M&K Medien & Kommunikationswissenschaft

herausgegeben vom
Leibniz-Institut für Medienforschung | Hans-Bredow-Institut

Themenheft „Die Datafizierung der Kommunikation – neue methodische Zugänge und Herausforderungen“ / „The Datafication of Communication – New Methodological Approaches and Challenges“

Gastherausgeber*innen

Julia Niemann-Lenz / Tim Schatto-Eckrodt / Emese Domahidi / Merja Mahrt

Mit Beiträgen von

Denise Sommer / Jörg Hagenah / Anna-Sophie Brucks

Patrick Zerrer / Paul Pressmann / Cornelius Puschmann / Philipp Krieter

Philipp Kessling / Felix Victor Münch

Nadezhda Ozornina / Mario Haim

Annekatrin Bock

1 2026
74. Jahrgang
ISSN 1615-634X

Nomos Inlibra

Medien & Kommunikationswissenschaft

2026
74. Jahrgang
Heft 1

Herausgegeben vom Leibniz-Institut für Medienforschung | Hans-Bredow-Institut

Inhalt

Editorial

Introduction to the Special Issue "The Datafication of Communication – New Methodological Approaches and Challenges"

Julia Niemann-Lenz / Tim Schatto-Eckrodt / Emese Domahidi / Merja Mahrt*

Over the past decade, methodological developments in communication research have accelerated considerably. New approaches and tools have been introduced under different labels, including "big data," computational communication science, and, more recently, generative artificial intelligence. Each of these terms signals shifts in how communication processes are empirically accessed and analyzed, often accompanied by promises of increased analytical power.

These methodological developments are rooted in a fundamental transformation of communicative environments and research conditions. The critical foundation of these transformations lies in processes of *datafication*. Datafication describes a broader sociotechnical process through which social action, communication, and interaction are translated into quantifiable data that can be stored, processed, and repurposed (Mayer-Schönberger & Cukier, 2013). Importantly, datafication is not a neutral process, but is embedded in asymmetrical power relations: Decisions about which forms and aspects of communication are captured, how they are classified, and who gains access to the resulting data are shaped by technical infrastructures as well as organizational and economic interests. These infrastructures and interests exert a substantial influence on communication processes themselves. For instance, a growing body of research shows that the affordances of algorithmic infrastructures in social media affect different forms of communication, including interpersonal, journalistic, and political communication (Hase et al., 2023; Li et al., 2024; Teocharis et al., 2023; Welbers & Opgenhaffen, 2019). Moreover, central concepts of media and communication research—such as selection, gatekeeping, visibility, and the formation of public discourse—are increasingly intertwined with datafied and algorithmically mediated processes (Couldry & Hepp, 2017; van Dijck et al., 2018).

The datafication of communication not only alters the object of media and communication research itself but also affects which aspects of communication are considered empirically observable and, consequently, researchable. As communicative practices are translated into digital traces, forms of communication that are easily captured, quantified, and processed tend to gain analytical prominence, while others remain less visible or difficult to access. New forms of data access, automated data collection, and computational methods of analysis thus increasingly influence how communication phenomena are opera-

* Dr. Julia Niemann-Lenz, German Centre for Higher Education Research and Science Studies (DZHW), Lange Laube 12, 30159 Hannover, Germany, lenz@dzhw.eu, https://orcid.org/0000-0003-1991-8535;
Dr. Tim Schatto-Eckrodt, University of Hamburg, Journalism and Mass Communication, Von-Melle-Park 5, 20146 Hamburg, Germany, tim.schatto-eckrodt@uni-hamburg.de, https://orcid.org/0000-0003-1658-4373;
Prof. Dr. Emese Domahidi, Technische Universität Ilmenau, Institut für Medien und Kommunikationswissenschaft, Fachgebiet Computational Communication Science, Ehrenbergstraße 29, 98693 Ilmenau, emese.domahidi@tu-ilmenau.de, https://orcid.org/0000-0003-3530-797X;
PD Dr. Merja Mahrt, Weizenbaum Institute e. V., Hardenbergstraße 32, 10623 Berlin, Germany, merja.mahrt@weizenbaum-institut.de, https://orcid.org/0009-0009-4246-734X.

DOI: 10.5771/1615-634X-2026-1-3

tionalized, measured, and interpreted. Under the label of computational communication science, digital trace data have moved to the center of empirical research in the field (Domahidi et al., 2019). Approaches such as automated content analysis, network analysis, machine learning, and, more recently, large language models (LLMs) promise new insights into communicative dynamics in a datafied society. These approaches build on large-scale datafied communication as their empirical foundation and further automate processes of classification, modeling, and interpretation. In doing so, they extend established methodological repertoires while simultaneously challenging existing conventions of data collection, analysis, and validation.

The increasing prominence of data-driven research has also prompted critical debate. Scholars have questioned the assumption that social reality can be adequately captured through large-scale data alone, pointing to the socially constructed nature of data and the socio-technical conditions under which they are produced (boyd & Crawford, 2012). From this perspective, digital data are not neutral representations of human behavior but are shaped by platform architectures, algorithms, institutional practices, and power relations. Similarly, the methodological convenience of accessing social media data has led to a proliferation of studies focusing on easy-to-research spaces. This may have contributed to an overestimation of these platforms' role in misinformation research and to the neglect of other areas, such as the role of established media in disseminating false information (Altay et al., 2023). Critical accounts have further emphasized the role of large platforms in extracting and controlling data, conceptualizing these practices as forms of "data colonialism" that parallel historical patterns of resource extraction (Couldry & Mejias, 2019). Relatedly, the ideological foundations of datafication have been described as "dataism," referring to the belief that human actions can be objectively quantified and understood through data (van Dijck, 2014). In addition to these epistemological critiques, the use of new data sources and analytical techniques raises legal and ethical concerns, particularly with regard to privacy, consent, transparency, and accountability (Spirling, 2023).

In light of these critiques and acknowledging the dual role of datafication as both a methodological precondition and a constitutive context of communication, the Digital Communication and Methods of Media and Communication Research sections of the German Communication Association (DGPuK) jointly organized a conference in Hamburg in 2024 to address these developments. The conference brought together 70 scholars working on different aspects of datafied communication research in order to reflect on methodological innovations, empirical practices, and critical perspectives across a range of subfields. The presentations and discussions quickly made clear that the topic engages colleagues from both sections, as they are centrally affected by datafication. Large language models were a focal point of many panels, as scholars employed different approaches to explore new methodological opportunities (as well as potential pitfalls) and to carve out appropriate roles for these approaches. This applies both to digital communication research and to the development of methods for data collection, organization, and analysis. It also became apparent that the field is currently highly productive and fast-moving, underscoring the need for more long-term investigations to complement the often tightly focused and practically oriented studies on developing or applying new datafied approaches.

Following the fruitful discussions at the conference, this Special Issue addresses some of the broader methodological implications of datafication for communication research. At the same time, it brings together contributions that focus more specifically on different stages of the research process, examining how digital data are generated, accessed, analyzed, and critically reflected upon within the discipline. Rather than advocating for a singular methodological position, the Special Issue aims to highlight the diversity of approaches that

currently shape datafied communication research. The contributions engage with different stages of the research process and illustrate how methodological innovation and critical reflection can be productively combined. While this results in a broad thematic scope, the Special Issue deliberately embraces this breadth in order to capture the heterogeneity of datafied communication research and the multiple ways in which datafication affects empirical practice.

The Special Issue opens with a contribution in German language by Denise Sommer, Jörg Hagenah, and Anna-Sophie Brucks who trace the long-term effects of datafication on empirical communication research. They conducted a content analysis of the three major German-language journals from 2003 to 2023 to investigate changes in method and data choices over time. Computational approaches have only recently seen an—albeit steep—increase in popularity, while the growing availability of larger datasets for secondary analysis does not appear to have resulted in more studies drawing on these resources. The authors conclude that compared to neighboring disciplines such as sociology, communication research, at least in these three journals, does not yet appear to be strongly data-driven, and that data infrastructures should be strengthened to enable scholars in German-speaking countries to keep pace with international developments.

Next are Patrick Zerrer, Paul Pressmann, Cornelius Puschmann, and Philipp Krieter, who present a holistic framework for researching mobile media use. In their article, Zerrer et al. highlight the relevance of mobile device use for media use research more broadly and reflect on the current state of the field. Accessing mobile media use involves various hurdles for researchers, ranging from non-trivial questions of technical access to participants' devices, the ethical considerations concerning user privacy to the methodological challenge of operationalizing media exposure beyond the analysis of timestamped browser histories alone.

Through two case studies, Zerrer et al. demonstrate the possibilities and limits of mobile media use tracking by combining screen recording, a keyword tracker that automatically detects relevant terms appearing on screen, and the matching of observed exposure to media content through scraping and (computational) analysis of, for example, relevant social media accounts. Based on these case studies, the authors derive five conceptual core principles for in-app tracking, concerning 1) user privacy, 2) the management and interpretation of the raw data generated by these methods, 3) potential biases arising from the interconnectedness of the socio-technical systems under study, 4) the need to track not only singular media exposures but also cascades of subsequent exposure in order to understand how information travels across platforms, and 5) the assurance of replicability through precise and comprehensible documentation of the reported approaches.

In the third article, Philipp Kessling and Felix Victor Münch address a central methodological challenge in the study of the digital networked public sphere: the systematic and transparent collection of data across multiple platforms. Rather than focusing on specific technologies or access points, the contribution proposes a conceptual framework that abstracts from concrete data sources and instead structures the data collection process itself. By conceptualizing data collection as a sequence of modular components, the article provides a common language for describing, comparing, and documenting exploratory data collection in cross-platform research. This modular perspective is particularly valuable in research contexts characterized by heterogeneous data access conditions and rapidly changing technical environments. It allows researchers to disentangle conceptual decisions from technical implementations and thereby enhances transparency, reproducibility, and interoperability. Beyond its immediate applicability, the framework contributes to ongoing

methodological debates by offering a way to systematically reflect on how networked communication is sampled and empirically reconstructed under conditions of datafication.

Nadezhda Ozornina and Mario Haim examine the influence of target language selection on topic modeling in multilingual settings—an issue that has so far received little attention in computational communication research. The study compares two methodological approaches: consolidating texts into one of the original document languages versus translating all texts into an intermediary language that is not part of the original corpus. These approaches are tested on a parallel corpus of United Nations documents in Russian and German and are subsequently replicated using a second bilingual dataset. The findings indicate that translation into an intermediary language—English in this case—offers certain advantages, including a more symmetrical topic distribution and a higher degree of overlap among top-ranked terms. At the same time, this strategy substantially reduces the available vocabulary compared to consolidation in an original language, which may have implications for downstream analyses, such as cross-lingual emotion detection. Particularly noteworthy are the detailed discussion of the findings and the formulation of best-practice recommendations, which provide valuable guidance for future research in this area.

Last but not least, Annekatrin Bock explores the benefits of combining the theoretical framework of critical data studies (CDS) with the research agenda of communication studies, while offering insights into the general assumptions of CDS, their relevance for communication research, and the methodological application of these principles. In her article, Bock critically reflects on the use of digital and computational methods by highlighting the implications of methodological frameworks that rely on platform data, which can be understood as made (i.e., not raw or value-free) and invisibly entangled within the socio-technical systems that both produce these data and exercise interpretative sovereignty over them.

