sri-lanka-implements-singapore-style-law-to-control-fake-news/.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

harmony). So Singapore has had a longstanding set of approaches to governing freedoms of expression on the Internet.[5]

Fake news and misinformation is a leading area world-wide where due to mounting concerns in recent years governments have been prepared to step in. This is certainly the explicit rationale for Singapore's Protection Against Online Falsehoods and Misinformation Act (POFMA), which it has insisted upon, in the face of criticisms that the law is the latest instance of Singapore's emphasis on keeping a tight rein on freedom of expression, especially with the new possibilities of communication via Internet, blogs, and mobile and social media platforms. The public discussion and consultation on the POFMA reforms, and especially the most dramatic moments in the parliamentary committee proceedings showed an implacable willingness of key government figures to send a strong message to digital platform operators. This is all the more impressive, given the power of the transnational operators of these firm – and also in the face of Singapore's keen desire to establish itself as the preferred Asia-Pacific headquarters of marquee tech firms.

Having conceded on POFMA (or having little choice but to do so), digital platforms based in Singapore are keen to head off at the pass, so to speak, other tendencies in platform regulation gathering momentum elsewhere.[6] The digital platforms are clearly more comfortable with Singapore's stance on privacy and data regulation (although privacy concerns have surged with data collection and use in COVID-19 public health surveillance – e.g. via contact tracing apps and QR check-ins) or its light to moderate consumer protections in relation to digital services, products, and platforms. So, it is likely that the platforms see POFMA as a continuation of Singapore's long-running efforts in censorship and control of information, content, and types and contexts of expression that is deemed inappropriate or offensive.

This article aims to explain how the POFMA came to pass.

---

5 Howard Lee and Terence Lee, "From contempt of court to fake news: public legitimisation and governance in mediated Singapore," *Media International Australia* 173, no. 1 (June 2019), https://doi.org/10.1177/1329878X19853074; Peng Hwa Ang and Berlinda Nadarajan, "Censorship and the Internet: a Singapore perspective," *Communications of the ACM* 39, no. 6 (June 1996), https://doi.org/10.1145/228503.228520.

6 Terry Flew et al., "Return of the regulatory state: A stakeholder analysis of Australia's Digital Platforms Inquiry and online news policy," *The Information Society* 37, no. 2 (2021), https://doi.org/10.1080/01972243.2020.1870597.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Chapter 2. Context

Back in 2014, the first author had encountered Metamorphosis, an NGO in Skopje, Macedonia (now called North Macedonia), that was fact-checking Macedonian newspapers. It was such a novel idea that the first author remembered it but as it was not part of the research agenda, it was not pursued.

Then in the 2016 US Presidential election, the BBC uncovered a city in Macedonia that was "getting rich from fake news".[7] Law Minister K. Shanmugam said in Parliament in April 2017 that the Singapore Government was "seriously considering" a law to combat "fake news" as current laws were inadequate.[8] Two months later, he added that such a law was a "no-brainer".[9]

Then in an unusual move, the government convened a Parliamentary Select Committee in 2018 to study the issue and seek feedback from experts and the public.[10] Altogether 167 written representations were received of which 65 individuals and organisations gave oral presentations that eventually lasted two working weeks.[11]

Two persons' feedback stood out in the hearing. The first was Facebook's vice-president of public policy for Asia-Pacific, Simon Milner. He admitted that the company had been remiss in its handling of the Cambridge Analytica issue.[12]

The second was Oxford visiting scholar Thum Ping Tjin, who in his submission to the Select Committee asserted that "'fake news' has not, historically, had much of an impact in Singapore — with one major exception: the People's Action Party government has, historically, spread 'fake news' for narrow party-political gain".[13] Thum, on the last day of the Committee's meeting, was then subjected to a six-hour exchange with

---

7 Emma Jane Kirby, "The city getting rich from fake news," BBC, accessed June 23, 2021, https://www.bbc.com/news/magazine-38168281.

8 Rachel Au-Yong, "Parliament: Government to review laws to tackle fake news," The Straits Times, accessed June 23, 2021, https://www.straitstimes.com/politics/parliament-government-to-review-laws-to-tackle-fake-news.

9 Seow Bei Yi and Nur Asyiqin M. Salleh, "Shanmugam sets out strategies in battle against fake news," The Straits Times, accessed June 23, 2021, https://www.straitstimes.com/singapore/shanmugam-sets-out-strategies-in-battle-against-fake-news.

10 Seow, 2018a.

11 Sim, 2018.

12 Seow, 2018b.

13 Ping Tjin Thum, "Submission to the Select Committee on Deliberate Online Falsehoods, Parliament of Singapore. Written Representation 83," 1, accessed

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Minister Shanmugam in his capacity as Select Committee member who questioned Thum's research and his position at the University of Oxford. Among the issues raised by Minister Shanmugam was the veracity of a 2013 paper in which Thum alleged that false information was used by the government to justify preventive detention in Operation Coldstore in 1963.[14]

The Singapore-based human rights organisation, Maruah, said that the Committee appeared to be "overly focused, through a process of intense interrogation, on showing that the witnesses were propagators of 'falsehoods'".[15] The Chair of the Committee later said that "no weight" had been given to Thum's views.[16]

Chapter 3. What is PoFMA/ How Does PoFMA Work

The Act defines a "statement of fact" as "a statement which a reasonable person seeing, hearing or otherwise perceiving it would consider to be a representation of fact" (S. 2(2a)). This uses the fabled "reasonable person" as the yardstick to determine facticity. This would be an objective standard.

On the other hand, what is false or misleading is not clearly defined. Under the Act, a statement is deemed to be false "if it is false or misleading, whether wholly or in part, and whether on its own or in the context in which it appears" (S. 2(2b)). Oddly, the reasonable person is absent, which suggests a subjective standard.

June 23, 2021, https://www.parliament.gov.sg/docs/default-source/sconlinefalsehoods/written-representation-83.pdf.

14 In Operation Coldstore, 113 persons were arrested in a covert security operation that the government said was "aimed at crippling the Communist open front organisation" that threatened Singapore's internal security. Current scholarship differs on the degree of the Communist threat. While Thum argues that the Communist threat was inflated, another scholar (Ramakrishna, Kumar (2015). *Original Sin: Revising the Revisionist Critique of the 1963 Operation Coldstore in Singapore*. Singapore: ISEAS Publishing) argues otherwise.

15 Low Youjin, "Maruah slams Select Committee's 'confrontational stance'," today, accessed June 29, 2021, https://www.todayonline.com/singapore/maruah-slams-select-committees-confrontational-stance.

16 Faris Mokhtar, "No weight given to historian Thum Ping Tijn's views and he 'clearly lied' about credentials, says committee," today, accessed June 29, 2021, https://www.todayonline.com/singapore/no-weight-given-historian-thum-ping-tjins-views-he-not-credible-representor-select.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Under the Protection from Online Falsehood and Manipulation Act, which passed in 2018, any Minister can issue a "correction direction" to statements made online that are false in his or her view and if the Minister thinks that it is in the public interest to issue such a direction (Section 11).

The public interest test is defined in S. 8(3) as false statements that may:

- be prejudicial to the security of Singapore or any part of Singapore;
- be prejudicial to public health, public safety, public tranquillity or public finances;
- be prejudicial to the friendly relations of Singapore with other countries;
- influence the outcome of an election to the office of President, a general election of Members of Parliament, a by-election of a Member of Parliament, or a referendum;
- incite feelings of enmity, hatred or ill-will between different groups of persons; or
- diminish public confidence in the performance of any duty or function of, or in the exercise of any power by, the Government, an Organ of State, a statutory board, or a part of the Government, an Organ of State or a statutory board,

While subparagraphs (a) to (e) are mentioned in the freedom of speech clause in Article 14 of the Singapore Constitution, subparagraph S. 8(3)(f) (diminish public confidence) may be questionable because it is not specifically mentioned in that Article.

As practised, the correction direction means posting a correction in a prominent position but the original post stays. Such a direction, in the face of a resistant author, would require the cooperation of the online host or platform.

In the government's view, because "[c]ensorship entails banning or suppressing offending material" and the government has "not banned or suppressed" the post, there is no censorship.[17]

Minister Shanmugam said that the correction direction "actually encourages greater democracy" because it encourages more information. He

17 Justin Ong, "In letter to Washington Post, Govt refutes Pofma criticism, saying it 'has not suppressed anything'," today, accessed June 29, 2021, https://www.todayonline.com/singapore/letter-washington-post-govt-responds-pofma-criticism-saying-it-has-not-suppressed-anything.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

said, "You can argue censorship only if your article is taken down. But your article is there. So, what are you embarrassed about?"[18]

The law also empowers the Minister to issue a "stop communication direction" (S. 12). So far, no such direction has been issued. The correction direction could also be "targeted" (S. 21) so that only those who received the original post would see the correction. It could also be "general" (S. 23) so that those who visit the platform or site but not the specific page would also view the correction. Most of the correction directions have been targeted.

A general correction was issued in May 2021 after an Indian politician asserted that a new Singapore variant especially dangerous for children was spreading to India.[19] In such an order, Facebook, Twitter and the newspaper and magazine conglomerate Singapore Press Holdings (SPH) was directed to post such a general correction such that all users would see the correction, even if they had not seen the original post.

For offenders who do not comply with correction or stop directions, access to the site or platform may be blocked through an "access blocking order" that is given to an intermediary or service provider (S. 28). For recalcitrant offenders, even if they comply with the correction or stop directions, access may also be denied they have had more than three such directions in a six-month period. Alex Tan, who had run as an opposition candidate during the general election of 2011, is the only person who has been issued such a blocking order. In February 2020, Facebook was ordered to block access to Tan's page as he had not posted corrections following at least three correction orders dating from November 2019. The more recent posts concerned the COVID-19 situation in Singapore. Tan's page was designated a Declared Online Location (S. 32) and Facebook was issued a disabling order to block the page.[20]

---

18 Aqil Haziq Mahmud, "POFMA encourages democracy, does not disadvantage opposition: Shanmugam on upcoming General Election," Channel News Asia, accessed June 29, 2021, https://www.channelnewsasia.com/news/singapore/pofma-democracy-disadvantage-opposition-election-ge-shanmugam-12857488.

19 "Pofma correction directions to be issued to Facebook, Twitter, SPH Magazines over 'Singapore' variant of Covid-19 falsehood," Today, accessed June 29, 2021, https://www.todayonline.com/singapore/pofma-correction-directions-facebook-twitter-sph-magazines-singapore-variant-falsehood.

20 "Facebook blocks Singapore users' access to States Times Review page," Channel News Asia, accessed June 29, 2021, https://www.channelnewsasia.com/news/singapore/facebook-blocks-singapore-users-access-states-times-review-pofma-12446952; Info-Communications Media Development Authority, *Protection from Online Falsehoods And Manipulation Act 2019 (Act 18 Of 2019) Notice of Declaration Under*

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Of all the orders, Facebook appears most concerned about the access blocking order. It said: "We believe orders like this are disproportionate and contradict the government's claim that POFMA would not be used as a censorship tool" (BBC News 2020).[21]

Chapter 4. Issues

The Protection Against Online Falsehoods and Misinformation Bill was the subject of much controversy when it was first presented. Academics from the authors' university, including the first author, petitioned to say that academic research could be caught under the Act because new research when first presented could run counter to conventional wisdom and so may be deemed as false. There had been an incident years before when two authors were criticised for being inaccurate in their findings that had used data from the government's website that were incomplete.[22]

The current Education Minister Ye Kung Ong said that the two economists would not have been caught under the new law because they did not fabricate data nor cause public alarm.[23]

The most significant concern was over the power of any minister to decide whether a statement was false and to order a correction. Thus, the government can invoke the law, but ordinary citizens may not.

Also contentious was the determination of facticity. The Minister for Law Shanmugam in the second reading of the Bill reiterated the distinction between facts, which the law was intended to cover, and opinion, which the Bill did not. He added that the ultimate arbiter would not be

---

Section 32(5) of Act. (424). [POFMA/DC/2020/02-02; AG/LEGIS/SL/256B/2015/4 Vol. 1]. https://www.egazette.com.sg/pdf.aspx?ct=gg&yr=2020&filename=20gg0442.pdf.

21 "Facebook expresses 'deep concern' after Singapore orders page block," BBC News, accessed June 29, 2021, https://www.bbc.com/news/world-asia-51556620.

22 "Singapore attacks 'foreigners get most new jobs' claim," The Star, Accessed June 29, 2021, https://www.thestar.com.my/news/regional/2003/08/02/singapore-attacks-foreigners-get-most-new-jobs-claim.

23 Janice Lim, "Education Minister explains why fake news laws don't apply to erroneous 2003 study on job creation," Today, accessed June 29, 2021, https://www.todayonline.com/singapore/academics-will-not-be-caught-proposed-laws-if-they-abide-research-discipline-education.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

the government but a judge.[24] But after the Bill was passed, the Deputy Attorney-General in a court case said that the law did cover matters of interpretation and that a correction direction may be issued based on the minister's interpretation.[25]

Further, S. 61 of the Act empowers the Minister of Communication and Information, under whose purview to Act falls, to exempt "any person or class of persons from any provision of this Act." It has been pointed out that, taken at face value, this means the Minister may exempt all his or her fellow ministers from having to meet the requirement of truthfulness or potential harm when issuing a direction. Taken in good faith, the writers Wijaya and Thuraisingam suggest that the exemption may be for criminal liability. But then this would be interfering with the judicial process. In any event, the provision could do with clarity through legislation or judicial review (2019).