By outlining two approaches from the methodological canon of CDS—the *walkthrough approach* and *data journeys*—Bock demonstrates how empirical communication research can benefit from adopting a perspective of critical data literacy. This perspective allows researchers to foreground the dynamic and relational aspects of data, as well as the ways in which data practices are embedded in broader social and technological systems. In an outlook, Bock addresses how researchers might deal with the implications of different data sources and calls for the adoption of open data practices and robust documentation to further transparency and accountability, as well as embracing interdisciplinary methodological (re-)adjustments where necessary.

The collection of articles thus illustrates how deeply and in manifold ways datafication affects communication research. The data, methods, and tools that are and become available steer research trends and may, in the long run, even shift entire paradigms. During the conference on datafication in Hamburg in the fall of 2024, this was often discussed through comparisons between the approaches presented and potential opportunities, similarities, and differences vis-à-vis related generative AI-based approaches. At the time, this line of research was still in its early stages, but the field has since been highly productive, as evidenced by numerous conferences and publications. We assume that many of the issues discussed in the five contributions collected here will require continued investigation, while others, such as the intransparency of especially commercial generative AI applications in particular, are likely to become even more pressing.

References

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on Misinformation: Conceptual and Methodological Challenges. *Social Media + Society, 9*(1), 205630512211504. https://doi.org/10.1177/20563051221150412

boyd, d., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society, 15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Couldry, N., & Hepp, A. (2017). *The Mediated Construction of Reality*. Polity.

Couldry, N., & Mejias, U. A. (2019). T*he Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press.

Dijck, J. van. (2014). Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. *Surveillance & Society, 12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Dijck, J. van, Poell, T., & Waal, M. de. (2018). *The Platform Society: Public Values in a Connective World*. Oxford University Press. https://doi.org/10.1093/oso/9780190889760.001.0001

Domahidi, E., Yang, J., Niemann-Lenz, J., & Reinecke, L. (2019). Computational Communication Science | Outlining the Way Ahead in Computational Communication Science: An Introduction to the IJoC Special Section on "Computational Methods for Communication Science: Toward a Strategic Roadmap". *International Journal of Communication, 13*, 3876–3884. https://ijoc.org/index.php/ijoc/article/view/10533

Hase, V., Boczek, K., & Scharkow, M. (2023). Adapting to Affordances and Audiences? A Cross-Platform, Multi-Modal Analysis of the Platformization of News on Facebook, Instagram, TikTok, and Twitter. *Digital Journalism, 11*(8), 1499–1520. https://doi.org/10.1080/21670811.2022.2128389

Li, M., Suk, J., Zhang, Y., Pevehouse, J. C., Sun, Y., Kwon, H., Lian, R., Wang, R., Dong, X., & Shah, D. V. (2024). Platform Affordances, Discursive Opportunities, and Social Media Activism: A Cross-Platform Analysis of #MeToo on Twitter, Facebook, and Reddit, 2017–2020. *New Media & Society, 28*(1), 119–147. https://doi.org/10.1177/14614448241285562

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (S. 242). Houghton Mifflin Harcourt.

Spirling, A. (2023). Why Open-Source Generative AI Models Are an Ethical Way Forward for Science. *Nature, 616*(7957), 413. https://doi.org/10.1038/d41586-023-01295-4

Theocharis, Y., Boulianne, S., Koc-Michalska, K., & Bimber, B. (2023). Platform Affordances and Political Participation: How Social Media Reshape Political Engagement. *West European Politics, 46*(4), 788–811. https://doi.org/10.1080/01402382.2022.2087410

Welbers, K., & Opgenhaffen, M. (2019). Presenting News on Social Media: Media Logic in the Communication Style of Newspapers on Facebook. *Digital Journalism, 7*(1), 45–62. https://doi.org/10.1080/21670811.2018.1493939

Empirische Aufbruchsstimmung?

Verdatung und Datafizierung als Impulsgeber kommunikationswissenschaftlicher Forschung?

Denise Sommer / Jörg Hagenah / Anna-Sophie Brucks*

Der Beitrag untersucht den Einfluss von Verdatung und Datafizierung auf die kommunikationswissenschaftliche Forschungspraxis im deutschsprachigen Raum. Vor dem Hintergrund medialer Wandlungsprozesse wie der Mediatisierung und der Digitalisierung wird gefragt, ob der Anstieg an (prozessproduzierten) Daten in Deutschland nicht nur zu mehr Forschungsmöglichkeiten, sondern auch zu einer empirischen Aufbruchsstimmung im Fach geführt hat? Eine Inhaltsanalyse von 935 Originalbeiträgen in den Zeitschriften Publizistik, M&K und SCM (2003–2023) zeigt einen konstant hohen Anteil empirischer Arbeiten (ca. 70 %), wobei quantitative Methoden dominieren. Gleichzeitig ist ein Rückgang sekundäranalytischer Studien sowie längsschnittlicher Designs zu beobachten – im Gegensatz zur Soziologie, wo diese zunehmen. Computational Methods gewinnen seit 2013 sichtbar an Relevanz, während experimentelle Designs abnehmen. Der Beitrag verweist auf strukturelle Defizite bei der Bereitstellung und Nachnutzung kommunikationswissenschaftlich relevanter Forschungsdaten, trotz punktueller Fortschritte. Die Ergebnisse unterstreichen die Notwendigkeit, institutionalisierte Dateninfrastrukturen zu schaffen, um datennutzende Forschung nachhaltig zu fördern und Anschluss an internationale Entwicklungen zu sichern.

Schlüsselwörter: Datafizierung, Verdatung, Mediatisierung, Inhaltsanalyse, Fachzeitschriften, Sekundärforschung, Dateninfrastruktur, empirische Kommunikationsforschung

An Empirical Awakening?

Datafication and Datafication Processes as Impulses for Communication Research?

This article examines the impact of datafication and the growing availability of data on research practices in communication studies in German-speaking countries. Against the backdrop of change processes such as mediatization and digitalization, it asks whether the increase in (process-produced) data has led not only to expanded research opportunities but also to renewed empirical momentum within the discipline. A content analysis of 935 original articles published in the journals Publizistik, M&K, and SCM (2003-2023) reveals a consistently high proportion of empirical studies (approx. 70 %), with quantitative methods clearly predominating. At the same time, a decline in secondary analyses and longitudinal research designs occurs—contrasting with developments in sociology, where such approaches have increased. Since 2013, computational methods have gained noticeable relevance, while experimental designs have declined. The article points to structural deficits in the provision and reuse of research data relevant to communication studies, despite isolated advances. The findings underscore the need to establish institutional-

* Prof. Dr. Denise Sommer, Ostfalia Hochschule für angewandte Wissenschaften, Hochschule Braunschweig/Wolfenbüttel, Campus Salzgitter, Karl-Scharfenberg-Str. 55/57, 38229 Salzgitter, Deutschland, de.sommer@ostfalia.de, https://orcid.org/0009-0003-5169-7880;
Dr. Jörg Hagenah, Ostfalia Hochschule für angewandte Wissenschaften, Hochschule Braunschweig/Wolfenbüttel, Campus Salzgitter, Karl-Scharfenberg-Str. 55/57, 38229 Salzgitter, Deutschland, hagenah@ostfalia.de, https://orcid.org/0000-0003-4563-6611;
Anna-Sophie Brucks, M. A., Ostfalia Hochschule für angewandte Wissenschaften, Hochschule Braunschweig/Wolfenbüttel, Campus Salzgitter, Karl-Scharfenberg-Str. 55/57, 38229 Salzgitter, Deutschland, an.brucks@ostfalia.de, https://orcid.org/0009-0006-2664-3612.

 DOI: 10.5771/1615-634X-2026-1-8

ized data infrastructures in order to sustainably promote data-driven research and ensure alignment with international developments.

Key words: datafication, digitization, mediatization, content analysis, academic journals, secondary research, data infrastructure, empirical communication research

1. Einleitung

Der Beitrag zeichnet die Forschungstrends und Praktiken im Umgang mit Daten in der deutschsprachigen Kommunikationswissenschaft der vergangenen 20 Jahre nach. Er fragt, ob die wahrgenommene Vervielfachung an produzierten Forschungsdaten – verursacht durch einen Anstieg an öffentlicher und medialer Kommunikation und ihrer entsprechenden Vermessung – den Forschungsbedarf im Fach angetrieben und zu einer empirischen Aufbruchsstimmung geführt hat. Untersucht wird dafür der veröffentlichte Fachdiskurs der Scientific Community, der sich an den Befunden einer inhaltsanalytischen Untersuchung von deutschsprachigen Fachzeitschriftenbeiträgen (n = 935) der Jahre 2003 bis 2023 ablesen lassen sollte. Die Analyse des Anteils und der Vielfalt empirischer Forschung sowie ihrer Merkmale gibt Aufschluss darüber, ob die breiten und verschiedenartigen Datenräume im Bereich der öffentlichen Kommunikation die wissenschaftliche Forschung vorantreiben und verändern.

Die Vermessung und Quantifizierung von Medien- und Publikumsstrukturen zeichnet die mediale Kommunikation von Beginn an aus (Schneider & Otto, 2007). Sie wird als *Verdatung* verstanden und u. a. aufgrund ihrer spezifischen Marktlogiken und Datenschutzfragen auch kritisch beleuchtet (Pohle, 2019; Schneider & Otto, 2007). Mit der Digitalisierung der öffentlichen Kommunikation haben sich zudem Prozesse der *Datafizierung* auf der Basis von automatisierten Erfassungsmöglichkeiten entwickelt, die neue Datenformen und Forschungsfragen generieren, aber ebenfalls sorgfältiger Einordnung und Reflektion bedürfen – bspw. aus erkenntnistheoretischer und ethischer Sicht (Breiter & Hepp, 2018; Filipovic, 2015). Beide Entwicklungen erzeugen prozessproduzierte Daten, die nicht theoriegeleitet erhoben werden. Damit ermöglichen sie die Erforschung des indirekt oder direkt gemessenen Medienverhaltens und werfen zugleich Anschlussfragen aus theoretischer und methodologischer Sicht auf. Das legt die Vermutung nahe, dass Verdatung und Datafizierung Forschungsanlässe und -impulse für die Kommunikationswissenschaft generiert haben, sei es in der sekundäranalytischen Verwendung marktorientierter Repräsentativdaten (Verdatung) oder in der Exploration und Reflektion der Bedeutung von digitalen Spuren im Netz (Datafizierung).

Im Folgenden werden die Prozesse der Verdatung und Datafizierung genauer dargestellt und im Vergleich mit benachbarten sozialwissenschaftlichen Disziplinen (Soziologie, Politikwissenschaft) in ihrem potenziellen Einfluss auf die kommunikationswissenschaftliche Forschung reflektiert. Die Untersuchung der deutschsprachigen Fachzeitschriftenpublikationen zeigt, wie sich die empirischen Aktivitäten in der wissenschaftlichen Community entwickelt haben, und lässt erste Rückschlüsse auf den Einfluss der Datentrends auf die angewandten Forschungspraktiken zu.

2. Verdatung und Datafizierung in der deutschen Kommunikationsforschung

2.1 Von der Verdatung zur Datafizierung: Begriffsbestimmung und -abgrenzung

Die Vermessung des Medienangebotes und seines Publikums gehört, nicht zuletzt seit der Einführung der sog. *Leser-Analyse* 1954 (später: *Media-Analyse* (MA)), zu den aufwendigsten Projekten der angewandten Markt- und Mediaforschung. Das Ziel dieser Untersuchun-

gen liege nach Schneider und Otto (2007) in einer *Verdatung* der (medialen) Gesellschaft, um das Nutzungsverhalten empirisch zu messen und in mehr oder weniger harte Medienwährungen (vgl. Hagenah et al., 2006, S. 129) zu überführen. Es handelt sich bei der Verdatung um einen primär medienwissenschaftlich geprägten Begriff, der den historischen Prozess der zunehmenden empirischen Datenproduktion und Quantifizierung menschlichen Verhaltens (marktorientierte Angebot-und-Nachfrage-Logiken, Vermessung von Konsument:innen) kritisch in den Blick nimmt (Schneider & Otto, 2007, S. 11; von Oertzen, 2017). Eine neue, engere Beschreibung der Verdatung fokussiert auf die wachsende Datenlandschaft durch die breiter und tiefer werdende Vermessung des medialen Verhaltens als Folge der durch die Mediatisierung wachsenden Medienlandschaft. Dabei wird Verdatung als aktiv gesteuerte Forschung verstanden, die dem medialen Wandel folgend eine zunehmende Anzahl an kontinuierlichen Trend- und Panelstudien oder Statuserhebungen produziert.

Insbesondere der Beginn der aktiven Verdatungsprozesse zeigt die mittlerweile eher aus historischer Perspektive notwendige begriffliche Trennung zwischen Verdatung und Datafizierung: Für die ersten Leser-Analysen ab 1954 wurde mit aufwendigen Befragungen das Lesen von Publikumszeitschriften wie dem Spiegel oder des Kickers mit persönlich-mündlichen Befragungen durchgeführt (Hagenah & Akinci, 2003), in Form von Lochkarten festgehalten und mittels Maschinen ausgezählt. Auftraggeber war die Arbeitsgemeinschaft Leser-Analyse, deren Mitglieder die großen Verlage, die werbetreibende Industrie und die entstehenden Werbe- und Media-Agenturen darstellten (agma, 2025a). Mit zunehmender medialer Bedeutung wurden Abfragen zur Nutzung von Tageszeitungen, Hörfunk- und Fernsehsendern in die Befragungsprogramme integriert, die Leser-Analyse wurde schließlich zur Media-Analyse und als Mitglieder wurden nach und nach weitere Medienhäuser und die öffentlich-rechtlichen und privaten Rundfunksender sowie die Anbieter von Online-Medienangeboten integriert (agma, 2025b).

Die derzeitige *Datafizierung* kann als *digitale Form* von Verdatung verstanden werden und zeigt sich u. a. durch Spuren des medialen Verhaltens im digitalen Raum (Krüger, 2021). Während die ersten Jahrzehnte der Markt- und Mediaforschung eher nicht der Datafizierung zuzurechnen sind, verschwimmen mit dem Aufkommen digitaler Medienangebote wie Spiegel online im Jahr 1994 (Bönisch, 2006, S. 6) auch die Begriffe Verdatung und Datafizierung. Neben (halb-)automatisierten Inhaltsanalysen digitalisierter oder digitaler Medienangebote (Quandt, 2008; Scharkow, 2011) ist nun auch eine Vermessung des Verhaltens im medialen Raum etwa durch Klickverhalten oder Verweildauern möglich (Breiter & Hepp, 2018). Mittlerweile haben nahezu alle traditionellen Medien neben ihren analogen Angeboten auch digitale in Form von Homepages, E-Paper, Apps, Mediatheken oder Streamingportalen etc. Zusätzlich sind neue internetspezifische Angebote entstanden, die den traditionellen Angeboten ähneln (Strippel & Paasch-Colberg, 2020, S. 158), sie ersetzen (z. B. digitaler Freitag) oder thematische Nischen besetzen und community-getragen funktionieren (z. B. transfermarkt.de).

Mit Hilfe der Computational Communication Science (Haim & Jungbluth, 2023) werden große Corpora an Forschungsdaten auf Basis programmierter Pipelines gesammelt. Von großer Bedeutung sind die sozialen Netzwerke, die häufiger durch vorhandene Application Programming Interfaces (APIs) Zugriff auf ihre digitalen Daten zulassen, wenn die Plattformbetreiber die Dokumentationen für (Forschungs-)Software freigeben. Allerdings kann dieser Datenzugang jederzeit vom Plattformbetreiber geschlossen werden. Beispielsweise wurde die GoogleReader-API 2013 für Forschende (überraschend) geschlossen, so dass sog. Feedabos nicht mehr funktionierten, mit denen sämtliche Beiträge des Feeds automatisiert durch den Server von Google regelmäßig abgerufen und im GoogleReader-Account des Nutzers hinterlegt wurden (Trilling 2014, S. 8). Ähnlich wurde die – bis dahin

häufig für kostenlose Untersuchungen von digitalem Verhalten genutzte – Twitter-API geschlossen, als Elon Musk 2023 eine Umbenennung in „X" vornahm und die strategische Ausrichtung der Plattform änderte (Bastos, 2025; Bruns, 2019). Nach Haim (2023, S. 4–5) erweitert zudem das sog. Internet der Dinge die damit zusammenhängende Datenlandschaft.

Während die aus der Verdatung stammenden Daten den *klassischen sozialwissenschaftlichen Erhebungsmethoden* wie Befragung, Inhaltsanalyse oder Beobachtung zuzuordnen sind, gehören zur Datafizierung *neue, technisch erzeugte Datenformen*, die sich häufig nicht mit den klassischen Analyseverfahren untersuchen lassen. Sie bringen neue Grundgesamtheiten und Auswertungsstrategien mit sich, die Schnittmengen mit der Informatik aufweisen (Haim, 2023, S. 2). Zudem bestimmen die verwendeten Software- und Hardwarestrukturen die Auswertungsmöglichkeiten. Bei der Markt- und Mediaforschung gestalten die Auftraggebenden die Strukturen verdateter Daten, bei den aus der Datafizierung stammenden Daten sind es v. a. die Medienanbieter:innen und die Plattformbetreibenden.

2.2 Vervielfachung von Daten und Forschungsimpulse

Aus kommunikationswissenschaftlicher Sicht führen die durch das Datenwachstum erhöhten Forschungsmöglichkeiten zu einem steigenden Analysebedarf. Er bezieht sich einerseits auf eine kommunikationswissenschaftliche Einordnung der sich verändernden Datenlandschaften (Breiter & Hepp, 2018; Filipovic, 2015; Schneider & Otto, 2007) und andererseits auf die Überführung der Daten in relevante Forschungsfelder des Faches (Breiter & Hepp, 2018; Domahidi et al., 2019; Strippel et al., 2018). Wenn der Forschungsprozess aufgrund neu entstandener Daten zunehmend auch datengetrieben erfolgt, müssen theoretische Konstrukte identifiziert werden, die sich einordnen und untersuchen lassen (Breiter & Hepp, 2018). Das führt erstens zu einer Suche nach passenden theoretischen Ansätzen im Sinne einer Theorieselektion oder zu neuer Theoriebildung, um den üblichen Forschungsparadigmen zu folgen (Waldherr, 2019). Zweitens lassen sich aufgrund der nicht selbst bestimmten Datenlage neue methodische Herausforderungen vermuten (Breiter & Hepp, 2018; Schneider & Otto, 2007). Bei den im Rahmen der Verdatung entstandenen MA-Trenderhebungen sind beispielsweise umfangreiche kommunikationswissenschaftliche längsschnittliche Harmonisierungen durchgeführt worden (Hagenah & Akinci, 2003), um neuartige Sekundäranalysen auf Basis von kumulierten Gesamtdatensätzen über mehrere Jahrzehnte zu ermöglichen (Hagenah & Meulemann, 2009). Im Zuge der Datafizierung zeigt sich diesbezüglich der sogenannte Computational Turn (Domahidi et al., 2019; Waldherr, 2019), weil neue, automatisiert erhobene Daten entsprechende Auswertungsstrategien erfordern, die oftmals anderen Disziplinen entstammen (Strippel et al., 2018). Nach Waldherr (2019) sind es auch und gerade die Forschungsmethoden, die Wissenschaft vorantreiben: Mit neuen Untersuchungsmöglichkeiten und Daten entwickeln sich neue Fragestellungen und tragen zur Theorieentwicklung bei. Auf institutioneller Ebene entwickeln sich Schulen und Paradigmen, die, nicht zuletzt unterstützt durch den technologischen Fortschritt, zur Entstehung neuer Infrastrukturen führen (sollten).

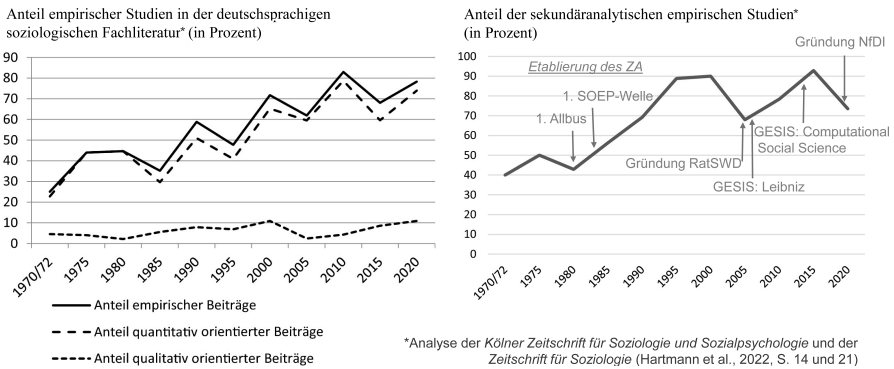
3. Verdatung, Datafizierung und Forschungsdateninfrastruktur

3.1 Verdatung und Datafizierung in den Sozialwissenschaften

Um den Wandel in der Kommunikationswissenschaft genauer in den Blick nehmen zu können, ist es hilfreich, Trends in verwandten Sozialwissenschaften zu beobachten. Eine inhaltsanalytische Untersuchung der wichtigsten deutschsprachigen sozialwissenschaftlichen

Zeitschriften *Kölner Zeitschrift für Soziologie und Sozialpsychologie* und *Zeitschrift für Soziologie* zeigt einen Anstieg an empirischen und quantitativ orientierten Studien ab etwa 1985 und einen langsamen, aber kontinuierlichen Anstieg der sekundäranalytischen empirischen Beiträge ab 1980. Seit den 1990er Jahren liegt deren Anteil bei 70 bis 90 Prozent der Beiträge (Hartmann et al., 2022). Diese Trends könnten einerseits mit einem Anstieg der Verdatung des sozialen Verhaltens in der Soziologie und mit einer steigenden Verbesserung der Forschungsdateninfrastruktur zusammenhängen (Abb. 1).