Chapter 5. Use

Since POFMA came into force in October 2019 to July 2020, 71 orders have been issued. The most frequent recipients of the orders have been activists and opposition political figures.[26] The first POFMA order was directed to Brad Bowyer, an opposition political figure, for a Facebook post that questioned the independence of government-linked investment companies.[27] The next three correction directions were issued to persons who were affiliated with opposition political parties. This led Nominated Member of Parliament Walter Theseira to ask if "the Government was setting up 'speed traps where opposition politicians drive and not elsewhere'". Information Minister S. Iswaran replied that the use of POFMA against

24 Parliament Singapore, *Singapore Parliamentary Debates*, *Vol. 94*, *Sitting No: 104*, *Sitting date: May 7*, https://sprs.parl.gov.sg/search/fullreport?sittingdate=07-05-2019.

25 Rei Kurohi, "Fake news law does cover matters of interpretation: AGC," *Straits Times*, January 18, 2020, https://www.straitstimes.com/singapore/fake-news-law-does-cover-matters-of-interpretation-agc.

26 Andrea Carson and Liam Fallon, *Fighting Fake News: A Study of Online Misinformation Regulation in the Asia Pacific* (Melbourne: La Trobe, 2021), https://doi.org/10.26181/60640ea43558f.

27 "POFMA Office directs Brad Bowyer to correct Facebook post in first use of 'fake news' law," Channel News Asia, accessed June 29, 2021 https://www.channelnewsasia.com/news/singapore/brad-bowyer-facebook-post-falsehood-pofma-fake-news-12122952.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

politicians was "an unfortunate convergence or coincidence", adding that it was "just the consequence of their actions".[28]

Of the correction directions issued from November 2019 to July 2020, 12 were directed at foreign entities. And of these, 10 were directed at Alex Tan's Facebook page and his clutch of websites; Tan was affiliated with an opposition political party but had moved to live in Australia.

Chapter 6. Comparison with Other Jurisdictions

How does the POFMA compare with the laws passed elsewhere?

Under Germany's 2018 Network Enforcement Act ("Netzwerkdurchsetzungsgesetz", colloquially referred to as the "Facebook Law" or "NetzDG") social media platforms must remove "illegal content" (such as hate speech and pro-Nazi ideology) or face fines of up to €50 million. NetzDG empowers the authorities to remove content that are illegal under existing laws. Singapore's POFMA law creates new offences for the intentional malicious spread of falsehoods. A speedy response, instead of removal of the content, is the primary focus.[29]

France's law empowers judges to remove misinformation during the election campaign upon the complaint of any political party or candidate. The judge must decide within 48 hours of the complaint if the information is manifestly false, was being disseminated widely online, and might compromise the outcome of the election. The law applies only during an election campaign. Unlike the Singapore law, any political party or candidate may invoke the law. On the other hand, the French law only provides for blocking of the content instead of a correction notice.[30]

More recently, correction directions have been given over information surrounding the COVID-19 pandemic. A year after the passage of the bill, the state-owned TV news station, Channel News Asia, evaluated the effectiveness of the law. It suggests that by enabling the control of the

---

28 Janice Lim, "'Unfortunate coincidence' initial Pofma actions directed at opposition parties, affiliated figures: Iswaran," Today, accessed June 29, 2021, https://www.todayonline.com/singapore/unfortunate-coincidence-first-four-pofma-actions-directed-opposition-politicians.

29 Sashi Jayakumar, Benjamin Ang and Nur Diyanah Anwar, „Fake news and disinformation: Singapore perspectives," in *Disinformation and Fake News,* eds. Sashi Jayakumar, Benjamin Ang and Nur Diyanah Anwar (Singapore: Palgrave Macmillan, 2020), 137-158, https://doi.org/10.1007/978-981-15-5876-4_11.

30 Jayakumar, Ang and Anwar, „Fake news and disinformation".

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

spread of misinformation, POFMA may have contributed to the relative success of Singapore in taming the pandemic with low infection and low mortality rates.[31]

Chapter 7. What Next

That misinformation in the news may have serious practical consequences in life has been made most evident by the pandemic. However, the ways to battle the spread of misinformation have yet to be fully understood. For example, while social media have often been blamed for the rapid propagation of misinformation, some research suggests that the use of social media may in fact reduce the spread of misinformation when other variables are controlled.[32] The reason is that the use of social media affords wider exposure to other information.

In that light, the process and outcome of the deployment of POFMA could well be reviewed. In the political sphere, it is not clear if there were any winners. Correction directions were issued during the 2020 General Election campaign period. Singapore's electioneering period is only nine days, the minimum specified by law. Because of the fact, that the turnaround time to file an appeal in court is nine days, a correction direction during the electioneering period has almost no chance of being reversed by the court. Did POFMA affect the campaigning or the election outcome? Opposition political leaders appear divided. One said that they could "take advantage" of the law to ferret out information by "forcing" the government to issue corrections on controversial statements. Another opposition figure received four correction directions, which meant that his social media posts had to be amended to include the correction from the government.[33] The view that POFMA was aimed at political figures was

31 Aqil Haziy Mahmud, "IN FOCUS: Has POFMA been effective? A look at the fake news law, 1 year since it kicked in," Channel News Asia, accessed June 29, 2021. https://www.channelnewsasia.com/news/singapore/singapore-pofma-fake-news-law-1-year-kicked-in-13163404.

32 Daniel Halpern, Sebastián Valenzuela, James Katz, and Juan Pablo Miranda, "From Belief in Conspiracy Theories to Trust in Others: Which Factors Influence Exposure, Believing and Sharing Fake News," in *Social computing and social media. Design, human behavior and analytics*, ed. Gabriele Meiselwitz (Cham: Springer, 2019), 217-232, https://doi.org/10.1007/978-3-030-21902-4_16.

33 Bhavan Jaipragas, "Has Singapore's fake news law passed the election test?" *South China Morning Post*, July 7, 2020, https://www.scmp.com/week-asia/politics/article/3092228/has-singapores-fake-news-law-passed-election-test.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

strengthened when there were no directions issued for several months after the 2020 General Election. On the other hand, informal conversations suggests that such directions during electioneering may backfire by garnering underdog support for the opposition parties.

The somewhat technical but critical point of burden of proof itself will require review. At the time of writing, there was a point of law that had yet to be settled: on whom does the burden of proof lie to prove the truthfulness of a statement? That is, if a Minister were to issue a correction direction, would he or she have to prove that the statement in question was false? Or is the burden of proof on the individual to prove that the statement is true? Two conflicting cases have led to an appeal that has apparently yet to be decided.[34]

Finally, using correction as the chief mechanism to address misinformation will require further follow up and research. It is known that there is a "boomerang effect" in persuasive messages where such messages have the opposite effect of the intended outcome. It would appear that the truth indeed is out there.

Bibliography

Aborisade, Sunday. "Anti-social media bill: Senator defends alleged plagiarism of Singapore statute." Punch Newspapers. Accessed June 23, 2021. https://punchng.com/anti-social-media-bill-senator-defends-alleged-plagiarism-of-singapore-statute/.

Ang, Peng Hwa and Belinda Nadarajan. "Censorship and the Internet: a Singapore perspective." *Communications of the ACM* 39, no. 6 (June 1996): 72-78. https://doi.org/10.1145/228503.228520.

Angwalkar, Shreetesh and Roxanne Powell. "Culture Matters: Sri Lanka Implements Singapore Style Law to Control Fake News." Spherex. Accessed June 23, 2021. https://spherex.com/regulation/sri-lanka-implements-singapore-style-law-to-control-fake-news/.

Au-Yong, Rachel. "Parliament: Government to review laws to tackle fake news." The Straits Times. Accessed June 23, 2021. https://www.straitstimes.com/politics/parliament-government-to-review-laws-to-tackle-fake-news.

34 Lydia Lam, "Judgment reserved in The Online Citizen, SDP's POFMA appeals, as court grapples with legal issues including burden of proof," Channel News Asia, accessed June 29, 2021, https://www.channelnewsasia.com/news/singapore/toc-sdp-pofma-appeals-judgement-court-appeal-13121094.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Audenhove, Leo Van and Karen Donders. (2019). Talking to Carson, A., & Fallon, L. (2021). Fighting Fake News: A Study of Online Misinformation Regulation in the Asia Pacific. *La Trobe University*. https://www.latrobe.edu.au/data/assets/pdffile/0006/1204548/carson-fakenews.pdf.

BBC News. "Facebook expresses 'deep concern' after Singapore orders page block." Accessed June 29, 2021. https://www.bbc.com/news/world-asia-51556620.

Carson, Andrea and Liam Fallon. *Fighting Fake News: A Study of Online Misinformation Regulation in the Asia Pacific*. Melbourne: La Trobe, 2021. https://www.latrobe.edu.au/__data/assets/pdf_file/0006/1204548/carson-fakenews.pdf.

Channel News Asia. "POFMA Office directs Brad Bowyer to correct Facebook post in first use of 'fake news' law." Accessed June 29, 2021. https://www.channelnewsasia.com/news/singapore/brad-bowyer-facebook-post-falsehood-pofma-fake-news-12122952.

Channel News Asia. "Facebook blocks Singapore users' access to States Times Review page." Accessed June 29, 2021. https://www.channelnewsasia.com/news/singapore/facebook-blocks-singapore-users-access-states-times-review-pofma-12446952.

Garrett, R. Kelly, Nisbet, Erik C., & Lynch, Emily K. (2013). Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naive theory. J*ournal of Communication*, 63(4), 617-637. https://doi.org/10.1111/jcom.12038.

Halpern, Daniel, Sebastián Valenzuela, James Katz, and Juan Pablo Miranda. "From Belief in Conspiracy Theories to Trust in Others: Which Factors Influence Exposure, Believing and Sharing Fake News." In *Social computing and social media. Design, human behavior and analytics,* edited by Gabriele Meiselwitz, 217-232. Cham: Springer, 2019, https://doi.org/10.1007/978-3-030-21902-4_16.

Flew, Terry, Rosalie Gillett, Fiona Martin and Lucy Sunman. "Return of the regulatory state: A stakeholder analysis of Australia's Digital Platforms Inquiry and online news policy." *The Information Society* 37, no. 2 (2021): 128-145. https://doi.org/10.1080/01972243.2020.1870597.

George, Cherian (2007). Consolidating authoritarian rule: Calibrated coercion in Singapore. *The Pacific Review*, *20*(2), 127-145. https://doi.org/10.1080/09512740701306782.

Info-Communications Media Development Authority. (2020). *Protection from Online Falsehoods And Manipulation Act 2019 (Act 18 Of 2019) Notice of Declaration Under Section 32(5) of Act.* (424). [POFMA/DC/2020/02-02; AG/LEGIS/SL/256B/2015/4 Vol. 1]. https://www.egazette.com.sg/pdf.aspx?ct=gg&yr=2020&filename=20gg0442.pdf.

ICJ, "Singapore: Parliament must reject internet regulation bill that threatens freedom of expression," April 4 2019, accessed July 6, 2021, https://www.icj.org/singapore-parliament-must-reject-internet-regulation-bill-that-threatens-freedom-of-expression/.

Jaipragas, Bhavan. "Has Singapore's fake news law passed the election test?" *South China Morning Post,* July 7, 2020. https://www.scmp.com/week-asia/politics/article/3092228/has-singapores-fake-news-law-passed-election-test.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Jayakumar, Sashi, Benjamin Ang, and Nur Diyanah Anwar. „Fake news and disinformation: Singapore perspectives." In *Disinformation and Fake News,* edited by Sashi Jayakumar, Benjamin Ang and Nur Diyanah Anwar, 137-158. Singapore: Palgrave Macmillan, 2020, https://doi.org/10.1007/978-981-15-5876-4_11.

Kirby, Emma J. "The City Getting Rich From Fake News." BBC. Accessed June 23, 2021. https://www.bbc.com/news/magazine-38168281.

Kok, Raphael Chi Ren. "Suppressing Fake News or Chilling Free Speech Are the Regulatory Regimes of Malaysia and Singapore Compatible With International Law?" *Journal of Malaysian and Comparative Law*. Vol 47. 1 (2020). 2021 June 29. Accessed 6 July 2021. https://ejournal.um.edu.my/index.php/JMCL/article/view/30840.

Kurohi, Rei. "Fake news law does cover matters of interpretation: AGC." *Straits Times*. January 18, 2020. https://www.straitstimes.com/singapore/fake-news-law-does-cover-matters-of-interpretation-agc.

Lam, Lydia. "Judgment reserved in The Online Citizen, SDP's POFMA appeals, as court grapples with legal issues including burden of proof." Channel News Asia. Accessed June 29, 2021. https://www.channelnewsasia.com/news/singapore/toc-sdp-pofma-appeals-judgement-court-appeal-13121094.

Lee, Howard and Terence Lee. "From contempt of court to fake news: public legitimisation and governance in mediated Singapore." *Media International Australia* 173, no. 1 (June 2019): 81-92. https://doi.org/10.1177/1329878X19853074.

Lim, Janice. "Education Minister explains why fake news laws don't apply to erroneous 2003 study on job creation." Today. Accessed June 29, 2021. https://www.todayonline.com/singapore/academics-will-not-be-caught-proposed-laws-if-they-abide-research-discipline-education.

Lim, Janice. "'Unfortunate coincidence' initial Pofma actions directed at opposition parties, affiliated figures: Iswaran." Today. Accessed June 29, 2021. https://www.todayonline.com/singapore/unfortunate-coincidence-first-four-pofma-actions-directed-opposition-politicians.

Ministry of Law. (2018). Select Commitee On Deliberate Online Falsehoods: Causes, Consequences and Countermeasures. January 5. Retrieved from https://www.mlaw.gov.sg/news/press-releases/select-committee-deliberate-online-falsehoods.