Abbildung 1: Soziologische Forschung: Entwicklung der Zeitschriftenpublikationen in einer sich wandelnden Forschungsdateninfrastruktur



*Analyse der *Kölner Zeitschrift für Soziologie und Sozialpsychologie* und der *Zeitschrift für Soziologie* (Hartmann et al., 2022, S. 14 und 21)

Um die aktuelle Forschungslandschaft in der Soziologie (und der Politikwissenschaft) einordnen zu können, ist es zielführend, die Geschichte der sozialwissenschaftlichen Forschungsdateninfrastruktur zu betrachten. 1960 wurde das Zentralarchiv für empirische Sozialforschung (ZA) an der Universität zu Köln gegründet (Mauer, 2012, S. 198), als eine der ersten Einrichtungen, die sich auf die Archivierung und Bereitstellung von sozialwissenschaftlichen Umfragedaten spezialisierte und somit die *Verdatung* der Gesellschaft – vor allem durch repräsentative Umfragen zu soziologischen oder politikwissenschaftlichen Themen – für die wissenschaftliche Nachnutzung verfügbar machte. Ausgehend von einer sich etablierenden Infrastruktur wurde, initiiert vom ZA, in Zusammenarbeit mit anderen Forschungseinrichtungen 1980 mit dem ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften; Mochmann, 2022; S. 290) ein kontinuierliches sozialwissenschaftliches Trendsurveyprogramm eingeführt. Der ALLBUS enthält repräsentative Daten über Einstellungen, Verhalten und soziale Strukturen der in Deutschland lebenden Bevölkerung und wird in der Regel alle zwei Jahre erhoben. Damit kann er als erste große sozialwissenschaftliche „Large Scale Study" in Deutschland bezeichnet werden (Kämper & Brüggemann, 2009). Zusätzlich zu den Sekundäranalysen einzelner archivierter Studien der Markt- und Meinungsforschung sorgt sie für ein von der Wissenschaft für die Wissenschaft konzipiertes kontinuierliches Datenerhebungsprogramm. 1984 wurde die erste Welle des Sozioökonomischen Panels (SOEP) erhoben (Wagner et al., 2007, S. 148), eine repräsentative Längsschnittstudie privater Haushalte in Deutschland, die ursprünglich vom Deutschen Institut für Wirtschaftsforschung (DIW) Berlin in Zusammenarbeit mit mehreren Forschungseinrichtungen konzipiert wurde und seitdem jährlich mit neuen Wellenbefragungen aktualisiert wird (ebd., S. 146; Goebel et al., 2019). In der Folge sind einige weitere „Large Scale Studies" auf den Weg gebracht worden (Kämper & Brüggemann, 2009, S. 3), z. B. das sog. Familienpanel (Panel Analysis of Intimate Relations and the Family – PAIRFAM), das *Na-*

tionale Bildungspanel (National Education Panel Study, NEPS), die sog. Wahlstudie (*German Longitudinal Election Study,* GLES) oder Erhebungsprogramme mit internationaler Ausrichtung, wie der *European Social Survey* (ESS) und der *Survey of Health, Ageing and Retirement in Europe* (SHARE).

Der Einfluss der mit diesen Studien verbundenen methodischen Entwicklungen und daraus resultierenden Infrastrukturen spiegelt sich in den Forschungspraktiken der jeweiligen wissenschaftlichen Disziplinen merklich wider: Für den Zeitraum von 1970 bis Mitte der 1990er Jahre lässt sich ein kontinuierlicher Anstieg von Sekundäranalysen in den soziologischen Zeitschriften nachweisen: Der Anteil steigt von 40 auf 90 Prozent. Von 2000 bis 2005 ist der Anteil an sekundäranalytischen Studien auf 70 Prozent leicht gesunken. Die Verdatung der Gesellschaft und die damit verbundene Institutionalisierung der dazugehörigen Bedarfe hat offensichtlich zu einem wachsenden Forschungsbedarf in den Sozialwissenschaften geführt, der flankiert wird durch eine wachsende Forschungsdateninfrastruktur. Nach der Gründung des Rates für Sozial- und Wirtschaftsdaten im Jahr 2004 (konsorts wd.de, 2025a; zum politischen Bedarf: Hauser et al., 1998) sind zahlreiche akkreditierte (Forschungs-)Datenzentren entstanden (konsortswd.de, 2025b).

Parallel dazu wurde im Jahr 2007 beschlossen, das Zentralarchiv für empirische Sozialforschung neben dem Informationszentrum Sozialwissenschaften (IZ) und dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) als Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e. V. (GESIS) in die Leibniz-Gemeinschaft aufzunehmen (Wissenschaftsrat, 2021, S. 7), seitdem sind Infrastruktureinrichtungen zu allen Phasen des empirischen Forschungsprozesses in den Sozialwissenschaften unter einem Dach vereinigt.

Die empirischen Folgen der *Datafizierung* in den Sozialwissenschaften sind – aufgrund der kürzeren Datengeschichte – weniger klar nachzuzeichnen. Die Abteilung *Computational Social Science* (CSS) bei der GESIS wurde 2013 gegründet und seitdem erheblich ausgebaut. Laut Wissenschaftsrat (2021, S. 12) lässt sich diese Entwicklung mit der steigenden Relevanz digitaler Verhaltensdaten begründen, erkennbar beispielsweise am Anteil von Publikationen zu digitalen Verhaltensdaten in hochrangigen Fachzeitschriften der Soziologie und Politologie, der sich seit 2014 mehr als verdoppelt hat. Außerdem befasse sich die CSS seitdem bei der Bearbeitung sozialwissenschaftlicher Fragestellungen mit neuen Methoden wie Webscraping, Textmining, Sensordatenerhebung, Netzwerksimulationen sowie Machine Learning. In einem Brückenschlag versucht die GESIS zudem die aus der Verdatung der Gesellschaft stammenden (Befragungs-)Daten, wie das seit 2014 existierende GESIS-Panel (Bosnjak et al., 2017), mit digitalen Verhaltensdaten zu verknüpfen und hat dafür das *GESIS Panel.dbd Digital Behavioral Data Sample* eingeführt, das u. a. Web-Tracking-Daten enthält (gesis.org, 2025).

3.2 Dateninfrastruktur und Fachentwicklung in der Kommunikationswissenschaft

Wesentlich kürzer sind die Infrastrukturprozesse der medialen Verdatung und Datafizierung in der Kommunikationswissenschaft einzuordnen. Das ist insofern bemerkenswert, als dass vor allem die deutschsprachige und europäische Kommunikationswissenschaft eine starke inhaltliche Tradition der Erforschung von Mediatisierungsprozessen aufweist (Couldry & Hepp, 2017; Döveling et al., 2024; Krotz, 2017). Nach Strippel (2021, S. 136) hat es „die Kommunikationswissenschaft im deutschsprachigen Raum bislang weitgehend versäumt, Forschungsinfrastrukturen aufzubauen, die den besonderen Anforderungen der Kommunikations- und Medienforschung im Fach gerecht werden". Zu Beginn des neuen Jahrtausends forderte Edmund Lauf (2002) in einem Aufsatz in dieser Zeitschrift „Freiheit für die Daten!" und bescheinigte der Kommunikations- und Medienwissenschaft Defizite bei der Freigabe, nachhaltigen Nutzung und Archivierung von Daten. Insbesondere wies er

darauf hin, dass archivierte Daten, wie etwa aus dem Zentralarchiv, nur selten für kommunikationswissenschaftliche Sekundäranalysen herangezogen wurden. Wenn Daten weitergegeben und nachgenutzt wurden, geschah dies viel häufiger über persönliche Netzwerke einzelner Forschender (ebd., S. 255). Das liegt aber nicht an einem fehlenden Forschungsinteresse: Die angewandte Medienforschung hat schon frühzeitig mit einer Verdatung des medialen Verhaltens begonnen. Im Rahmen einer Kölner Initiative hat sich insbesondere der Soziologe Heiner Meulemann in der Zeit von 2002 bis 2012 für eine kommunikationswissenschaftliche Aufbereitung der wichtigsten Reichweitendaten im Medienfeld eingesetzt (Strippel, 2021, S. 141–142). In vier Drittmittelprojekten wurden die Daten der Media-Analyse in SPSS konvertiert, über die Zeit von 1954 bis 2009 harmonisiert und zu kumulierten Gesamtdatensätzen aufbereitet und der wissenschaftlichen Community für Sekundäranalysen zur Verfügung gestellt (Hagenah, 2005, 2006a, 2006b, Hagenah et al., 2006, Hagenah & Meulemann, 2009). Als gewähltes Mitglied im RatSWD hat Meulemann eine „Agenda für Aktivitäten der DGPuK und des RatSWD" (Strippel, 2021, S. 141–142) und ein Gutachten mit Empfehlungen zur Dateninfrastruktur geschrieben (Meulemann & Hagenah, 2010). Ähnlich wie eine „Initiative zur Einbindung der DGPuK in den RatSWD" und eine „Initiative Infrastruktur Inhaltsanalyse" (Strippel, 2021, S. 143-145) blieb die Kölner Initiative letztlich ohne langfristige Folgen. Gleichwohl hat sich gezeigt, dass die temporär vorhandene Forschungsdateninfrastruktur zu einer sichtbaren „Forschungswelle" geführt hat, die danach wieder abgeebbt ist (Hagenah & Meulemann, 2006, 2008, 2012).

Obwohl viele Datenerhebungen zur Mediennutzung und einige wenige inhaltsanalytische Datenerhebungen identifiziert wurden (Meulemann & Hagenah, 2010; Hagenah et al., 2023), fehlt eine systematische kommunikationswissenschaftliche Aufbereitung der vielfältigen Datenlandschaft in der angewandten Medienforschung. Da es keine von der kommunikationswissenschaftlichen Scientific Community verantwortete „Large Scale Studies" vergleichbar mit dem ALLBUS oder dem SOEP gibt (Strippel, 2021, S. 137), kann nur auf die wenigen (unregelmäßigen) Abfragen zur Mediennutzung in den soziologischen Trend- und Panelstudien verwiesen werden. Generell gilt, dass die vorhandene Forschungsdateninfrastruktur auch offen für kommunikationswissenschaftliche Forschungsdaten ist. Beispielsweise werden die Daten der Media-Analyse seit Jahrzehnten bei der GESIS archiviert. Die soziologische und politikwissenschaftliche Ausrichtung führt aber dazu, dass die MA-Datensätze kaum noch genutzt werden. Der notwendige kommunikationswissenschaftliche Datenaufbereitungsaufwand konnte bisher nur durch externe Drittmittelprojekte aus dem Fach geleistet werden, bei denen eigene Publikationsaktivitäten (z. B. Hagenah & Meulemann, 2012) die Arbeiten auch dokumentierten und zu Sekundäranalysen anregten. Die letzte externe Datenaufbereitung hat für eine Nutzbarkeit von ausgewählten Datenteilen zur Presse- und Radionutzung bis zur MA 2015 geführt (Brentel & Jandura, 2018; Jandura & Brentel, 2018). Die geringe Nachnutzung gilt auch für die sog. *MA Intermedia PLuS*, die neben fusionierten Daten aus den Tranchen *MA Audio, MA Tageszeitungen, MA Pressemedien, MA Out of Home* und *AGF-Fernsehforschungspanel* auch Daten aus der *MA Internet* beinhaltet (agma-mmc.de/media-analyse/ma-intermedia-plus), die mit digitalen Verhaltensdaten angereichert ist (www.agma-mmc.de/media-analyse/dna). Hier verschmelzen die Verdatung des medialen Verhaltens, die durch die angewandte Medienforschung vorangetrieben wird, und die Datafizierung der digitalen Mediennutzung immer stärker: Bei der *AGF-Fernsehforschung* und der *MA Audio* wird zunehmend auch die Streamingnutzung der reichweitenstärksten Anbieter integriert.

Die Datafizierung hat mit einer zunehmenden Digitalisierung der Gesellschaft auch zu einer Explosion an theoretisch zu analysierenden Daten geführt (Breiter & Hepp, 2018; Domahidi et al., 2019). Parallel findet ein reger Diskurs über die Anwendung von Open Science

im Fach statt (Dienlin et al., 2021; Schäfer & Rössler, 2014), begleitet von Befragungen der wissenschaftlichen Community zum Status Quo ihrer Anwendung (Bowman et al., 2022; Sommer et al., 2025). So sollen Forschungsdaten den FAIR-Prinzipien zufolge *findable, accessible, interoperable* und *re-usable* sein, was erneut auf den Bedarf an entsprechender Forschungsdateninfrastruktur verweist (Strippel, 2021). Aufgrund des Fehlens von zentralen, genuin kommunikationswissenschaftlichen Dateninfrastruktureinrichtungen muss der Blick zur Beschreibung der Datafizierungsprozesse in unterschiedliche Richtungen gehen. Bisher gehören bspw. digitale Verfahren in der Kommunikationswissenschaft noch nicht zum publizierten Lehrstandard, wenn man die geltenden Empfehlungen der AG Methodenausbildung der DGPuK betrachtet (Matthes et al., 2011, 2012), die offensichtlich aktualisiert werden müssten. Ein grober Blick auf die Tagungen der Fachgruppe *Methoden* und die Jahrestagungen der DGPuK zeigt jedoch, dass hier eine systematische Erfassung der (offensichtlich gewachsenen) bestehenden Bedarfe und Angebote noch aussteht. Prägend für das Fach sind Verfahren wie die automatisierte Inhaltsanalyse (Scharkow, 2013) und die Nutzung von Social-Media-Daten (z. B. Twitter; zuletzt: Telegram). Web-Scraping/-Tracking oder die Analyse von sonstigen digitalen Spuren bspw. durch Datenspenden liegen seit Jahren im Trend und können als Folge der Digitalisierung und Datafizierung angesehen werden (Haim, 2023). Eine zentrale Forschungsdateninfrastruktur für die Kommunikationswissenschaft im Feld der Digitalisierung und Datafizierung wird häufiger gefordert, aber auch hier eher der GESIS und ihrer Abteilung für *Computational Social Science* (CSS) zugeschrieben. Erste (zentrale) Ansätze sind aber beim *Media Research Methods Lab* des *Leibniz-Instituts für Medienforschung | Hans-Bredow-Institut* (HBI) zu sehen, zumal es (ähnlich wie die GESIS) etablierte sozialwissenschaftliche Methoden mit neuartigen digitalen Verfahren verknüpft.

Die deutschsprachige Kommunikationswissenschaft hat sich von einer Integrationsdisziplin mit diversen akademischen Wurzeln (Brosius, 1994) sukzessive zu einer „Normalwissenschaft" (Brosius & Haas, 2009) entwickelt, die sich durch ein klares Selbstverständnis im akademischen Feld behauptet (Brosius, 1994; Brosius & Haas, 2009; Donsbach et al., 2005). Zugleich hat sich mit der starken Expansion seit den 1990er Jahren eine große Ausdifferenzierung des Faches ergeben (Koenen & Sanko, 2017; Waldherr, 2019), die begleitet wurde von einem erheblichen Internationalisierungsschub (Strippel & Domahidi, 2016) sowie der kollaborativen Wende hin zur Forschung im Verbund (Volk et al., 2022). Nicht zuletzt angestoßen von Digitalisierung und Datafizierung hat sich der Fachdiskurs in den vergangenen zehn Jahren wiederholt um die Frage gedreht, wie sich das Fach zukünftig positioniert, wenn Theorien, Begriffe und Methoden Fachgrenzen überschreiten und neu gedacht werden (Strippel et al., 2018). Verdatung und Datafizierung könnten hier – ähnlich wie in der Soziologie – zu einer empirischen Aufbruchsstimmung geführt haben. Während die Verdatung der öffentlichen Kommunikation eher einen Anstieg an Sekundäranalysen nahelegt, weil hier Daten vorliegen, die mit dem etablierten methodischen Instrumentarium der Kommunikationswissenschaft erhoben wurden und schon vergleichsweise lange im System verfügbar sind, geben datafizierte Datenbestände vermutlich eher einen Impuls für neue methodische Ansätze explorativer Natur und inspirieren die Anwendung von Computational Methods.

Um solche Prozesse nachzuzeichnen, widmet sich die nachfolgend dargestellte Untersuchung, orientiert an den inhaltsanalytischen Fachbeobachtungen von Brosius und Kolleg:innen (Brosius, 1994; Brosius & Haas, 2009; Donsbach et al., 2005), den publizierten Forschungspraktiken in den deutschsprachigen Fachzeitschriften der Kommunikationswissenschaft.

4. Methode

In Anlehnung an die bereits bestehenden Inhaltsanalysen zur Fachentwicklung (u. a. Brosius, 1994; Donsbach et al., 2005; Brosius & Haas, 2009; Haas et al., 2018) wurde aufbauend auf Methodik und Daten der Vorgängerstudien bis 2017 (Haas et al., 2018) eine Vollerhebung aller wissenschaftlichen Originalbeiträge in *Publizistik, Medien & Kommunikationswissenschaft (M&K)* und *Studies in Communication and Media (SCM)* der Jahre 2003 bis 2023 vorgenommen (n=935). Der von Haas et al. (2018) zur Verfügung gestellte Datensatz nebst Codebuch diente als Grundlage und wurde für die Neu- und Nacherhebungen ab 2017 angepasst. Die Tabelle im Anhang visualisiert, welche Variablen (teilweise unter anderem Namen) bereits vorhanden waren und welche Variablen von den Autor:innen ergänzt wurden. Unberücksichtigt in der Tabelle sind die Variablen, die sich auf Sekundäranalyse und -daten beziehen, da diese im Rahmen dieser Studie per se neu erfasst wurden.

Die SCM fließt neu in die Analysen ein. Die Studie beschränkt sich auf diese drei Fachzeitschriften, weil sie die aktuelle Forschung der Kommunikationswissenschaft reflektieren, die publizierten Beiträge in der Regel wissenschaftlichen Qualitätskriterien unterliegen und sich alle drei Zeitschriften nicht auf ein bestimmtes Thema innerhalb der Kommunikationswissenschaft beschränken. Sie fungieren damit als „nerves of the disciplines" (Riffe & Freitag, 1997) und können ein authentisches Abbild der deutschsprachigen Publikationskultur in dieser Zeit liefern, auch wenn die zunehmende Internationalisierung die Praktiken des wissenschaftlichen Veröffentlichens zweifellos verändern (s. u.).

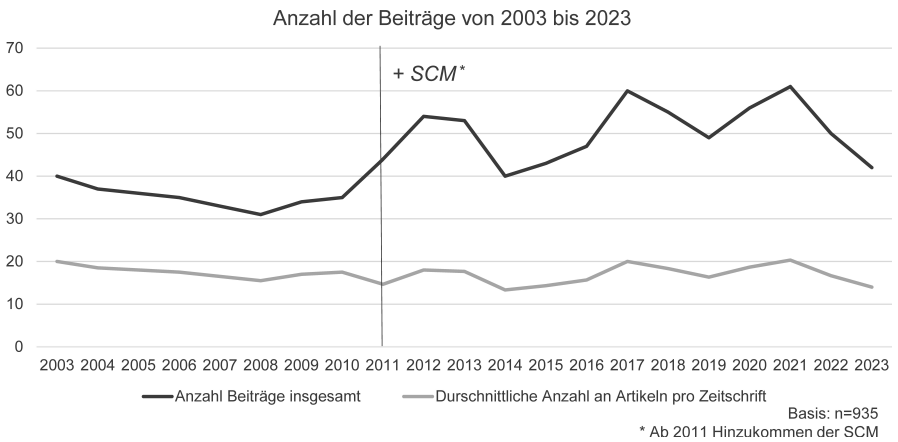
Der Untersuchungszeitraum umfasst zwei Jahrzehnte, die von einer umfassenden Datafizierung gekennzeichnet sind. Der Beginn der Untersuchung lässt sich einerseits mit der Struktur der weitergeführten Datenerhebung (Haas et al., 2018) erklären und andererseits mit den durch die Digitalisierung gewachsenen Social-Media-Strukturen, die v. a. durch die Gründung von Twitter im Jahr 2006 (Simon & Bernhardt, 2008, S. 39–44) und die Facebook-Innovation im Jahr 2004 zu einer sozial relevanten Einflussgröße wurden (Newman et al., 2017). Bezogen auf die Verdatung und Forschungsdateninfrastruktur schließt der Zeitraum die Arbeit des MLFZ ebenso mit ein wie die seit der Gründung des RatSWD intensiveren Entwicklungen der sozialwissenschaftlichen Infrastruktur.

In die Untersuchung gingen alle wissenschaftlichen Originalbeiträge der drei Zeitschriften der Jahre 2003 bis 2023 (SCM seit 2011) ein. Unberücksichtigt blieben Buchbesprechungen, Rezensionen, Editorials, Chroniken, Mitteilungen und Errata. Insgesamt ergab sich für den Zeitraum von 2003 bis 2023 ein Sample von 935 Aufsätzen, wobei 346 auf die *Publizistik*, 375 auf die *M&K* und 214 auf die *SCM* entfallen. Die durchschnittliche Anzahl an Beiträgen pro Jahr und pro Zeitschrift liegt bei 17 Artikeln (Abb. 2).