Mokhtar, Faris. "No weight given to historian Thum Ping Tjin's views and he 'clearly lied' about credentials, says committee." Today. Accessed June 29, 2021. https://www.todayonline.com/singapore/no-weight-given-historian-thum-ping-tjins-views-he-not-credible-representor-select.

Mahmud, Aqil Haziq. "POFMA encourages democracy, does not disadvantage opposition: Shanmugam on upcoming General Election." Channel News Asia. Accessed June 29, 2021. https://www.channelnewsasia.com/news/singapore/pofma-democracy-disadvantage-opposition-election-ge-shanmugam-12857488.

Mahmud, Aqil. "IN FOCUS: Has POFMA been effective? A look at the fake news law, 1 year since it kicked in." Channel News Asia. Accessed June 29, 2021. https://www.channelnewsasia.com/news/singapore/singapore-pofma-fake-news-law-1-year-kicked-in-13163404.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Ong, Justin. "In letter to Washington Post, Govt refutes Pofma criticism, saying it 'has not suppressed anything'." Today. Accessed June 29, 2021. https://www.todayonline.com/singapore/letter-washington-post-govt-responds-pofma-criticism-saying-it-has-not-suppressed-anything.

Seow, Bei Yi. "Parliament: House votes unanimously to form committee looking into fake news." *The Straits Times*. January 11, 2018a. Accessed July 6, 2021. https://www.straitstimes.com/politics/parliament-proposal-to-appoint-select-committee-to-examine-online-falsehoods.

Seow, Bei Yi. "Facebook admits it should have told users earlier about breach of policy." *The Straits Times*. March 22, 2018b. Accessed July 6, 2021. https://www.straitstimes.com/politics/facebook-admits-it-should-have-told-users-earlier-about-breach-of-policy.

Seow, Bei Yi and Nur Asyiqin M. Salleh. "Shanmugam sets out strategies in battle against fake news." The Straits Times. Accessed June 23, 2021. http://www.straitstimes.com/singapore/shanmugam-sets-out-strategies-in-battle-against-fake-news.

Sim, Royston. "Select Committee on fake news: 22 recommendations unveiled to combat online falsehoods." *The Straits Times*. Sept 20, 2018. Accessed July 6, 2021, https://www.straitstimes.com/singapore/select-committee-on-fake-news-22-recommendations-unveiled-to-combat-online-falsehoods.

Parliament Singapore. *Singapore Parliamentary Debates*, *Vol. 94, Sitting No: 104, Sitting date: May 7*. https://sprs.parl.gov.sg/search/fullreport?sittingdate=07-05-2019.

The Star. "Singapore attacks 'foreigners get most new jobs' claim." Accessed June 29, 2021. https://www.thestar.com.my/news/regional/2003/08/02/singapore-attacks-foreigners-get-most-new-jobs-claim.

Thum, Ping Tjin. "Submission to the Select Committee on Deliberate Online Falsehoods, Parliament of Singapore. Written Representation 83." Accessed June 23, 2021. https://www.parliament.gov.sg/docs/default-source/sconlinefalsehoods/written-representation-83.pdf.

Xoujin, Low. "Maruah slams Select Committee's 'confrontational stance'." Today. Accessed June 29, 2021. https://www.todayonline.com/singapore/maruah-slams-select-committees-confrontational-stance.

Today. "Pofma correction directions to be issued to Facebook, Twitter, SPH Magazines over 'Singapore' variant of Covid-19 falsehood." Accessed June 29, 2021. https://www.todayonline.com/singapore/pofma-correction-directions-facebook-twitter-sph-magazines-singapore-variant-falsehood.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Conclusions: Regulatory Responses to Communication Platforms: Models and Limits

Judit Bayer, Bernd Holznagel, Päivi Korpisaari, Lorna Woods (eds)

1. Communication Platforms

In this book, we focused on "communication platforms". Platforms have come to define and dominate several areas of social existence, primarily commerce and communication. Ample research has been published about how social media – in particular combined with the use of a smartphone – changed the communication habits of individuals. The accumulation of these individual actions and habits have resulted measurable changes in societies and politics. We were interested in how platforms effect public communication around the world – as opposed to their market and economic effect – and what the regulatory responses have been in different jurisdictions.

In this, we recognise that there is no one agreed legal definition of "communication platform", and that there may be some variations in the scope of services considered to fall within this category from jurisdiction to jurisdiction. Rather, we proceed on the basis that platforms are services that organise and distribute the information-based content of third parties to a potentially large audience. The value added by platforms is the service of content ranking, personal content recommendations, prioritising and deprioritising, and other currently developing services. The latter activity is what makes platforms so powerful in forming the public discourse. This facilitating action is less than the editorial activity of traditional media service providers, but more than mere "dissemination of information" which does not express the potential of influence and manipulation that is inherent nature of the online platform activity.

At the moment, platforms have some very different definitions, provided by legal instruments that approach different aspects of online platforms, and set different aims. Such are the so-called „platform-to-business regu-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

lation“[1] which focused on the commercial angle of online platforms,[2] the draft Digital Services Act,[3] and the draft Digital Markets Act. The German Media State Treaty approaches online platforms from the perspective of public opinion building and defines them (“media intermediary,” Medienintermediär) as an online service[4] that aggregates, selects and presents for the general public among others also journalistic-edited content, without combining them into a complete supply.[5]

2. *Effects of modern platform economy on public communication*

In its influential decision from July 21, 2021, the German Federal Constitutional Court analysed the impact of the modern network and platform economy on the process of public opinion-forming as follows:

---

1 Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (Text with EEA relevance), OJ L 186, 11.7.2019, p. 57.

2 Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (Platform to business regulation), Recital 1.

3 Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. Where “dissemination to the public” means making information available, at the request of the recipient of the service who provided the information, to a potentially unlimited number of third parties. Article 2. (h-i). The draft Digital Markets Act uses the expression „platform services“ without offering a definition. Instead, it lists the „core platform services“ as examples. (Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) COM/2020/842 final

4 So-called „telemedium” which is any electronic information and communication service, unless they are telecommunications services under section 3 no. 24 of the Telecommunications Act, consisting entirely of the transmission of signals via telecommunications networks, or telecommunications-based services under section 3 no. 25 of the Telecommunications Act or broadcasting under sentences 1 and 2. See MStV. § 2. 16. and 13.

5 Somewhat confusingly, the German Media State Treaty also uses the word „Medienplattform”, however, this term applies to streaming services like Netflix or Amazon Prime: „any telemedium insofar as it combines broadcasting, broadcast-like telemedia or telemedia pursuant to section 19 subsection 1 into an overall offer determined by the provider. The combination of broadcasting, broadcast-like telemedia or telemedia pursuant to section 19(1) is also the combination of software-based applications which essentially serve the direct control of broadcasting, broadcast-like telemedia, telemedia pursuant to section 19(1) or telemedia within the meaning of sentence 1.” MStV. § 2. 14.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

"[…]Where services are for the most part financed through advertising, they do not necessarily foster journalistic quality; even on the Internet, the large audiences sought by the advertising industry can only be reached by way of programmes that appeal to the masses. In addition, there is the danger that content can be deliberately tailored to users' interests and preferences, also by means of algorithms, which leads to the reinforcement of the same range of opinions. Such services do not aim to reflect diverse opinions; rather, they are tailored to one-sided interests or the rationale of a business model that aims to maximise the time users spend on a website, thus increasing the advertising value of the platform for its clients. […]

This all leads to increased difficulty in the separation of fact from opinion, content from advertisement, as well as to new uncertainties regarding the credibility of sources and assessments. Individual users themselves must now process and assess the information provided by the mass media, which would traditionally have passed through the filter of professional selection in the spirit of responsible journalism. […]."[6] In conclusion, the Court attributes the described changes in the process of forming public opinion – such as the difficulty in the separation of fact from opinion, new uncertainties regarding the credibility of sources and assessments, new burden on individual users to assess the information provided by the internet and social media – to a business model of the platforms, which is financed by advertising and thus has to generate high attention for the content. Maximising attention is achieved through the use of algorithms that address groups of users based on their behaviour and thus inherently carry the risk of manipulation. While for the one-to-many traditional mass media the financing by advertising and the selection of information through gatekeepers (journalists, publishers, broadcasters) were two distinct functions, for the platforms these two functions are governed by the same tool: algorithms are used to optimise the allocation of advertising, and the allocation of content as well. The logic is the same: to generate maximum attention for advertisements. While this logic also existed previously in the traditional mass media, it is realised at a higher efficiency rate with the new characteristics of platform communication. Some of these characteristics are entirely new, like the vanished entry barrier (see point a. below). Others are old features with an enhanced power.

a. *No gatekeeping:* first, contrary to public communication as we knew it in the 20th century, entry barriers vanished with the emergence

6 BVerfG, Judgment of the First Senate of 20. July 2021 - 1 BvR 2756/20 -, Rn. 1-119, http://www.bverfg.de/e/rs20210720_1bvr275620.html.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

of platforms. Platforms provide a simple and user-friendly interface which allows anyone to publish content even without literacy (e.g., pictures, videos, sound or simply sharing others' content). All content that is published has the potential to reach a global public. In contrast, content that was meant to reach the public had gone through several layers of filtering in the pre-internet age: owners, editors, journalists kept the public communication under their control. The publishing system naturally enforced a certain financial and educational barrier. Platforms have taken over only some of the gatekeeping roles, the extent of this is still under discussion by policymakers, legislators and platforms themselves.[7]

b. *Personal data:* Second, platforms' activity is driven by personal data. As a primary tool to improve their performance, they collect, aggregate and utilise data, for example in order to optimise their ranking, targeting and recommending systems. It is personal data which drives the placement of advertisements, which is our third point.
c. *Attention-driven advertising:* the competition for the audience's attention has always been the goal of public communication. This has also been a widely criticised pitfall of commercial media. The advertisement-financed content offer's main goal was to maximise the number of financially solvent viewers, which, according to the German Federal Constitutional Court, led to a reduction of content quality.[8] Discussing complex topics would have resulted a loss in audience, therefore priority was given to general themes, and easily accessible content.[9] The goal for platforms is the same, but the means to the end, and the consecutive result are different. Polarising themes can be targeted at susceptible audiences. In absence of the entry barriers (see point a.), this becomes a race to the bottom. With the help of algorithms (see in d.) finding the right person for the right content can be perfected, and thereby the attention of users can be exploited in a much more effective way than by traditional commercial media. The format of some platforms leads to shorter communications, which may also be less sophisticated in analysis. Some communication tools, e.g., emojis

7 This gatekeeping role in public communication is not to be confused with the emerging gatekeeping role of platforms in regard of platform communication.
8 BVerfG, Judgment of the First Senate of 20. July 2021 - 1 BvR 2756/20 -, Rn. 1-119, http://www.bverfg.de/e/rs20210720_1bvr275620.html.
9 See among others, for example: McChesney, Robert W. "Corporate Media and the Threat to Democracy", *Penguin Random House*, 1997; Curran, James, "Media and Democracy", *Routledge*, 2011.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

and 'likes', can lead to swift communication, but may lead to many different interpretations and the risk of misunderstanding. A further concern is that the constant strive for positive feedback (likes, upvotes and other signs of public approval) affects the types of content produced; research suggests that users are more likely to share sensational disinformation than truthful content.[10] With the ubiquitous presence of social media through our smartphones, this brings about the problem of information overconsumption. Attention is becoming a scarce resource and not all users are capable to manage it wisely.

d. *Algorithms and AI:* the governance of content distribution, personal data aggregation, advertisement auctions, targeting, recommending, ranking and many more actions on which social media is built, would not be possible without algorithms and AI solutions. Automation is also applied in content moderation, although human supervision appears still inevitable in that regard. AI has also appeared as „artificial users", social bots, ad bots, pol (political) bots and trading bots,[11] which are potential influencers of public communication trends. In this sense AI has the potential to manipulate public opinion building.[12]
e. *Concentration:* finally, the public communication sphere is dominated by some giant companies. There are thousands of small companies, but a small number of large companies hold the biggest market shares. The reason for this development are network effects of communication platforms.[13] This phenomenon has been deeply analysed in the

10 Vosoughi, Soroush, Roy, Deb, Aral, Sinan. „The spread of true and false news online." *Science* 09 Mar 2018. Vol. 359, Issue 6380, pp. 1146-1151. DOI: 10.1126/science.aap9559. See also: Islam, A., Laato, S., Talukder, S., & Sutinen, E. (2020). Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective. Technological forecasting and social change, 159, 120201. https://doi.org/10.1016/j.techfore.2020.120201.

11 Caprolu, Maruantoni, Cresci, Stefano, Raponi, Simone, Di Pietro, Roberto. „New Dimensions of Information Warfare: The Economic Pillar—Fintech and Cryptocurrencies." *Risks and Security of Internet and Systems: 15th International Conference, CRiSIS 2020, Paris, France, November 4-6, 2020, Revised Selected Papers*. Springer Nature, 2021. p. 3.

12 BVerfG, Judgment of the First Senate of 20. July 2021 - 1 BvR 2756/20 -, Rn. 1-119, http://www.bverfg.de/e/rs20210720_1bvr275620.html.

13 Recital 55 Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC. See also: Gillespie, Tarleton. "Content Moderation, AI, and the Question of Scale." Big Data & Society, (July 2020). https://doi.org/10.1177/2053951720943234. Ofcom, "Use of AI in online content moderation". 2019 Report. https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

telecommunications sector.[14] The platform giants have unprecedented numbers of users (e.g., 2,85 billion for Facebook in March 2021[15]) and impressive profit rates (30-40% 2020-2021[16]). On the one hand, this audience reach is theoretic: typically, not all users see the same content. How many people see a certain piece of content is defined by many factors. However, with the help of algorithms and the available personal data, the platforms are in the position to influence this reach. On the other hand, the power of the giant companies is unprecedented in public communication – in comparison with traditional media companies –, with substantial consequences on their lobbying power against regulatory initiatives. Besides, these few giant platforms increasingly act as gatekeepers between business users and end users, and the misuse of their dominant position can be suspected.[17] The significant difference in power between small and large platforms justifies the differentiated treatment of large platforms, as it is envisaged in the draft Digital Services Act.