Dabei wurde die Untersuchungsebene von der ursprünglichen Autor:innenzentrierung (Brosius, 1994; Donsbach et al., 2005; Haas et al., 2018) auf die Beitragsebene verschoben, wenngleich einige beitragszentrierte Variablen, wie etwa die Erhebungsmethode, im Originaldatensatz bereits erfasst worden waren. Zusätzlich zu den Variablen der früheren Studien (z. B. empirisch/nicht-empirisch, Erhebungsmethode) wurden detaillierte methodische Codierungen vorgenommen, die etwa zwischen *Primär- und Sekundäranalyse, Quer- und Längsschnitt* differenzieren und die *verwendeten Datenarten und -quellen* sowie die *Auffindbarkeit der Daten* erfassen.[1] Nicht zuletzt aufgrund der Verschiebung der Untersuchungsebene war eine Anpassung des „alten" Codeplans von Haas et al. (2018) an das neue Kategoriensystem notwendig, wodurch sich auch im Zeitraum bis 2017 umfangreiche

1 Das Codebuch kann online unter https://osf.io/9sr7m/?view_only=9542a23ac9f449d781f98030735b53ea abgerufen werden.

Abbildung 2: Anzahl der Beiträge von 2003 bis 2023 (n=935)



Nacherhebungen ergeben haben. Insgesamt lassen sich die pro Beitrag erfassten Variablen in drei Bereiche einteilen:

1. Zunächst wurden formale und inhaltliche Merkmale der Aufsätze erfasst. Dazu gehören der thematische Schwerpunkt, das untersuchte Medium, das Forschungsobjekt sowie die Anzahl an deutsch- und englischsprachigen Quellen.
2. Außerdem wurden detaillierte methodische Merkmale wie *empirisch/nicht-empirisch*, die Art der Methode, die Verwendung von Primär- oder Sekundärdaten, die Datenherkunft etc. erhoben.
3. Zuletzt wurden wichtige Merkmale der Autor:innen erfasst, wobei eine Fokussierung auf den Erstautor:innen lag. So wurden Name, gelesenes Geschlecht, Institution, Stellung und Anzahl an Publikationen der jeweiligen Erstautor:innen erfasst. Zusätzlich dazu wurden auch die Anzahl an mitwirkenden Autor:innen und die jeweiligen Stellungen und gelesenen Geschlechter erfasst.

Die Codierung wurde vorwiegend von zwei wissenschaftlichen Mitarbeiterinnen durchgeführt und koordiniert (Intercoderreliabilität nach Holsti = 0,93[2]). Die formalen Variablen, wie Anzahl der Autor:innen, Titel des Aufsatzes, Anzahl der Quellen etc. wurden von einem Team aus studentischen Hilfskräften unter Supervision der Autor:innen codiert.[3]

5. Ergebnisse

Um den möglichen Einfluss von Verdatung und Datafizierung auf die kommunikationswissenschaftlichen Forschungspraktiken zu eruieren, wurde zunächst der Anteil veröffentlichter empirischer Beiträge über die vergangenen 20 Jahre hinweg betrachtet. Ausgangspunkt war hier die Annahme, dass die zunehmende Menge und Vielfalt an Daten Forschungsanlässe und -impulse gibt (Kap. 2). Anschließend wurden diese empirischen Arbeiten genauer charakterisiert, um einen Einblick zu erhalten, inwiefern die größere Datenviel-

2 Spannweite von 0,5 bis 1 (siehe Tabelle 1 im Anhang).

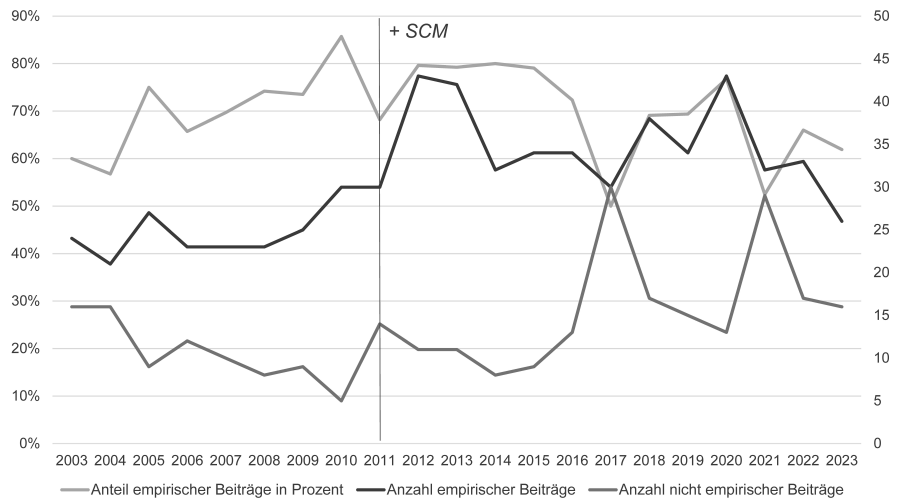
3 Wir danken Svenja Marquardt für ihre wertvolle Unterstützung bei der Konzeption, Koordination und Durchführung der Inhaltsanalyse sowie Mika Maria Eckel, Jana Harig, Lavinia Pissula, Ruben Reckel, Daniel Schell und Emely Vysoudil für ihre Beiträge zu Recherche und Codierung.

falt die angewandten Analysemethoden verschiebt, z. B. in Richtung explorierender oder computergestützter Auswertungsverfahren. Hier sollten sich vor allem die jüngsten Trends der Digitalisierung und Datafizierung identifizieren lassen. Im letzten Schritt wurde der ältere und zeitlich weniger eindeutige Verdatungstrend aufgegriffen und im Zusammenhang mit bestehenden und sich entwickelnden Dateninfrastrukturen geprüft, ob sich analog zur Soziologie ein Aufschwung von sekundäranalytischer Forschung mit der Zunahme an Datenbeständen zeigt (Kap. 3).

5.1 Empirische Aufbruchsstimmung?

Rund 70 Prozent der analysierten Beiträge (n=647) wurden als empirische Untersuchungen eingeordnet. Der Anteil lässt sich im Zeitverlauf als stark schwankend charakterisieren. In den Nullerjahren des neuen Jahrtausends lässt sich ein starker Anstieg verzeichnen, der durch das Hinzukommen der *SCM* im Jahr 2011 noch verstärkt wird (Abb. 3). Danach ist allerdings ein Abwärtstrend zu beobachten, so dass sich der Anteil wieder bei circa 60 Prozent einpendelt. Insbesondere zeigt sich, dass die absolute Anzahl nicht-empirischer Beiträge nach einem Rückgang in den Nullerjahren in den Jahren nach 2011 wieder zugenommen hat. Die Spitzenwerte nicht-empirischer Beiträge in 2017 und 2021, die pro Jahr circa 30 Aufsätze beinhalten, lassen sich durch zwei Themenhefte[4] mit theoretischem Schwerpunkt erklären.

Abbildung 3: Anteil und Anzahl empirischer Beiträge von 2003 bis 2023 (n=935)



Im direkten Vergleich der drei Zeitschriften zeigt sich, dass die *SCM* mit einem Anteil von 78 Prozent die stärkste empirische Ausrichtung aufweist, während *Publizistik* und *M&K* jeweils etwa zwei Drittel empirische Beiträge enthalten (vgl. Tab. 1 im Anhang).

Anteil an dieser Entwicklung dürften dabei die Publikationsformate und -bedingungen der Zeitschriften haben. So hat die *SCM* etwa eine Rubrik *Research in Brief*, die für kurze

4 Konstruktivismus in der Kommunikationswissenschaft (M&K, 2017, Heft 2); Forschungssoftware in der Kommunikations- und Medienforschung (M&K, 2021, Heft 1).

Tabelle 2: Anteil der empirischen Beiträge im Zeitschriftenvergleich (n=935) in Prozent

Zeitschrift	Anteil empirischer Beiträge	Anteil nicht-empirischer Beiträge
Publizistik	66,8 %	33,2 %
M&K	66,4 %	33,6 %
SCM	78,0 %	22,0 %
Gesamt	**69,2 %**	**30,8 %**

empirische Arbeiten in Form von Werkstattberichten und Zwischenständen größerer empirischer Erhebungen genutzt wird. Zudem entsteht die Zeitschrift überhaupt erst zu einem Zeitpunkt, als sich standardisierte sozialwissenschaftlich-empirische Forschungspraktiken im Fach schon etabliert haben (Brosius & Haas, 2009), und ist seit ihrem Bestehen 2011 als Open-Access-Zeitschrift angelegt. 2021 folgten M&K und Publizistik. Im Zuge des Trends zu Open Science erhöht sich vermutlich die Attraktivität einer Veröffentlichung von Forschungsergebnissen und den zugrundeliegenden Daten in einer kostenlos zugänglichen Fachzeitschrift. Da der Untersuchungszeitraum mit dem Jahr 2023 endet, sollte zukünftige Forschung prüfen, ob sich die Anteile der empirischen Studien deshalb in den kommenden Jahren angleichen.

5.2 Merkmale der empirischen Aufbruchsstimmung

Nachdem eine potenzielle empirische Aufbruchsstimmung insbesondere durch die gestiegene absolute Zahl an empirischen Studien angenommen werden kann, werden ihre Merkmale genauer betrachtet. Dabei werden die empirischen Arbeiten (n=647) anhand ihrer Forschungsanlage und Datenbasis charakterisiert. Wie aus Abb. 4 ersichtlich wird, dominieren die empirisch-quantitativen Forschungsanlagen die analysierten Beiträge deutlich, sind allerdings in ihrem Anteil über die 20 Jahre rückläufig. Insbesondere in der zweiten Hälfte des Untersuchungszeitraums steigt dagegen der Anteil qualitativer Zugänge merklich an. Kombinationen beider Herangehensweisen sind in den Nullerjahren noch stärker oder in gleichem Maße auffindbar wie rein qualitative Ansätze, verlieren aber zugunsten der qualitativen Verfahrensweisen ab den Zehnerjahren an Bedeutung. Mit Blick auf die Datafizierung kann hier vermutet werden, dass neue Datenformen und -spuren den Bedarf an explorativen Zugängen erhöhen, bevor standardisierte Vorgehensweisen entwickelt und etabliert werden (vgl. Breiter & Hepp, 2018).

Im Zeitverlauf ergibt sich ein Anstieg von Querschnittstudien bei einem gleichzeitigen Rückgang von Längsschnittstudien (Abb. 4), was vor allem auf den parallel verlaufenden Rückgang an Sekundäranalysen zurückzuführen ist. Der vergleichbare Verlauf der beiden Graphen deutet auf einen Zusammenhang hin, der sich bei der Betrachtung der Kreuztabellen als höchst signifikant ($p < 0{,}001$) manifestiert. So sind 24,8 Prozent der Längsschnittstudien auch Sekundäranalysen, wobei dies nur auf 9,2 Prozent der Querschnittstudien zutrifft.

Neben der Datenart wurde die methodische Herangehensweise der veröffentlichten empirischen Beiträge untersucht, um mögliche Veränderungen in den Untersuchungsanlagen über die Zeit zu prüfen. Hier zeigt sich zunächst ein erwartbares Bild: Die Befragung stellt die am häufigsten angewandte Methode dar, dicht gefolgt von der Inhaltsanalyse. Ein leichter prozentualer Rückgang bei der Verwendung von Befragungen wird über die Zeit deutlich, wobei die absoluten Zahlen stabil bleiben (Abb. 5). Die Inhaltsanalyse erlebt zugunsten des prozentualen Rückgangs bei der Befragung dagegen einen leichten Aufschwung (2003: 46 %; 2023: 58 %), der sich auch in den absoluten Zahlen abbildet. Experimentelle Untersu-

chungsanlagen nehmen im Zeitraum der vergangenen 20 Jahre ab (2003: 21 %; 2023: 4 %), wohingegen die Computational Methods[5] insbesondere ab 2013 einen Aufschwung erleben (2003: 0 %; 2023: 27 %)[6].