3. *Platform harms*

The impact that platforms exercise on public communication, cannot be easily categorised. All induced changes carry elements that can be evaluated positively or negatively. The circumstances and the context define whether a certain way of usage causes positive or negative effects for a certain individual, or a group of people.[18] For example, the spread of conspiracy theories in the context of the Covid-19 pandemic is celebrated

---

cambridge-consultants-ai-content-moderation.pdf. See further: Bradshaw, S. (2019). Disinformation optimised: gaming search engine algorithms to amplify junk news. *Internet Policy Review,* 8(4). http://dx.doi.org/10.14763/2019.4.1442.

14 Kühling, Jürgen, Schall, Tobias, Biendl, Michael, "Netzwerkeffekte ausführlich dargestellt", Telekommunikationsrecht, no. 2, 2014, Pages 50-53.

15 Statista, *Number of monthly active Facebook users* worldwide, https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

16 Facebook Profit Margin (Quarterly): 36,29% for March 31, 2021. YCharts. https://ycharts.com/companies/FB/profit_margin.

17 European Commission, "Antitrust: Commission opens investigation into possible anticompetitive conduct of Facebook". 4 June 2021. https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2848.

18 For example, the possibility of anonymous content sharing and getting connected to like-minded people brought up the me-too movement, and helped victims of stigmatising crimes to speak and find support. The same features which help

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

by some as an expression of their freedom of expression, while others see it as a danger to themselves and to public health. When a piece of content is clearly criminal, like child pornography or terroristic content, there is often a broad consensus in society and across legal instruments that it should be removed. However, it is not always easy to evaluate, whether content is criminal, illegal under another law, or legal. For example, defamation is criminal in some states and a civil wrong in others; beyond that, it is contextual and its evaluation might depend on several defences.[19] Hate speech can be used to cover a vast swathe of comment – from mere slurs at one end of the scale to incitement to genocide at the other.[20] Within this range, the placement of the boundary for criminal offences may lie at different points in different states. Moreover, it is context-dependent in most cases.

The example of the draft Digital Services Act illustrates some of the difficulties in this area. Article 2(g) of the draft Digital Services Act speaks about 'illegal content' meaning "any information, which, in itself or by its reference to an activity, including the sale of products or provision of services is not in compliance with Union law or the law of a Member State, irrespective of the precise subject matter or nature of that law". Recital 12 DSA further clarifies that the term 'illegal' is a broad one. It may refer to information, that under the applicable law is either itself illegal, or which relates to activities that are illegal.[21] For a digital service provider, it could be difficult to judge, whether they should block or delete information from a platform due to "illegality" – especially because the "illegality" might also differ in content from one Member State to another. The removal raises technical and procedural questions that impact users' rights, such as notification of the content provider, and the possibility to put back the content.[22] Ultimately a court should decide

---

marginal groups to organise themselves also foster political extremism, hate speech or hate crime.

19 Law Commission of Ontario, *Defamation Law in the Internet Age*, March 2020 at 77, <www.lco-cdo.org/wp-content/uploads/2020/03/Defamation-Final-Report-Eng-FINAL-1.pdf> (accessed 15 July 2021).

20 United Nations Strategy and Plan of Action on Hate Speech, May 2019 at 12, https://www.un.org/en/genocideprevention/hate-speech-strategy.shtml (accessed 13 August 2021).

21 See Rec 12 of the draft DSA that makes clear that the term ‚illegal' is a broad one.

22 See also *Glawischnig-Piesczek v Facebook Ireland Limited* (C-18/18), where the CJEU judged that EU law does not preclude a host provider from being ordered to remove identical and, in certain circumstances, equivalent comments previously declared to be illegal.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

about illegality, while supervision authorities and agencies may play a role in notifying relevant illegal content and ordering their removal.[23] Jurisdictional differences complicate the picture for the global platforms.

The position as regards other than criminal content is even more complex. Some information could be termed 'illegal' but not criminal, because (depending on jurisdiction) they are contrary to other types of law: e.g., misleading advertising. And, a wide range of information can be termed 'harmful', i.e., content that does not trigger a legal response outside the platform environment. This last category of content might still be dealt with by platforms enforcing their community standards. Increasingly, however, there are concerns about content (e.g., COVID denial) that in offline context is potentially harmful but has little opportunity to spread, however, in a platform environment is accessible to a large audience and in many cases actively promoted by the platform systems. The problem occurs from the interplay between the content and the platform's distribution system (and their features (a)-(e) noted above). This has led to suggestions that the "online ecosystem" as such should be regulated (see more on this below). Where human rights are in issue – as here – it must be remembered that state measures must always be specific and proportionate. Legal measures may only exist if there are legitimate reasons (such as the protection of minors, fairness in business transactions, protection of reputation). It is therefore advisable to determine precisely which online harms require which countermeasures for which reasons.

It is also important to take into consideration, that according to the practice of the ECtHR, freedom of expression also applies to expressions that offend, shock or disturb, including untrue facts. Therefore, all limitations to freedom of expression have to be construed strictly, and the need for any restriction must be established convincingly.[24] However, states have positive obligations to ensure protection of privacy, and also the chances of a plural information environment. According to the case law of the ECtHR, it may be justified to restrict expression for these purposes, e.g., ECtHR upheld restrictions against misleading advertisements,[25] and

23 See also: Advocate General's Opinion in Case C-401/19, Poland v Parliament and Council. https://curia.europa.eu/jcms/upload/docs/application/pdf/2021-07/cp210138en.pdf.

24 For example, *Hertel v Switzerland* App no 25181/94 (ECtHR, 25 August 1998) Reports of Judgments and Decisions 1998-VI; *Steel and Morris v the United Kingdom* App no 68416/01 (ECtHR, 15 December 2005) ECHR 2005-II; *Stoll v Switzerland* [GC] App no 69698/01 (ECtHR, 10 December 2007) ECHR 2007-V;.

25 Hertel v Switzerland (2002) App. No. 53440/99, 17 January 2002.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

against a campaign by the Raëlien Movement which fostered believes that life on Earth was created by extraterrestrials, among others.[26]

The high volume of hate speech and disinformation is currently seen as shaking the foundations of our democracies. Political disinformation has been seen capable to influence elections, induce riots, lynching, mobbing, and even genocide.[27] Health disinformation may cost lives and hamper the defence against deadly diseases.[28] Hate speech and harassment against vulnerable groups intimidate their victims and have induced violent attacks against several of them.[29] Beyond the actual harms in the life and safety of the victims, the mentioned content has been causing fissures in the social cohesion and the functioning of democracy.[30] Truth and trust have become concepts that we have become unable to authentically identify.

The low entry barrier into public communication enabled by social media opened the possibility for masses of people to let their voice heard, and react to others. In a quest for popularity, some leaders target people from whom they hope the widest support. Vulnerable minorities are easy target points for both as scapegoats (inciting other users to attack them) and as targets of misleading advertising. Lacking editorial responsibility, and

---

26 *Mouvement Raëlien Suisse* v *Switzerland* (2013) 56 EHRR 14 para 62.

27 116th Congress Senate Report of the Select Committee on Intelligence US Senate on Russian Active Measures Campaigns and Interference in the 2016 US Election. Volume 2. Russia's Use of Social Media With Additional Views. https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.; House of Commons, Digital, Culture, Media and Sport Committee. Disinformation and 'fake news': Final Report. 18 February 2019. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.; UN Human Rights Council Report of the independent international fact-finding mission on Myanmar. A/HRC/39/64. https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf.

28 Bayer, Judit, Holznagel, Bernd, Lubianiec, Katarzyna,, Pintea, Adela,, Schmitt, Josephine B, Szakács, Judit, Uszkiewicz, Erik, (2021) Disinformation and propaganda: impact on the functioning of the rule of law and democratic processes in the EU and its Member States - 2021 update. EP/EXPO/INGE/FWC/2019-1/LOT6/R/07.

29 Bayer, Judit and Bárd, Petra: Hate speech and hate crime in the EU and the evaluation of online content regulation approaches. Study for the European Parliament, Policy Department C: Citizens' Rights And Constitutional Affairs. 2020. ISBN 978-92-846-6902-8 | doi:10.2861/28047.

30 McKay, Spencer and Chris Tenove. "Disinformation as a Threat to Deliberative Democracy." Political Research Quarterly, (July 2020). https://doi.org/10.1177/1065912920938143. ; See also: Luttrell, Regina - Xiao, Lu – Glass, Jon (eds), "Democracy in the Disinformation Age. Influence and Activism in American Politics." Routledge. 2021.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

pillars of truth, spreading of populistic disinformation becomes easy. Technological capacity allows the amplification and manipulation of messages, e.g., through the use of bots, trolls, disinformation networks or deep fakes. The attention-based advertising model advantages sensational content and disadvantages rational presentation of facts. In our days, this mechanism defines all public communication, including political communication.

The new social gap appears along the lines of rational thinkers and believers. On the one hand, it is important that all citizens feel represented in democracies, and all people have the right to believe and think what they do. On the other hand, social functioning cannot be based on false facts and conspiracy theories. All the freedom of expression theories have been based on the presumption that people are rational human beings and that in an open discussion, truth will prevail.[31] A minority of extremists can and should be tolerated by the majority, and their contest of ideas is supposed to lead to better solutions. However, if more than a small minority follows extremist ideas, that is bound to disrupt the functioning of democracy. Thus, the challenge of our age is to turn the tide: to reduce the spreading of false beliefs and conspiracy theories without prohibiting them and without stigmatising the people who believe in them. The goal should be to reduce their representation to a level which is tolerable in a constitutional democracy.

Therefore, the action ground ought to be the distribution logic of this platform-based public communication system, rather than fighting against certain content or the people who like and share them. One way could be that certain rules and conditions were amended so that verified information has better chances to be accessed than disinformation. But which rules and conditions would those be, and how would the truthfulness of information be verified in a rapid communication environment?

The possibilities offered by the rapidly developing platform technology are complex for legislative policy making. Legislation takes years to get finalised, and the development rushes by. Freedom of pursuing business, and other freedoms are also factors to be respected. Against this background, there are strong forces in Europe to develop a counterweight, safeguarding diversity and providing guidance in the post-truth information environment. As the German Constitutional Court argues, public broadcasters are even becoming more significant in "times of increased complex information on the one hand and one-sided representations, fil-

31 Mill, John Stuart: On Liberty. Boston. 1863. p. 50-58.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

ter bubbles, fake news, deep fakes on the other".[32] The British regulator Ofcom, similarly, called for updating the system of public broadcasting.[33] An in-depth consultation has been pursued exploring the possibilities of how to adapt the system to the changing informational environment.[34] Meanwhile, the Finnish government proposed a bill, limiting the Finnish Broadcasting Company (Yle) to publish longer texts only in support of video or audio broadcast, rather than independently.[35] The move is to preserve fair competition between commercial media and Yle. According to the director of Yle, the change can also foster reform and strengthening of Yle.[36] Self-regulation appeared to be a route that builds on the know-how of those who best understand what platforms are able to do: platforms themselves. Platforms did indeed large efforts to introduce measurements in their communication systems to reduce the visibility of disinformation and hate speech.[37] The assessment of the self-regulation showed that the efforts were partly successful, but they were diverse across platforms and countries and also incalculable. The European Regulators' for Audio-visual Media Services emphasised the inconsistent application and the insufficiency of the oversight mechanism.[38]

---

32 BVerfG, Judgment of the First Senate of 20. July 2021 - 1 BvR 2756/20 -, Rn. 1-119, http://www.bverfg.de/e/rs20210720_1bvr275620.html. For further reference, see: report of the Enquête Commission on Artificial Intelligence of the German Parliament (Bundestag) of 28 October 2020, BTDrucks 19/23700, p. 447 ff.).

33 Ofcom „Ofcom calls for stronger system of public service media fit for the digital age." July 15 2021. https://www.ofcom.org.uk/about-ofcom/latest/media/media-releases/2021/stronger-public-service-media-system-for-digital-age

34 Ofcom „Small Screen, Big Debate. Consultation. The Future of Public Service Media." December 8. 2020. https://www.smallscreenbigdebate.co.uk/__data/assets/pdf_file/0032/208769/consultation-future-of-public-service-media.pdf

35 Yle „Gov't aims to limit Yle web publications." June 16. 2020. https://yle.fi/uutiset/osasto/news/govt_aims_to_limit_yle_web_publications/11405119

36 Ibid.

37 Washington Post "Facebook says it has taken down 7 million posts for spreading coronavirus misinformation. The company also labeled 98 million posts with warning notices about coronavirus misinformation between April and June." August 11. 2020. https://www.washingtonpost.com/technology/2020/08/11/facebook-covid-misinformation-takedowns/. More recently the Centre for Countering Digital Hate's report, Failure to Protect, https://www.counterhate.com/failuretoprotect, suggests that 84% of antisemetic posts were not taken down.

38 European Commission Staff Working Document, 'Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement', SWD(2020) 180 final, p. 22.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Under the current scheme, platforms have a substantial income from the spreading and flourishing of disinformation and extreme content. A systematic restructuring of the communication patterns would result that platforms lose part of their revenues, unless they also restructure their income base. Considering the stellar profits that giant platforms make, obviously there is ample room for manoeuvre in this area. But expecting that platform companies would proactively cut their own profit appears reasonable only if they are given clear expectations with the possibility of enforcement. The worldwide attempts that are reflected in this book, to draft some kind of control on social media platforms, can be interpreted as a signal that the time is ripe for this move.

4. How to deal with the harms?

This book highlights several snapshots of legal approaches and instruments which aim at dealing with the dangers caused by online platforms. Some states are dealing with issues through the lens of data protection (e.g. Russia), or by focussing on market regulation (to some extent, the USA). Both these in some way relate to the business model of the platforms, as do proposals that focus on the design of the platforms. There are also approaches that focus on content regulation. Below, we attempt to typify and order the approaches that we have encountered during the project.[39]

a. Defining a general duty of care standard

We know now that the inherent structure of platform communication carries the risk of distorting the social discourse. The individual violations of rights cannot reflect accurately the systemic distortion of the communication scene. Addressing only the individual violations of law, or even the individual pieces of harmful content, will not change the systemic harms. While these systemic harms may be indirect, their effects are more than

39 The workshop series on Hate speech and platform regulation included seven workshops and several speakers who were not included in the volume. Report on the project in Bayer, Judit, Kalbhenn, Jan, „Masse und Macht – Auf der Suche nach Regeln für digitale Kommunikationsplattformen“, ZUM – Zeitschrift für Urheber- und Medienrecht, No. 4 (2021).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

subtle[40] and threaten the operation of democracies, the basic foundation of which is free, but also rational, discourse on common matters. Therefore, there is good reason to view platforms, especially social media platforms as systems which carry an inherent systemic risk, like rail, automobile, or powerplants. Their operators should be aware of the risks and do all necessary efforts in their competence to minimise those risk.

This approach is followed by the United Kingdom's „Duty of Care" principle found in the draft Online Safety Bill, and the EU's draft Digital Services Act's „risk assessment" obligation, with some meaningful differences. First, as opposed to the scheme in relation to rail and automobiles, platforms' liability for the damage caused through the platform as a vehicle, is generally exempted. They do not bear direct liability to cover the losses caused by illegal content, for example. But they still might be made *responsible* by law, to design and apply the preventive measures to minimise the risk.[41] The UK model includes an obligation to *design a system* that allows for content notified as illegal to be taken down swiftly.[42] As regards children, platforms are under an obligation to mitigate and manage the risks of harm and in some instances, using system design, prevent children from coming into contact with specified types of content.[43] Although the UK has left the EU, it currently maintains the immunity provisions derived from the e-Commerce Directive. The interplay between the two sets of provisions is not yet known. Whereas, the EU's draft Digital Services Act strictly orders removal of illegal content when platforms are notified of them, as a condition of their exemption from liability. The "due diligence" obligations apply to all other issues, including procedures, transparency, dealing with harmful content, and more.

---

40 Ibbetson, Connor "Where do people believe in conspiracy theories?" *YouGov Cambridge Globalism Project*. 18 January, 2021. https://yougov.co.uk/topics/international/articles-reports/2021/01/18/global-where-believe-conspiracy-theories-true.

41 On delineation of liability and responsibility see: Chapter 5.1. by Sarah Hartmann in: Bayer, Judit, Katsirea, Irini, Batura, Olga, Holznagel, Bernd, Hartmann, Sarah, Lubianiec, Katarzyna. The fight against disinformation and the right to freedom of expression. (Brussels: European Parliament, 2021) p. 59-63. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU(2021)695445_EN.pdf. The draft Online Safety Bill imposes an obligation to operate a system that minimises the presence of illegal content and to mitigate against the likelihood of children encountering content of a type assessed to be harmful to them.

42 Clause (3)(d) draft Online Safety Bill.

43 Clause 10(2) and (3) draft Online Safety Bill.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

b. Duty of standards in specific areas (sectors)

In the broader picture, it is important to note that the topics that need to be assessed as systemic risks by the platform providers, are manifold. Some of these topics are also regulated by separate acts, such as the use of artificial intelligence, data protection, the protection of children and advertising. Others may be regulated by some states, but are often left to self-regulation, such as hate speech and disinformation. However, recently, many countries took countermeasures in this area. For example, the Canadian government has left behind its reservations to regulate internet communication and has introduced a comprehensive law against hate speech. Intensive discussions are also taking place in other countries, on how effective action can be taken in particular against hate speech and disinformation, as it was reflected in our workshops (in Japan, Singapore, India, etc.).

A further systemic risk which is not left to self-assessment and self-regulation, but falls entirely in the realm of state regulation, is market concentration of platform operators and the risk to the fair economic competition.[44] In the USA in particular, there is intensive discussion about whether limiting the economic power of the large platforms could be an important prerequisite not only for more competition, but also for effectively combating hate communication and disinformation. There is also intensive discussion here about whether interoperability obligations, as we know them from telecommunications law, can contribute to increasing the number of communication platforms such as Facebook.[45] There is a more general concern about the operation of competition law with regard to the super-dominant tech companies. The Digital Markets Act envisages special rules to be applied to digital gatekeepers; the UK is considering similar measures. Germany[46] has already passed a new law. The common

44 The draft Digital Services Act approaches the areas of risk assessment from the perspective of activities: besides (a) dissemination of illegal information, it relates to (b) any negative effects for the exercise of the fundamental rights to respect for private and family life, freedom of expression and information, the prohibition of discrimination and the rights of the child; and (c) the intentional manipulation and exploitation of their service, with an actual or foreseeable negative effect on the protection of public health, minors, civic discourse, or actual or foreseeable effects related to electoral processes and public security.

45 See more in Chapter 1.4. Policy Developments in the USA to Address Platform Information Disorders by Sarah Hartmann.

46 Section 19a Kartellgesetz 2020. (German Competition Act).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

theme is that there is some element of ex ante obligations, an approach which may have been borrowed from the telecommunications regime.

c. *Enforcing the duty of care standard: self-regulation, co-regulation or state supervision*

This leads us to the second question: provided that the industry would design the measures to mitigate the harms, and given that the actors are not directly liable for direct harms, how is the efficacy of the system supervised and enforced? The direct harms are not the liability of platforms, and the indirect harms are not measurable – how to ensure that the risk minimising measures are successful? In the United Kingdom, this task would be allocated to Ofcom, the converged authority for media, telecommunication and post.[47] As a regulator, it would be competent to supervise and impose orders on the platform operators. The European Union's draft Digital Services Act provides for a complex set of supervisory and compliance measures. National competent authorities, with designated Digital Services Coordinators would have wide powers to investigate, seek information and impose orders, as well as penalties. In case the national procedure is insufficient (cases defined precisely in the Act), the European Commission may exercise delegated powers to investigate and enforce the Regulation and the relating decisions. Critiques find that authorities' role is exaggerated, in view of the freedom of expression standards which require access to courts. Yet, the Council of Europe has also recommended the establishment of regulatory authorities in the context of the broadcasting sector, albeit with strong emphasis on the independence of such authorities.[48] Provided that ultimate judicial review of the authority's decision remains possible, authorities may need to play a proportionate role in justified restrictions of freedom of expression rights also in the platform environment, considering the abundance of content which would otherwise overload the judicial system. The draft Digital Services Act provides for a set of "due diligence" obligations, which aim at different public policy objectives such as the safety and trust of the recipients of the service, including minors and vulnerable users, protecting the relevant

47 See Chapter 1.3. by Lorna Woods in this book.

48 Recommendation Rec(2000)23 of the Committee of Ministers to member states on the independence and functions of regulatory authorities for the broadcasting sector.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

fundamental rights enshrined in the Charter, empowering recipients and other parties.[49] Some of the rules under Chapter III. which sets out the due diligence obligations, are formulated strictly, such as the notice and action procedure (dealing with illegal content) and the transparency obligations.[50] Also, several of the obligations that apply to very large online platforms (Chapter III. Section 4) are straightforward and easily controllable, such as the transparency of recommender systems, additional online advertising transparency, data access, appointing compliance officers, and the further transparency obligations.[51] However, in the case of other obligations, checking adequate compliance may be a complicated endeavour. Such is the obligation to identify, analyse, assess the risks; to put in place reasonable, proportionate and effective mitigation measures; and to have an independent audit.[52] The exact content of these expectations from very large online platforms is left open, to be developed by the industry actors themselves, in particular in the Code of Conduct, and in the Advertising Code of Practice. No enforcement measures are planned in relation to the envisaged code of practice (and the advertising code). There is no clear provision on whether the Coordinator can decide if the measures taken to mitigate the risks are insufficient. This is supposed to be established by the independent audit. However, a negative audit report entails nothing more than the obligation to justify the reasons for not implementing the operational recommendations – and setting out „any alternative measures they may have taken to address any instances of non-compliance identified". At this stage, it is unclear whether the Digital Services Coordinators would have the power to declare that the operator did not adequately justify the reasons for not implementing the recommendations.

Digital Services Coordinators may start their procedure only in case of an infringement of the rules of the Regulation. In that case, they may adopt a decision on the infringement, and request the platform to draw up an action plan. The Digital Services Coordinator may then decide whether the action plan is appropriate, and it may request the platform to subject itself to an additional independent audit, with an appointed auditor.

In sum, the enforcement system of the draft Digital Services Act is very carefully designed, and sets out considerable fines in case of violation of the Act, but the Act defines only the basic obligations of platforms,

49 Recital 34-35. Digital Services Act.
50 Article 10-24. Digital Services Act.
51 Articles 29-33.
52 See Section 4 of Chapter III. See also Section 50. (1).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

whereas many details are referred to self-regulation. The self-regulatory codes are passed under the supervision of the European Commission, but the consequences of non-compliance with the Codes are not clarified in the Act.

d. Supervision: allocating competences between competent authorities

When it comes to the supervision and regulation of these – often overlapping – areas of systemic risks, the question of allocating competences between competent authorities emerges. It can already be observed that in particular the data protection authorities, the cartel authorities and the media and telecommunications regulators are arguing about who should be responsible for combating online harms. In the EU, in addition to this problem of horizontal distribution of supervisory responsibilities, there is also the problem of vertical distribution of competences between the European Commission or EU agencies, and the national authorities.

In the UK, Ofcom has responsibility for the range of communications industries, including now video sharing platforms and, when the draft Online Safety Bill comes into force, other social media platforms. It nonetheless needs to work with other regulators – notably the Competition and Markets Authority (dealing with competition and consumer protection), the Information Commissioner's Office (responsible for data protection and freedom of information) and even the Financial Conduct Authority (the financial services regulator). To do this, the Digital Regulators Cooperation Forum has been established. It remains to be seen how effective it will be. Extensive discussions within each jurisdiction are likely to be needed to develop an effective supervisory model.

5. Final remarks: do we need a global regulation?

Many of our expert authors have expressed the view that national regulation is not expected to be successful against the actions of global online platforms. It has even been noted that actual notices and requests by state authorities have been seen to be ignored by giant companies. This leads us to ask whether transnational regulation or international rules would deal with the mentioned social, individual and economic problems more efficiently. However, this has some obstacles. First, national legal frameworks are different, especially when it comes to content regulation.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The draft Digital Services Act plans to overcome this difficulty with the transnational hub of the Digital Services Board and the Commission – a scheme that has been applied in the General Data Protection Regulation before. But still, the evaluation on what is "illegal" and what is permitted, would be defined through national regulations. It may, however, be that regulation of the distribution of content (systems or ecosystem regulation) is marginally less contentious than direct content regulation. Second, the globe is divided in major attitudes towards regulation. China or Russia have vastly different standards than the United States, with Europe and other continents being also divergent. Inspite of these hindrances, there is some hope to come to common denominators provided there is an intention to do so. There are some soft law initiatives being developed at the international level (e.g., OSCE Guidance on AI in content moderation). A regional cooperation between democratic states would be possible and also desirable. Currently, all states appear to keep their eyes on other states, watching what those are initiating to tackle the problems which press so many societies worldwide. Therefore, there is considerable responsibility on the European Union and those states which lay the groundwork for a new regulatory regime.

Bibliography

116th Congress Senate Report of the Select Committee on Intelligence US Senate on Russian Active Measures Campaigns and Interference in the 2016 US Election. Volume 2. *Russia's Use of Social Media With Additional Views*. Washington D.C.: 116th Congress Senate. (2019). https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf.

Bayer, Judit, Bárd, Petra. *Hate speech and hate crime in the EU and the evaluation of online content regulation approaches*. Brussels: European Parliament, Policy Department C: Citizens' Rights And Constitutional Affairs, 2020. doi:10.2861/28047.

Bayer, Judit, Holznagel, Bernd, Lubianiec, Katarzyna,, Pintea, Adela, Schmitt, Josephine B, Szakács, Judit, Uszkiewicz, Erik. *Disinformation and propaganda: impact on the functioning of the rule of law and democratic processes in the EU and its Member States - 2021 update*. Brussels: European Parliament, 2021. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653633/EXPO_STU(2021)653633_EN.pdf.

Bayer, Judit, Kalbhenn, Jan, *Masse und Macht – Auf der Suche nach Regeln für digitale Kommunikationsplattformen*. ZUM – Zeitschrift für Urheber- und Medienrecht, No. 3 (2021): 185-194.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Bayer, Judit, Katsirea, Irin, Batura, Olga, Holznagel, Bernd, Hartmann, Sarah, Lubianiec, Katarzyna. *The fight against disinformation and the right to freedom of expression*. Brussels: European Parliament, 2021. https://www.europarl.europa.eu/RegData/etudes/STUD/2021/695445/IPOL_STU(2021)695445_EN.pdf.

Bradshaw, S. "Disinformation optimised: gaming search engine algorithms to amplify junk news." *Internet Policy Review*, 8(4) (2019). http://dx.doi.org/10.14763/2019.4.1442.