Abbildung 4: Quantitative, qualitative, kombinierte Forschungsansätze und Längs- und Querschnittdaten mit Sekundäranalyse im Zeitverlauf (n=647)

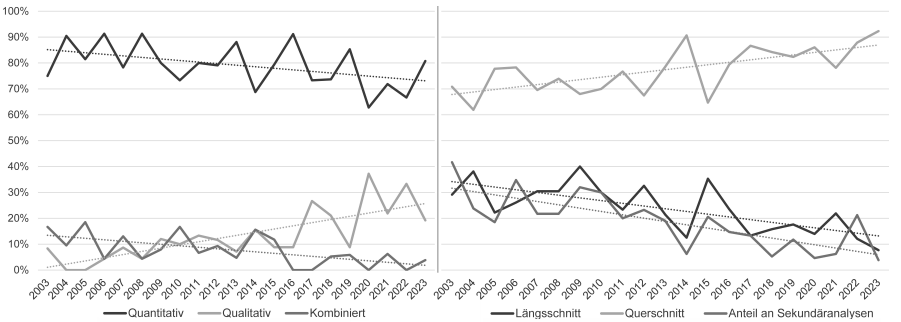
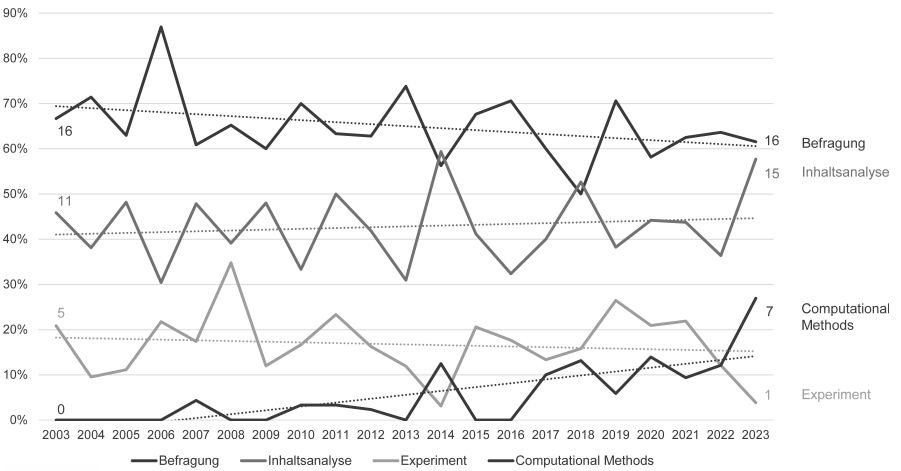


Abbildung 5: Anteile der angewandten Forschungsmethoden im Zeitverlauf (n=647)



Anmerkung: Die Zahlen in den Jahren 2003 und 2023 stellen zur Veranschaulichung die jeweils absoluten Werte dar.

5 Folgende Definition von Computational Methods wurde bei der Codierung beachtet: Computational Methods sind die Automatisierung wissenschaftlicher Verfahren durch maschinelle Ersetzung. Dabei geht es nicht um klassische Methoden, sondern um große und komplexe, digitale oder natürliche Datensätze. Computational Methods basieren auf algorithmischen Lösungen zur Datenerhebung, -aufbereitung, -analyse und interaktiven/individuellen Darstellung und stammen aus einer (un)überwachten technischen Quelle. Beispiele: Webscraping über Application Programming Interfaces (APIs), Dokument-Term-Matrizen (Netzwerkanalysen, Zeitreihenanalysen), Computersimulationen, Topic Modeling, geographische Analysen.

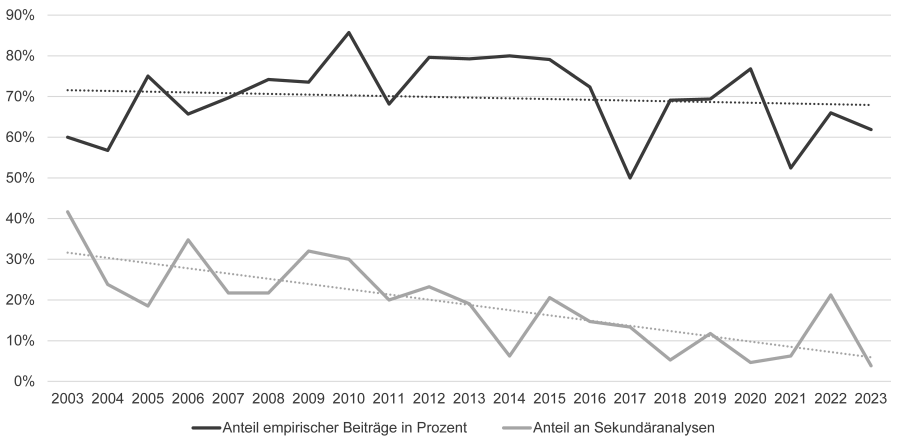
6 In der Codierung wurden auch die Beobachtung und die Gruppendiskussion als eigene Methoden erfasst. Sie kommen jedoch nur sehr vereinzelt vor und werden deshalb an dieser Stelle nicht gesondert ausgewiesen.

Damit lassen sich auf der Detailebene erste Auswirkungen der Datafizierung erahnen: Zum einen deutet der Anstieg von inhaltsanalytischen Untersuchungen parallel zum leichten Rückgang des Anteils an Befragungen darauf hin, dass mehr kommunikative Botschaften zur Verfügung stehen, deren inhaltsanalytische Auswertung aus wissenschaftlicher Sicht interessant für das Fach sind. Dies lässt sich mit den vielfältigen Spuren, die Mediennutzer:innen im Netz hinterlassen, plausibel begründen. Durch die digital fixierten Reaktionen von Mediennutzenden können Interaktionsprozesse und Medienwirkungen inhaltsanalytisch erfasst und direkt modelliert werden – ohne den Umweg der Befragung oder experimentellen Untersuchung. Zum anderen deutet der noch recht frische und sprunghafte Anstieg von Publikationen mit Computational Methods darauf hin, dass ein weiter Auf- und ggf. auch Umbruch der empirischen Kommunikationsforschung im Gange ist, der sich aufgrund der Datafizierung vermutlich weiter beschleunigen wird.

5.3 Verdatung als Treiber von Sekundäranalysen?

Dem mit kontinuierlichen Schwankungen über die Jahre relativ gleichbleibenden Anteil empirischer Beiträge steht – anders als in der Soziologie – ein deutlicher linearer Rückgang des Anteils an Sekundäranalysen gegenüber (Abb. 6). Waren 2003 noch 42 Prozent der empirischen Aufsätze sekundäranalytische Untersuchungen, sind es im Jahr 2023 nur noch vier Prozent. Dieses Bild bestätigen auch die absoluten Zahlen: Während von 2003 bis 2023 die absolute Anzahl der empirischen Beiträge insbesondere durch das Hinzukommen der *SCM* deutlich anstieg, verzeichnet die absolute Anzahl an Sekundäranalysen einen linearen Rückgang über die Zeit. Hier zeigt sich die „stiefmütterliche Behandlung" sekundäranalytischer Forschung im Fach, die Lauf (2002, S. 247) bereits zu Beginn der Nullerjahre beklagte.

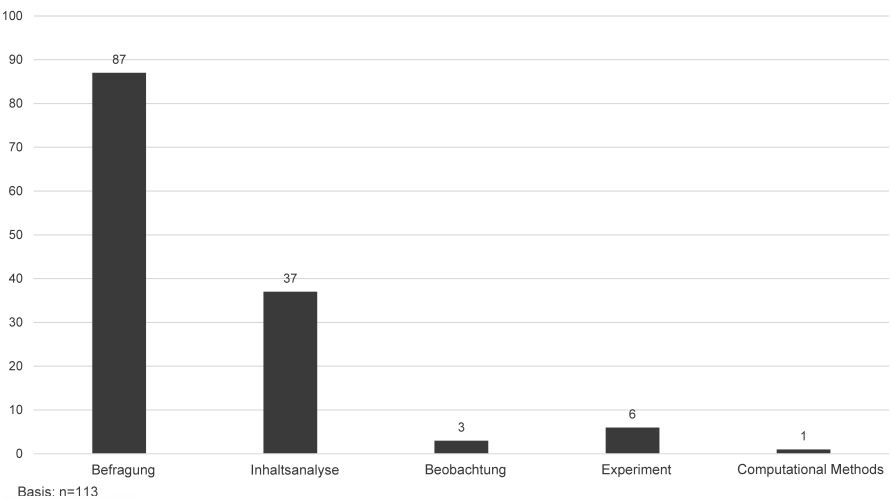
Abbildung 6: Anteil empirischer und sekundäranalytischer Beiträge (n=935)



Basis: n=935
Basis: n=647 empirische Beiträge

Um einen genaueren Einblick in die sekundäranalytische Forschungspraxis zu erlangen, wurden diese Beiträge (n=115) gesondert betrachtet. Zunächst wurde auch hier nach den methodischen Herangehensweisen gefragt, die die sekundäranalytisch verwendeten Daten hervorgebracht haben. Weit über die Hälfte der nachgenutzten Datensätze stammt dabei aus Befragungen, etwa ein Drittel aus Inhaltsanalysen. Damit dominieren die gängigen sozialwissenschaftlichen Methoden der kommunikationswissenschaftlichen Primärforschung auch die nachgenutzten Datenbestände (Abb. 7). Mit einem Aufschwung in den Nullerjahren und dem folgenden Rückgang lässt sich hier ein nachlassender Einfluss der klassischen Verdatungsprozesse annehmen.

Abbildung 7: Methoden, mit denen Sekundäranalysedaten erhoben wurden, von 2003 bis 2023 in absoluten Zahlen



Vor dem Hintergrund von Verdatung als Treiber von Forschungsdateninfrastruktur stellte sich zudem die Frage, woher die Daten kommen, die Gegenstand von Sekundäranalysen im Fach sind (Tab. 3). Die Ergebnisse decken sich weitgehend mit den Befragungsdaten von Lauf (2002) sowie aktuellen Befragungen zum Umgang mit Forschungsdaten im Fach (Sommer et al., 2025). Der überwiegende Teil der sekundäranalytisch ausgewerteten Daten entstammt eigenen Untersuchungen (33 %) oder wird von Kolleg:innen über persönliche Netzwerke bezogen (25 %). Dennoch spielen Daten der Media- und Marktforschung (19 % bzw. 15 %) sowie aus verschiedenen organisationalen Kontexten (16 %) keine unbedeutende Rolle in der sekundäranalytischen Forschung, was auf einen gewissen Datenbedarf von außen für die Kommunikationswissenschaft hindeutet. Angewandte und kommerzielle Quellen rangieren dabei tendenziell höher als die akademische Dateninfrastruktur (Rangplätze drei, fünf, sieben; ca. ein Drittel aller erfassten Studien). Dennoch zeigt sich auch ein Interesse an der Verwendung von Daten aus der wissenschaftlichen Forschungsinfrastruktur, wie der Verweis auf Datenreihen wie die GLES-Studie (7 %) oder die „Worlds of Journalism (WoJ)"- bzw. „Journalist:innen in Deutschland (JouriD)"-Studie (7 %) als eine der wenigen genuin kommunikationswissenschaftlichen Datenreihen nahelegen.