Caprolu, M., Cresci, S., Raponi, S., Di Pietro, R. "New Dimensions of Information Warfare: The Economic Pillar—Fintech and Cryptocurrencies." *Risks and Security of Internet and Systems: 15th International Conference, CRiSIS 2020, Paris, France, November 4-6, 2020, Revised Selected Papers*. Springer Nature (2021): 3.

Centre for Countering. "Digital Hate's report, Failure to Protect." Accessed August 12, 2021. https://www.counterhate.com/failuretoprotect.

Curran, James. *Media and Democracy*. Routledge. (2011).

European Commission. *Antitrust: Commission opens investigation into possible anti-competitive conduct of Facebook*. Brussels: European Commission, 4 June 2021. https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2848.

European Commission. *Assessment of the Code of Practice on Disinformation – Achievements and areas for further improvement*. Brussels: European Commission, 2020. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=69212.

Gillespie, T. "Content Moderation, AI, and the Question of Scale." *Big Data & Society* (July 2020). https://doi.org/10.1177/2053951720943234.

House of Commons, Digital, Culture, Media and Sport Committee. *Disinformation and 'fake news': Final Report.* London: Hose of Commons, Digital Culture, Media and Sport Committee, 2019. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf.

Ibbetson, C. "Where do people believe in conspiracy theories?" *YouGov Cambridge Globalism Project.* (January 2021). https://yougov.co.uk/topics/international/articles-reports/2021/01/18/global-where-believe-conspiracy-theories-true.

Islam, A., Laato, S., Talukder, S., & Sutinen, E. "Misinformation sharing and social media fatigue during COVID-19: An affordance and cognitive load perspective". *Technological forecasting and social change*, 159, 120201. (2020). https://doi.org/10.1016/j.techfore.2020.120201.

Kühling, Jürgen, Schall, Tobias and Biendl, Michael. "Netzwerkeffekte ausführlich dargestellt" *Telekommunikationsrecht* no. 2 (2014): 50-53.

Law Commission of Ontario, *Defamation Law in the Internet Age*, Ontario: Law Commission, 2020. www.lco-cdo.org/wp-content/uploads/2020/03/Defamation-Final-Report-Eng-FINAL-1.pdf.

Luttrell, Regina, Xiao, Lu, Glass, Jon (eds). "Democracy in the Disinformation Age. Influence and Activism in American Politics." *Routledge.* (2021).

McChesney, Robert W. *Corporate Media and the Threat to Democracy*. Penguin Random House, 1997.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

McKay, S., Tenove, C. "Disinformation as a Threat to Deliberative Democracy." *Political Research Quarterly*, (July 2020). https://doi.org/10.1177/106591292093814 3.

Mill, John Stuart. *On Liberty*. Boston: Tricknor and Fields. (1863).

Ofcom, "Use of AI in online content moderation". (2019). https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf.

Statista. "Number of monthly active Facebook users worldwide." Accessed July 29, 2021. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

UN Human Rights Council. *UN Human Rights Council Report of the independent international fact-finding mission on Myanmar*. Geneva: UN Human Rights Council, 2018. https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_64.pdf.

Vosoughi, S., Roy, D., Aral, S. "The spread of true and false news online." *Science* , Vol. 359, Issue 6380 (March 2018): DOI: 10.1126/science.aap9559.

Washington Post "Facebook says it has taken down 7 million posts for spreading coronavirus misinformation. The company also labeled 98 million posts with warning notices about coronavirus misinformation between April and June." *Washington Post*, August 11, 2020. https://www.washingtonpost.com/technology/2020/08/11/facebook-covid-misinformation-takedowns/.

YCharts. "Facebook Profit Margin (Quarterly): 36,29% for March 31, 2021.". Accessed July 21, 2021. https://ycharts.com/companies/FB/profit_margin.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

The Authors and Editors

Alejandro Aréchiga Morales holds a law degree from the University of Guadalajara, Mexico, and a Magister degree from University of San Andrés, Argentina, where he completed the Joint Masters Program in Intellectual Property and Innovation (MIPI). His research focuses on Author's Rights in Latin America, freedom of expression, non-conventional trademarks, among other subjects related with Intellectual Property. Currently, he practices law in Jalisco, Mexico, specializing in Intellectual Property.

Alexandre Alaphilippe is the Executive Director and co-founder of the EU DisinfoLab. He has coordinated work on some of the organisation's largest investigations into Information Operations linked to Russia, India and China, and is a member of a number of working groups in Brussels and Washington DC linked to platform regulation, transatlantic relations, and hybrid threats, where he emphasises the role of civil society in maintaining democratic values. He has published papers for the Brookings Institution and his work has been featured on CNN, BBC, Le Monde and Politico. Prior to founding the EU DisinfoLab, Mr. Alaphilippe worked in consultancy and served for four years as chief digital officer at the Clermont-Ferrand city hall. He holds a degree in Communications.

Ang Peng Hwa is professor at the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. His teaching and research interests combine law and communication, touching on internet law and policy, censorship, and the social impact of media. He is the author of *Ordering Chaos: Regulating the Internet* (Thomson, 2005). He co-founded the Global Internet Governance Academic Network and the Asia Pacific Regional Internet Governance Forum and served as inaugural chairs of both organisations.He has been a visiting Fulbright scholar at Harvard University and a visitor at Oxford University.

Ashwini Siwal holds Doctorate in Data Security Laws & LL.M. (Science, Technology & Law). He specializes in Intellectual Property Laws, Information Technology Laws and Anti-Trust Laws. He has research interests in technology & law. He is the Director of "LC-II Entrepreneurial Law

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Clinic" at Faculty of Law, University of Delhi. He frequently delivers talks at prestigious institutions like Central Bureau of Investigation Training Academy , India and other training centers on IP Laws, IT Laws and Anti-Trust Laws. He has developed "Survey course on IPR" which is available on CEC YouTube Channel: https://www.youtube.com/playlist?list=PLNsppmbLKJ8Kc0PffwRdrafC1ptO1oScO.

Bernd Holznagel is Professor for Public and Administrative Law and Director of the Institute for Information, Telecommunication and Media law (ITM) at the University of Münster. The main focus of his research work lies in Telecommunication and Media Law as well as in network regulation, notably energy law. He studied law and sociology at Freie Universität Berlin (Free University of Berlin) as well as at McGill University Montréal. Since 1997 he has been Professor of constitutional and administrative law at the University of Münster and head of the public section of the ITM. Professor Holznagel is a member of an academic consortium for questions of regulation of the Bundesnetzagentur (Federal Network Agency) and a member of Münchener Kreis (Munich Circle). Moreover he is co-editor of the law magazine Multimedia und Recht.

Elda Brogi is a Part-Time Professor at the European University Institute (EUI). Ph.D. in Public Law and Constitutional Law (University La Sapienza, Rome). At the EUI she is Scientific coordinator of the Centre for Media Pluralism and Media Freedom. Member of the Executive Board of the European Digital Media Observatory (EDMO). She teaches Communication Law at the University of Florence. Member of the Council of Europe (CoE) Committee of Experts on Media Environment and Reform (MSI-REF), co-rapporteur on the recommendation on electoral communication and media coverage of election campaigns. She was also member of the CoE MSI-MED and MSI-JO Committees.

Enni Ala-Mikkula, D.Sc.(econ.), acquired her PhD. at the Tampere University. Her research interests focus especially on different questions concerning occupational safety and health regulation and on labor law in general. In her doctoral thesis she studied employer's occupational safety and health responsibility by concentrating on employer's key responsibilities and on the requirements placed on employer's occupational safety and health activities. At the University of Helsinki Ala-Mikkula works as a part of the Hate and public sphere -research team and studies the rela-

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

tion between hate speech and employer's occupational safety and health responsibilities.

Gerard Goggin is Wee Kim Wee Professor in Communication Studies at Nanyang Technological University, where he is co-director of the Asian Communication Research Centre. He has a longstanding interest in communication and media policy, especially in relation to issues of consumer protection, public interest, accessibility and disability, and digital inclusion. Among other areas, Gerard is widely published on mobile communication research, with his most recent book being *Apps: From Mobile Phones to Digital Lives* (2021).

Giovanni De Gregorio is Postdoctoral Researcher working with the Programme in Comparative Media Law and Policy at the Centre for Socio-Legal Studies. His doctoral study has investigated the rise of European digital constitutionalism as a reaction and strategy against the predominance of digital private normativities. His research interests focus on online speech and artificial intelligence; privacy and data protection; digital policy. Giovanni has been Academic Fellow at Bocconi University, non-resident legal research for Columbia Global Freedom of Expression and visiting fellow at the Center for Cyber Law and Policy at the University of Haifa.

Izumi Aizu is Professor and Senior Research Fellow, Institute for InfoSocionomics, Tama University, Tokyo, and a researcher at the Institute for HyperNetwork Society, Oita, Japan. Mr. Aizu specializes the study of Information Society. He promoted PC based communications in the '80s and the Internet in the '90s. He participated in the global policy debate on Internet Governance and Information Society at Internet Cooperation for Assigned Names and Numbers (ICANN), World Summit for Information Society (WSIS) and Internet Governance Forum (IGF) meetings. In 2011, he engaged in the recovery from the East Japan Great Earthquake using ICT (Information and Communication Technology). His recent research focus includes social innovation at large.

Jacob Mchangama is the founder and executive director of Justitia and a visiting fellow at the Foundation for Individual Rights in Education in Washington. In 2018 he was a visiting scholar at Columbia's Global Freedom of Expression Center. He has commented extensively on free speech and human rights, has published in academic and peer-reviewed journals. Jacob is the host and narrator of the podcast Clear and Present

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Danger: a history of free speech and is the recipient of numerous awards for his work on free speech and human rights. Jacob's book 'Free Speech: A History from Socrates to Social Media' will be published by Basic Books in January 2022 (US and UK).

Jan Christopher Kalbhenn, LL.M. studied law and public international law in Osnabrück, Stockholm and Amsterdam. He completed his legal clerkship at the Berlin Court of Appeal with stations at the Berlin State Opera, Germanys Public Radio Broadcaster and the German Publishing Association. From 2014 – 2018 he was in-house lawyer at Deutsche Welle, the international broadcasting station of the Federal Republic of Germany. Since 2018 Jan has been managing director at the ITM under Prof. Dr. Bernd Holznagel and PhD candidate in the field of media law. He is an attorney at law and a lecturer for copyright law, data protection law, media law and cultural law at the University of Applied Sciences Münster.

Jörg Becker is Managing Director of the Institute for Information Systems at the University of Münster (WWU) and Academic Director of the European Research Center for Information Systems (ERCIS). He directs the Chair for Information Systems and Information Management and holds an honorary professorship at the National Research University – Higher School of Economics (NRU-HSE) in Moscow. He served as Vice-Rector for Strategic Planning and Quality Assurance of the WWU (2008-2016). Since 2016, he is Spokesman of the WWU Center for Europe. He is editor of various journals, has published in renowned outlets and authored and edited numerous books. His research fields comprise Information Management, Information Modelling, Retail IS, e-Government, and Business Intelligence.

Judit Bayer is associate professor of media law and international law at the Budapest Business School, Hungary, and at the time of writing this book, Schumann Fellow at the University of Münster. Her research interest is in human rights, freedom of expression, media freedom and pluralism, and privacy. She has a PhD in constitutional law (internet regulation) and habilitation in data protection. Bayer has authored several books and articles in the field of freedom of expression and the media, in particular on the liability of internet service providers, of social media platforms, freedom of expression on the internet, public service broadcasting, and human rights.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Juliya Kharitonova is Doctor of Science (Law), Professor of Legal Science, the Head of the research and education center "Center for Legal Studies of Artificial Intelligence and Digital Economy," Professor of Business Law Department, Law Faculty of Lomonosov Moscow State University (Moscow, Russia). Prof. Kharitonova is a member of the Scientific Advisory Board of the *Moscow Arbitration Court*; a member of the Commission of the Moscow Department of the All-Russian Non-Governmental Organisation "Association of Lawyers of Russia" for the legal regulation of economic activities. Arbitrator of the ICAC under the Chamber of Commerce and Industry of the Russian Federation.

Kilian Müller, M.Sc., studied Information Systems at the University of Münster (WWU) and is currently working as a doctoral candidate at the Chair for Information Systems and Information Management at the University of Münster. As a part of the MODERAT!-Project Kilian Müller studies the possibilities of automated hate speech detection in user-generated content, their applicability within the comment-moderation process, and their acceptance.

Konrad Bleyer-Simon is a research associate at the European University Institute's Centre for Media Pluralism and Media Freedom. He conducted doctoral research at the Human Rights Under Pressure joint program of the Freie Universität Berlin and the Hebrew University in Jerusalem, and holds a Master of International Affairs degree from Columbia University. Prior to working at CMPF, he worked for NGOs and news media in Berlin, Brussels, Bishkek and Budapest. At the CMPF, he works on the European Digital Media Observatory and the Media Pluralism Monitor.

Kristiina Koivukari is a postdoctoral researcher at the University of Helsinki. Her research covers areas such as criminal law, EU law, human rights and legal theory. In the research project on cyberhate (2020-2021) she studies hate speech and online shaming from the criminal law perspective. She has also extensive experience in teaching criminal law and working as a lawyer in different areas of law.

Kuo-Wei Wu From 2004 til 2016, he was the CEO of NIIEPA, consultant firm to develop internet policy for Taiwan government. He was member of the ICANN board from 2010-2016. In the year of 2016, he was appointed as a board member of ChungHwa Telecom – the major Telecom in Taiwan. He holds a BS degree in mathematics from Tunghai University,

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

a MS degree in mathematical science from University of Cincinnati, USA, and another MS degree in computer science from Columbia University of New York, USA. From 1999 until 2010, he was elected to be one of the executive council member of APNIC. In 2008, he was appointed to serve as a board member of Public Interest Registry (PIR) „.org" registry.

Larisa Sannikova, Doctor of Law (2007), Professor of Legal Science (2007), Professor of The Russian Academy of Sciences (2018). She is the Head of the Centre for Legal Research of Digital Technologies, The State Academic University for the Humanities (Moscow). Prof. Sannikova main expertise is in legal regulation of the use of digital technologies such as Blockchain, Artificial Intelligence, etc. She has authored more than 130 peer-reviewed articles and contributed to several textbooks and other works on civil law, banking law and digital law. Prof. Sannikova is a member of the Scientific Advisory Council of the Supreme Court of the Russian Federation; a member of the Financial Markets Commission of the Russian Bar Association; a member of the editorial Board of the journal «Gosudarstvo i pravo». She is an expert of the Russian Science Foundation (RSF) and the Russian Foundation for Basic Research (RFBR). She is editor-in-chief of the journal «Law & Digital Technologies».

Lorna Woods is Professor of Internet Law at the University of Essex and a member of the Human Rights Centre there. She has extensive experience in the field of media policy and communications regulation, including data protection, social media and the Internet, and she has published widely in this area. Her current research project with Carnegie UK Trust is on reducing harm arising on social media and she was awarded an OBE in recognition of her work. She is serving a second term as a member of the ESRC Peer Review College, is a member of the Digital Freedom Fund's Panel of Experts, is a senior associate research fellow at the Information Law and Policy Centre, Institute of Advanced Legal Studies, University of London and a fellow of the Royal Society for Arts.

María Carolina Herrera Rubio is an attorney, holding a law degree from the University of Medellin, where she also attained a postgraduate diploma in Intellectual Property Law. She is a magister candidate in the Joint Masters Program in Intellectual Property and Innovation (MIPI) at University of San Andrés, Argentina. Currently, she practices law in Medellín, Colombia, specializing in Intellectual Property and Commercial Law.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Maria de Lourdes Vazquez is Dean of the Law School at Universidad de San Andres in Buenos Aires, Argentina, where she directs the Joint Masters Program in Intellectual Property and Innovation, and the Center of IP & Innovation. She holds a law degree from Universidad Católica Argentina, an LLM from Harvard Law School, and completed her doctoral studies at the European University Institute in Italy. After serving as in-house counsel to Virgin Records (London) and EMI Records (New York). Maria was partner at the firm Marval O'Farrell & Mairal (Buenos Aires). She was a recipient of the "deFortabat Visiting Scholarship" in David Rockefeller Center for Latin American Studies at Harvard University.

Marten Schultz is professor of private law at Stockholm University, since 2011. Prior to that, professor of private law at Uppsala University. Founder and chairman of the Swedish Law and Internet Institute, a non-profit organisation working on digital human rights. Member of the Swedish National Digitalisation Council, expert for the Swedish Media Council and member of the Crime Victim Compensation Board. Legal commentator for Svenska Dagbladet. Author or numerous articles, government reports and books on freedom of speech, privacy protection and hate speech, on the internet.

Maximilian Hemmert-Halswick studied law at the University of Cologne, with a one year stay in Beijing for Chinese language studies. He worked as a research assistant at the Institute for Information, Telecommunications and Media Law (ITM), University of Münster where he finished his doctoral thesis. At the moment, he serves as the head of the Energy Law working group at the West Coast University of Applied Sciences in Heide.

Michael Geist is a law professor at the University of Ottawa where he holds the Canada Research Chair in Internet and E-commerce Law and is a member of the Centre for Law, Technology and Society.. He was appointed to the Order of Ontario in 2018 and has received numerous awards for his work including the Kroeger Award for Policy Leadership and the Public Knowledge IP3 Award in 2010, the Les Fowlie Award for Intellectual Freedom from the Ontario Library Association in 2009, the EFF's Pioneer Award in 2008, and Canarie's IWAY Public Leadership Award for his contribution to the development of the Internet in Canada.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Mrs. Päivi Korpisaari is Professor in Communication Law at the Faculty of Law, University of Helsinki. Her main fields of research are media and communication law, personal data protection, tort law, constitutional law, human rights law, and criminal law in respect to 'freedom of expression offences'. She is leading a research project relating to legal challenges of 5G network and smart city infrastructure and other project relating to online hate speech. One of her most recent project is on bio-medical law and legal challenges that impede or slow down the provision of new treatments for chemotherapy resistance in high-grade serous ovarian cancer patients. She has been a member in several committees preparing Finnish legislation. She has worked as an attorney and served at the district court and court of appeal before her career at the university.

Natalie Alkiviadou is senior research fellow at Justitia. Her work focuses on free speech, 'hate speech' and the far-right. Her books: 'The Far-Right in International and European Law' and 'Legal Challenges to the Far-Right: Lessons from England and Wales' were published by Routledge. She has several publications in a wide range of peer-reviewed journals on free speech, hate speech and online content moderation. Natalie has worked in higher education for the last eight years, teaching subjects related to International Human Rights Law and European Law. She has worked with civil society, educators and public servants in the framework of training and capacity building on the freedom of expression and related themes.

Nicole Stremlau is Head of Programme in Comparative Media Law and Policy at the Centre for Socio-Legal Studies, University of Oxford and she is Research Professor in the School of Communications at the University of Johannesburg. She leads a European Research Council project on Social Media and Conflict with a focus on Africa. Recent books include Media, Conflict and the State in Africa (Cambridge University Press); Speech and Society in Turbulent Times (ed with M Price) (Cambridge University Press); and UNESCO's flagship publication World Trends in Freedom of Expression and Media Development.

Poren Chiang is a research assistant at Institutum Iurisprudentiae, Academia Sinica (Taipei, Taiwan). He holds an LL.M. from UCLA School of Law, with a specialization in Digital Law and Policy. Simultaneously an active software developer, he has worked with civic tech communities and

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

FOSS initiatives on both legal and technical matters. His research agenda includes electronic voting, digital transformation, open source license, social media regulations, open data policies, and community governance.

Richard Janda is professor for extracontractual obligations, business associations, administrative process and environmental law. A former clerk to Justices Le Dain and Cory of the Supreme Court of Canada, he was also Director of the Center for the Study of Regulated Industries at McGill University. He is currently leading the Myko project (www.myko.org), which explores how to connect everyone to the environmental footprint of their choices in real time. He has written, among other things, on corporate social responsibility, digital law, and theories of justice.

Sarah Hartmann is an academic counsellor and senior research associate at the Institute for Information, Telecommunications and Media Law (ITM) at the University of Muenster, Germany. After studying law in Bremen, Paris and Muenster, she was awarded her doctorate in law for her dissertation on the material scope of the European AVMS directive and regulation in convergent media environments in 2018. Her research focuses on German and European media regulation, human rights and data protection law. She is currently working on her habilitation.

Shun-Ling Chen is an associate research professor and a co-director of the Information Law Center at Institutum Jurisprudentiae, Academia Sinica (Taipei, Taiwan). She holds an S.J.D. from Harvard Law School. She has worked closely with online collaboration and civic tech communities. Trained in both Law and Science and technology studies, her research interest include copyright, privacy, free speech, peer production, citizen science, community governance and platform governance.

Trisha Meyer is an Assistant Professor in Digital Governance and Participation at the Vrije Universiteit Brussel and a Professorial Fellow at the United Nations University CRIS in Bruges. Trisha researches the regulatory push toward and societal consequences of tech platforms taking proactive (and automated) measures to moderate content. In this context, she co-authored a study on use of AI to tackle disinformation for the European Parliament (2019) and a study mapping global responses to disinformation for UNESCO (2020).

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Index

A

advertisement **53**, **61 f.**, 64, 68, 83, 91, 108, 245, 265, **268 f.**, **334 f.**, 343, 341, 524, 528, **534**, **536**, 540 f., ***567–569***, 572–574, 580.

Africa 439, 441–443, **464–471**

age appropriate design code (AADC) 336 f.

AIQ 332 f.

algorithmic transparency *56*, 391

amplification of content **108**, 452, 510, 524, *574*

anti-fake news law 435, 439, *549* f.

antitrust 49, *67*, **104**, 111, 114, 218

authors' rights 236, **246–253**, 422, **425 f.**, 542 f.

B

BBC 552

big data 173, 175 f., 201, **206 f.**, 282, 315, **331**, 351, **358–363**

big tech 49, 114, 151, 174–176, 187 f., 200 f., 207, *569* f.

Bill C-10 291 f., **303–305**, **307–323**, 373, 379, 384

Bill C-36 369, **395–398**, 400

bots 64, 219, **267**, *269*, 285, 391, 394, 536, ***569***

broadcasting 131 f., 266, **280–287**, **291–323**, 373, 374, 379, *566*, *575*

bundling benefits 306

C

California Consumer Privacy Act 376

Cambridge Analytica **330–333**, 436, 512

Canada 291–323, 367–407

Canadian Commission of Democratic Expression 369, 378, 381, 389 f., **392–397**

Canada's Broadcasting Act **291–323**, 373

censorship 26, 43, 93, 106, 153, 183, 242, 254, 419, 430, 446 f., **451-456**, 467, 470, 551, *554–556*

chilling effect 225, 372, 478 f., 485

China **168–184**, 200 f., 211, 582

chinese platforms 167 f., **172–174**, **180–184**

choice architecture 77, 82, 94

Christchurch 368, 452

citizen **187–192**, **196–199**, 282 f., 359 f., **389–391**, **398–404**

co-regulation 187, 203, 297 f., 220–222, **579–581**

Code of Practice on Disinformation 16 f., 74, **531–541**

communication platforms 30 f., 34 f., 42, **47–51**, ***53–66***, 71, 88–92, 112 f., 121, 182 f., 216–221, 234–236, 279, 384–387, 399, 439, 510–512, 551, ***565–582***

community standards 28, 39, ***56–59***, 146–147, **154–157**, 160, 271–273, **276–278**, **423–428**, 453

competition 48–50, 67 f., 100–106, **111–115**, **123–125**, 141 f., 202 f., 280, 306–308, **338–344**, 478, 581

compliance approach 417-420

consultation paper 297, 381, 389, **398 f.**

consumer **119–141**, 177, 187 f., **245**, **338–344**, 536

consumer protection 106–109, **119–142**, 177, 188, 205–207, 245, **341–344**, *376*

consumer-friendly approach 119-142

content governance 25 f., 34, 42, 393, **433–447**, *569*

content moderation 25 f., 27, ***52–71***, 145, **148–161**, 234, **270 f.**, **276–278**, **434–440**, **453–458**, **510–514**, 544

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

content regulation 26, 29, 37, 77, 415–418, 420, 429 f., 453, 576, 581 f.
contract law 56, **154–161**
copyright **233–236**, **246–256**, 306, 371 f.
copyright retransmission rules 306
coronavirus 62, 199, 263–266, 280, 434 f., 514, 575
COVID-19 *see coronavirus*
criminal justice 377 f., 473, **479–486**
criminal law 86–88, 242 f., 273–275, 371, **377–379**, 400, 428–429, 454–455, **473–486**, **501–504**, 571 f.
criminalizing 479_486
CRTC **292–321**, 372 f., 379, 384 f.
culture 210, 216 f., 250 f., 298–303, 475
custodial services 112
cyber bullying 82, 198, 476
cyber hate 82, 489

D
data access **64 f.**, **67–71**, **112–115**, 173, 340 f., **351–363**
data portability 69, 99, **111–115**
data protection 67, 73 f., 113, 179, 187, 200, **205–207**, **240 f.**, **329–350**, **351–363**, **427–429**
democratic expression **369–407**, 454, 467 f., **473–475**, 480
democratic expression online *see democratic expression*
democracy 47–49, 75, 180, 184, 212, 218, 228–230, 250, 382 f., 389, 464–471, 573 f.
depersonalization 362 f.
design 47 f., 57 f., **63–71**, 77, **82–85**, 89–94, 114, 219, 267 f., **270–273**, 286 f., 329 f., 336 f., 344–350, 577
design code 336–341
design specifications 47 f., 57, 63–71, 270–273
digital advertisers 245, 334
digital markets **47–53**, 103 f., 111–113, **338–341**
digital Markets Act 36 f., 47–53, 67, 70 f., 113, 566
digital markets unit 338-340
digital news portals 228,
digital platform regulation 27, 36 f., 47–53, 57–67, 70f., 111–113, 184, 187–208, 215–229, 234 f., 245 f., 263–287, 329–350, 391, 401, 511, 531–535, 550 f., 565 f., 575–582
digital Services Act 27 f., 43, 47–65, 70 f., 102, 109 f., 270, 279, 286, 430, 486, 537, 566, 570 f., 577, 579–581
disinfodemic 433, **509–511**
disinformation 62 f., 74, 92 f., 102, 114–116, 168–170, 180 f., 267–270, 277–280, 389, 419, 435–447, 509–528, 531–545, 549–560, 573–582
disinformation campaigns 167–170, 180
disinformation monitoring programme 541
distribution 298 f., 306 f., 373, 378 f., 528, 581 f.
domestic boundaries 438
due diligence 47, 63 f., 224 f., **266 f.**, 286, 331, 577–580
duties 25–44, 85 f., 150, 192–195, 346–348
duty of care 43, **77–95**, 112, 346–348, 502–504, **576–581**
duty to act responsibly 379, 392 f., 400
duty to cooperate 428–430
duty to report 348, **383**, **428–430**

E
E-Commerce Directive 27, 49, 75, 110, 511
economic rights 66, 247, 277
ecosystem 136, 174, 219, 235, 282, 528, 533, 541, 572, 582
e-courts 391
EDMO 532, 537–540
election 25, 49, 64, 99–101, 168–170, **180–184**, **198–200**, 268–271, 300–303, 332 f., 369, 416, 442 f., **509–528**, 533, 552–560, 573
employer 85–87, **489–504**
employers' responsibility 489–504
entry barriers 111, 138, 543, 567 f.
e-tribunal 394 f.

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

EU 37 f., 44, 47–71, 78 f., 110–115, 130–142, 151 f., **168–184**, **282–287**, 317–322, 473–486, 491, 502, 510–514, **531–545**, 574–582
EU law **47–76**
Europe *see EU*
European Union *see EU*

F
Facebook 31–44, 49–71, 83 f., 103, 108–113, 119–125, 132 f., 138–142, **145–161**, 172–181, 200 f., 216, 235, 264–283, 331–334, 340–342, 384–406, 416, **423–430**, 438-443, 453, **510–528**, 557 f., 570
Facebook's Oversight Board 38 f., **145–161**, 386, 394, 406
fake news 81, 100, 188, 198 f., 276, 283, 384, 419, 439, 533, 536, **550–560**, 575
Federal Office of Justice 275, 424, **428**
federated protocols 122 f., 128 f.
filter bubbles 84, 283, 380
filter systems 254, 270–273
filtering 224, 254, 263, **273–279**, 373, 376, 383, 568
fine 177, 331–334, 397, 403, 421–428, 439
Finland 474–486, 489–504
foreign investment restrictions 306, 309
free speech 114, 184, 195, 214–225, 242–244, 295, 393, 440–442, 447 f., 451–471, 475, 479,
freedom of expression 25–33, 50, 64, 89, 92 f., 151–161, 192, 197, 199, 222, 238–256, 274–277, 291 f., 323, 348, 389, 417, 426, 435, 442–447, **451–471**, 473–486, 502, 532, 543–545, 549–551, 570–579
freedom of speech 148–161, 223, 228, 554
freedom of the press 154, 370, 532
Freiwillige Selbstkontrolle Multimedia-Dienste (FSM) **274–276**, 422f.

G
GAFAM 167–184
GDPR 50, 58, 74, 93, 113, 126, 168, 173, 179 f., 330–350, 358 f., 391
general basic data 357
general personal data *see Generel basic data*
geopolitics 167–174
Germany 30, 49, 60, 130, 138–142, 168, **263–287**, 319, 371, 383, 415 f., 430, 439–441, 463 f., 473 f., 531, 558, 578
global disinfodemic 509 f.
global market 37, 167–174
global platforms 470–472
global south 433
Good Samaritan Clause 106
Google 36, 51, 76, 103 f., 113, 118, 172, 175–180, 199–207, 238–240, 264, 279, 282, 287, 341, 382, 387 f., 390, 406, **510–528**, 532, 536, 540
guidance 65, 87 f., 92 f., 140, 147 f., 281, 294, 332, 343, 348, 396, 398, 457–460, 493–495, 425 f., 532–545, 574, 582

H
harassment 107, 155, 346, 368, 436, 476–479, **496–504**,
hard approach 187 f.
hate speech 26, 31–34, 47–50, 78, 82, 99 f., 152–155, 168, 187–200, 205, 208, 216, 236, 244 f., 272–277, 367–407, 415–430, 433–447, 451–471, 473–486, 489–504, 531, 558, 571–578
hate speech regulation 188, 193, 489–504
homosexuality 462, 467
horizontal effect 25, 28, 30 –32, 38–44
horizontal effect of human rights 25, 28, 30, **38–44**
hosting privilege 109 f.
human rights 25–30, 33, 37–44, 89, 148–161, 183, 191–197, 206–208, 224, 241–244, 255, 276, 371–377, 381–383, 395–398, 405, 442–447, **451–471**, 473–486, 543, 553, 572 f.,

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

I
ICO 330–342
illegal content 26–44, 52–71, 106, 114, 188, 273, 346 f., 373, 385, 401, 415–427, 452, 455, 510 f., 558, 571 f., 577, 580
immunity 26, 30, 34, 79 f., 92, **106–115**, 217, 264, 346, 380, 577
India **215–229**, 441–443, 448–452, 555, 578
infiltration 182 f.
information disorder 99 f., 535
information intervention 436, 445–447
intellectual property 92–94, 107, 236, 240, **246–249**, 255, 307, **314 f.**, 355, 511,
Intermediary Guidelines and Digital Media Ethics Code 215–229
intermediary liability 219, 222, **236–238**, 454 f.
intermediary rules 223
International Covenant on Civil and Political Rights (ICCPR) 43, 405, 444, **455–471**
internet shutdowns 218, **443 f.**
internet streaming 305–308, 311, 317
interoperability 37, 69, 99, **111–115**, **119–142**, 340, 578
interoperability and its legal possibilities under EU law 119
Interstate Media Treaty 263, **265–279**, **285 f.**
intervention 29, 138 f., 225, 294, 301, 340 f., 392, **428**, **433**, **436**, **444–447**, 540
IP policy 314
ISP 79, **372**

J
Japan 170, **187–213**, 578
Journalism Trust Initiative 542
journalistic standards 266 f.

K
Key Performance Indicators (KPIs) 533, 538 f., 545

L
Latin America **233–256**
legal approach 351–363
legislative initiative **47–76**
liability 26 f., 36, **77–81**, 90, 99 f., **104–115**, 147, 195, 217, 219, 222, 224 f., **236–242**, 333, 348 f., 354, 361, 372, 379 f., 384–386, 395, 453–455, 483 f., 504, 511 f., 531, 544, 557, 577, 579, 588
liability privilege 109 f.
limitation 104, 108, 195, 236, 248
limitation and content filtering 254–256, 376
local news 99 f., **102–105**, 114 f., 522
local news subsidies 99
lock-in effects 111, 120, 124, 203

M
market regulation 71; 576
media law 18; 20; 31; 48-50; 63; 74; **263-364**; **263-289**
media pluralism **531-547**
Media Pluralism Monitor **531-547**
Merkel 274-276
Merkel regime case 274-276
messenger services 51; 62; **119-143**; 418-419
microtargeting 269
misinformation 62; 83; 92; 101; 198-199; 216-219; 229; 264; 286; 320; 347; 400; 440; 442; 516; 522; 524; 535; 541; 544; 549; 550-551; 556-560
models **25-163**; **565-584**
monetization of content **108**
multiplicity 264; 349
must-carry regulations 306

N
national security **167-185**; 348; 436; 442; 444
NDP **382-383**
Netflix 180; 223; 279; 282; 287; 291; **295-322**
network economy 304
NetzDG, see Netzwerkdurchsetzungsgesetz

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

Netzwerkdurchsetzungsgesetz 30; 58; 168; 270; 273-279; 285-286; 371; 383; 392; 401; **415-432**; 439; 452-453; 455; 531; 558
news aggregators 103-104; 301
news desert 103
Nigeria 440; 550
notice and takedown 13; 27; 58; 110; 234; 367; 372; 379; 383; 386-388; 392; 395

O
occupational safety and health **489-506**
Ofcom 80; 92; 338; **345-349**; 575; 579; 581
Online Communication Services (OCSs) **399-405**
online dispute resolution 367; 386; 388-389; 394; 404
online harms 16; 77; 329; 344-345; 349; 388; 392; 453; 572; 581
online harms white paper (OHWP) **344-345**
online hate 20; 188; 193; 199; **367-413**; 433-440; 446-447; 459; 470; 473; 489-490
online safety 329; 349; 377; 581
online shaming **473-487**;
online speech 145; 369; 385; 406; 425; **433-438**; 442; 444; 447; 510; 515
OSHA **489-506**
OTT 121; 130; 142; 180-181; 222; 228; 295-296
overblocking 415; **421-422**; 424-425
Oversight Board 18; 38-39; **145-163**; 386; 394; 402; 404-406; 416
Over-the-top providers 121; 130; 142; 180-181; 222; 228; 295-296

P
payment data 356
payment system 316
paywalls 103-104
People's Republic of China 18; **167-185**; 200-201; 211; 582
personal data 19; 67; 73; 124-126; 132; 141; 152; 179; 187-188; 200; 205-207; 236; 240-241; 268; 331-332; 335; 337; 340; **351-364**; 390-391; 404; 568-570
personal data in the public domain **360-363**
personal data protection 179; 187; 200; 205-206; 240; 358
platform accountability **99-117**
platform governance **167-185**; 367; 393; **433-450**
platform information disorders **99-117**
platform power 512
platform regulation 18; 20; **25-163**; **187-213**; **215-231**; 367; 415; 511; 531; 551
platform responses **509-530**
platform unaccountability 43-44; 513

"platformization" **451-471**;

platforms 14-20; **25-163**; **329-350**; **367-506**; **531-547**; **565-584**
policy developments **99-117**; 329
portable computing devices 218
prejudicial 462-466; 469-470; 554
PriceWaterhouseCoopers report **310**
principles of criminal law 473
privacy policy 354
privacy protection 187
propaganda 212; 216; 219; 378; 382; 396; 466
Protection Against Online Falsehoods and Misinformation Act POFMA **551-560**
public European space 263
public infrastructure 172
public service broadcaster 263-265; 280

Q
QAnon 264; 517; 519; 526

R
recommendation systems 53; **57-58**; 64-65; 70; 263; 270-273; 279; 532; 534
recommender system 47; 55; 91; 286; 513; 542; 580

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

regulation **25-163**; **187-364**; **489-506**; **549-563**
regulation of content **233-259**
regulatory responses 20; 404; **565-584**
Republic of China 18; **167-185**; 200-201; 211; 582
Republic of Finland **489-506**
Republic of India 18; **215-231**; 441; 443; 452; 555; 578
Republic of Singapore 20; 170; 438-439; **549-563**; 578
Republic of South Africa 20; 433; 451; 456; 464-471
right to private life 473-475
rights **25-45**; **145-163**; **233-259**
risk assessment 40; 43; 47; 63-64; 77; 85; **89-94**; 347-348; 394; 577
risk mitigation 47; 64-66
risk of hate speech at work **489-506**
Russia *see Russian Federation*
Russian Federation 17; 19; 211; **351-364**; 452; 462-463; 576; 582
Rwanda 436; 445

S
safe harbour 104; 221; 224-225
safety 31; 61-64; 77; 81; 85-86; 151; 155; 329; 345; 347-349; 368; 395; 399-402; 416; 460-461; **489-506**; 554; 573; 577; 579; 581
safety measure **489-506**
SCL **332-333**
Section 230 Reform 13; 34; **99-117**; 106; 236; 378-379; 511
security safeguards 377
self-regulation 16-17; 27; 39; 92; 145; 173; 187-188; 195-197; 215; 221-222; 266-267; 269; 274; 279; 345; 392; 422; 430; 575; 578-579; 581
sensitive personal data 351; 358-359
shaming 442; **473-487**
shareholder 171; 406
shutdown 218; **433-450**
significant social media intermediary 215
simultaneous substitution policies **305**
Singapore 20; 170; 438-439; **549-563**; 578
social bots **267-268**; 269; 285; 569
social media **77-97**; **215-259**; **351-364**
social media councils 367; 388-389
social media intermediary 215
social media regulation 215; 233
social networks 47; 49; 51; 71; 121; 234-236; 246; 250; **263-289**; 351-352; 355; 358; 360-362; **415-432**
soft approach **187-213**
South Africa 20; 433; 451; 456; 464-471
Stadium Ban Decision 426
statutory duty **77-97**; 393
statutory duty of care 18; 43 **77-97**; 344-345; 388; 392
strategic market status 339
streaming 18; 250; 278; 281-282; 291; 301; **303-319**
supervisory 427-428; 501; 579; 581
Systems Approach 18; **77-97**

T
Taiwan 17-18; **167-185**; 200-201
Taiwan-China relationship 167; 169
targeted advertising 61; 83; 329
tax incentive 104; 114
TikTok 35; 51; 63; 174; 267; 320; 399; 509-510; 514; 524-526; 532; 536; 538; 540
trade agreement 307; 371; 378; 385
transparency 26; 28; 31; 49; 54-56; 58; 61; 65; 89-90; 109-110; 114; 127; 173; 180; 187; 202; 204; 207; 245; 263; 269; 271-272; 274; 279; 285-286; 302; 335; 337; 346-348; 383; 391; 394; 401; 421; 423-425; 427; 430; 442; 509; 511; 513; 528; 533-534; 537-540; 542; 577; 580;
trolls 574
Twitter 16; 32; 35; 38; 51; 113; 196; 216; 250; 267; 273; 382; 399; 416; 440; 441; 452; 509-514; 523-527; 532; 536; 555

U
UK 18; 29; 43; **77-97**; **329-350**; 392; 512; 526; 577-579; 581
UN Security Council 445-446

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA

United Kingdom 18; 29; 43; **77-97**; **329-350**; 392; 512; 526; 577-579; 581
United States of America 16; **99-117**; 158; 201; 519; 522; 576; 578
Universal Declaration of Human Rights (UNDHR) 444
US 2020 Elections 509
USA *see United States of America*
US Digital Millennium Copyright Act 372
user behaviour 81; 83; 85; 94; 130
user data 49; 112-113; 173; 351-352; 355-356; 391; 512
user generated content 88; 238; 251; 312; 319-322; 379; 442; 452
user-to-user service 346

V
very large communication platforms **47-76**
video on demand 282; 318; 322
VOD 282; 318; 322

W
White Paper on Online Harms 388; 392
work **367-413**; **489-506**
World Health Organization (WHO) 62; 510; 526

Y
Yle 575

Z
ZDF 275; 280; 282

https://doi.org/10.5771/9783748929789 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY-SA