Die prozessgenerierten Daten der verdateten Medien liefern also zunächst keine Grundlage für eine verstärkte sekundäranalytische Forschung zu kommunikationswissenschaftli-

Tabelle 3: Berichtete Datenherkunft bei den Sekundäranalysen (n=115)

	Berichtete Datenherkunft	**Absolute Häufigkeit**
1	eigene Daten	33
2	Kolleg:innen	25
3	Media-Trendstudien	19
4	Organisation	16
5	sonstige Marktforschung	15
6	Universitärer Kontext	9
7	GfK	7
8	GLES	7
9	WoJ/JouriD	6
10	DFG	5
11	Unbekannt	5

chen Fragestellungen. Auch von der verbesserten Dateninfrastruktur in den deutschen Sozialwissenschaften profitieren diese Forschungsansätze offenbar bislang eher wenig.

6. Diskussion

6.1 Empirische Aufbruchsstimmung?

Die untersuchten Forschungsarbeiten lassen sich als stark empirisch ausgerichtet beschreiben. Die Tendenz der genuin empirischen Beiträge ist in absoluten Zahlen steigend, im relativen Anteil gleichbleibend auf einem hohen Niveau von etwa 70 Prozent. Quantitativ orientierte Forschungsansätze dominieren deutlich. Vor dem Hintergrund der empirisch-sozialwissenschaftlichen Wende im Fach in den 1960er Jahren (Löblich, 2010; Waldherr, 2019) und im Abgleich mit den deutlich weiter zurückreichenden Untersuchungen für die Soziologie (Hartmann et al., 2022, s. o.) erscheint es plausibel, dass empirische Forschung in der Kommunikationswissenschaft nach der Jahrtausendwende in der Scientific Community als Standard angesehen wird. Eine Aufbruchsstimmung ist allerdings durchaus beobachtbar, denn mit der Einführung der *SCM* im Jahr 2011 bleibt der hohe Anteil an empirischen Arbeiten bestehen, absolut gesehen wächst also die Zahl der veröffentlichten Studien. Das deutet darauf hin, dass der Bedarf an Publikationsmöglichkeiten für Forschungsergebnisse gewachsen ist, so dass auch eine dritte Zeitschrift mit einem ähnlichen inhaltlich breiten Profil neben *Publizistik* und *M&K* zusätzliche empirische Arbeiten veröffentlicht.

6.2 Kommunikationswissenschaftliche Forschungspraktiken in datengetriebenen Zeiten

Im Unterschied zur soziologischen Forschung ist der wachsende Anteil der empirischen Studien in der Kommunikationswissenschaft jedoch nicht gekoppelt an eine zunehmende Zahl von Sekundäranalysen. Im Gegenteil verlieren sekundäranalytische und längsschnittliche Untersuchungen im Verlauf der vergangenen 20 Jahre an Bedeutung zugunsten von Querschnittstudien. Diese Entwicklung wird begleitet von einem Trend hin zur verstärkten Anwendung qualitativer Untersuchungsanlagen und einem leichten, aber sichtbaren Rückgang quantitativer Studien, die die Stichprobe anteilig jedoch dominieren. Hier zeigt sich eine gewisse Gegenläufigkeit zur Soziologie, die in der großen Dynamik des Forschungs-

feldes öffentlicher und medialer Kommunikation begründet liegen könnte. Die tiefe Mediatisierung der Gesellschaft (Couldry & Hepp, 2017) und die ihr inhärenten disruptiven Veränderungen öffentlicher Kommunikation verlangen eine schnelle und flexible Anpassung der Forschung an neue mediale Trends, die sich im Bedarf an neuen Messmethoden und Datensammlungen niederschlägt. Qualitative Exploration ist hier oftmals unerlässlich, ebenso wie kurzfristige Querschnittstudien (Breiter & Hepp, 2018).

Zugleich zeigt sich ein wichtiges Potenzial von sekundäranalytischen Arbeiten und einer entsprechenden Forschungsdateninfrastruktur: Insbesondere im Zuge des fortlaufenden Mediatisierungsdiskurses im Fach (Couldry & Hepp, 2017; Krotz, 2017) ist von einem Bedarf an längsschnittlicher Forschung auszugehen, um der These von Medienwandel als wesentlichem Treiber gesellschaftlicher Veränderungen empirisch gerecht zu werden. Zugleich ist jedoch die Sammlung längsschnittlicher Daten in größerem Umfang und über längere Zeiträume wissenschaftlich wie infrastrukturell herausfordernd. Hier bieten prozessgenerierte Daten der angewandten Medienforschung oder der amtlichen Statistik, die teilweise auf extrem ausgedehnte Zeitreihen zurückgreifen können, einen großen Datenschatz, der offenbar immer noch vielfach ungenutzt bleibt (Lauf, 2002).

In den analysierten Forschungspraktiken spielen die Methoden der Befragung und der Inhaltsanalyse die wichtigste Rolle – sowohl in den Primärerhebungen als auch in den Sekundäranalysen. Damit bildet die untersuchte Stichprobe sehr wahrscheinlich charakteristisch die breite kommunikationswissenschaftliche Forschung ab. Während Befragungen einen leichten Abwärtstrend aufweisen und experimentelle Untersuchungsanlagen an Bedeutung verlieren, zeichnet sich bei inhaltsanalytischen Verfahren ein Aufwärtstrend ab. Einen starken Zugewinn innerhalb der jüngsten Phase der Untersuchung verzeichnen die Computational Methods, die einen Hinweis auf zukünftige Entwicklungen der zunehmend datafizierten Forschung geben. Dieser in den Publikationen der deutschsprachigen Fachzeitschriften noch sehr neue Trend zeigt den Bedarf an weiterführender Forschung zum Thema und regt dazu an, die hier vorgestellte Zeitreihe fortzuschreiben.

Erste Hinweise liefern die Befunde außerdem mit Blick auf die Forschungsdateninfrastruktur im Fach. Während die leichte Zunahme inhaltsanalytischer Forschung darauf hindeutet, dass digitale Botschaften der datafizierten Medien als Forschungsanlässe bedeutender werden, liegen Untersuchungen nachgenutzter Daten und Meta-Studien mit längeren Zeitreihen derzeit nicht im Trend. Die Bereitstellung von digitalen Spuren und Verhaltensdaten aus dem Netz ebenso wie von Befragungsdaten zu den aktuellen Themen der angewandten Forschung erfolgt noch unzureichend und wenig strukturiert (Strippel, 2021). Hier sind weitere Bemühungen für eine qualitativ und forschungsethisch hochwertige Archivierung relevanter Datenbestände erforderlich, um Forschende in einem sich immer weiter ausdifferenzierenden Feld der öffentlichen und medialen Kommunikation zu unterstützen. Wenn Kommunikationsforscher:innen derzeit auf bereits erhobene Daten zur Nachnutzung zugreifen, tun sie das primär über persönliche Netzwerke. Daran hat sich seit der Jahrtausendwende nur wenig geändert, wie der Vergleich mit der Untersuchung von Lauf (2002) zeigt. Die zwischenzeitlich verbesserte Infrastruktur von Forschungsdaten in den Sozialwissenschaften kommt ebenso, wie schon damals ermittelt, nur marginal zum Einsatz. Zugleich wird auf etablierte Daten aus der angewandten Forschung wie auf genuin kommunikationswissenschaftliche Daten, die inzwischen zur Verfügung stehen (Worlds of Journalism/ Journalist:innen in Deutschland), durchaus zurückgegriffen, wenn eine Bereitstellung erfolgt ist.

6.3 Limitationen und Ausblick

Schließlich sei auf die Limitationen dieser Untersuchung verwiesen, die eine Reihe von weiterführenden Fragen aufwerfen. Alle drei untersuchten Fachzeitschriften zielen darauf ab, das Fach im deutschsprachigen Raum in seiner thematischen Breite zu repräsentieren. Dennoch liegt es nahe, dass die untersuchten Zeitschriften ein eingeschränktes Bild der kommunikationswissenschaftlichen Forschung im deutschsprachigen Raum zeichnen. Die Publikationskultur der wissenschaftlichen Fachgemeinschaft hat sich im Untersuchungszeitraum grundlegend gewandelt: Es wird mehr und anders publiziert, wobei die Publikationsstrategien ebenso wie die Einstellungen der Kolleg:innen in der Fachgemeinschaft dazu extrem divers ausfallen – das zeigen nicht zuletzt zahlreiche fachinterne Debatten zum Peer Review, Publish or Perish und den verwendeten Publikationssprachen (Debatin & Rath, 2014; Prinzing & Herczeg, 2023; Schäfer & Rössler, 2014). Das Angebot an internationalen Zeitschriften zu spezialisierten Bereichen ist extrem groß, so dass Publikationen dort international sichtbar für ganz spezifische Expert:innenkreise platziert werden können. Diese Möglichkeiten stehen jedoch auch einem deutlich größeren Wettbewerb gegenüber, der wissenschaftlichen Output in renommierten internationalen Publikationsorganen verlangt (Prinzing & Herczeg, 2023). Das wirft die Frage auf, welche Beiträge aus welchen Gründen in den untersuchten deutschsprachigen Zeitschriften veröffentlicht werden und welche weiteren Forschungsarbeiten aus der Community, insbesondere aus der datafizierten Forschung, bspw. direkt international publiziert werden.

Eine Erweiterung der Untersuchung auf weitere Zeitschriften ist im nächsten Untersuchungsschritt denkbar – beispielsweise könnten zusätzlich thematisch spezialisierte Zeitschriften oder ähnlich angelegte Journals aus anderen Ländern zum Vergleich herangezogen werden. Das würde allerdings an anderer Stelle neue Unschärfen produzieren, da sich Forschungsdateninfrastrukturen wie auch wissenschaftliche Schulen und Paradigmen im Fach zwischen den verschiedenen Wissenschaftskulturen mitunter deutlich unterscheiden (Averbeck-Lietz & Löblich, 2017). Dies ist auch der Grund dafür, weshalb der Untersuchungszeitraum nicht weiter zurückreicht. Auch wenn Daten ab 1983 vorliegen, wurde für diese Untersuchung der Zeitraum von 2003 bis 2023 gewählt, um Beiträge untersuchen zu können, die von ihrer Anlage und Gestaltung ebenso wie den durchlaufenen Review-Prozessen vergleichbar sind. Damit bildet die Untersuchung in gewisser Weise den Common Sense der Scientific Community über Forschungstrends und wissenschaftliche Qualität der untersuchten Zeitspanne ab, vernachlässigt jedoch mit Sicherheit einige thematische Bereiche des Faches, deren Diskurse beispielsweise traditionell in anderen Publikationsorganen geführt werden oder deren wissenschaftliche Herangehensweisen sich weniger gut für die Veröffentlichung in den untersuchten, recht standardisierten Fachzeitschriften eignen.

Letztlich verdeutlichen diese Einschränkungen den Nutzen weiterer Arbeiten zum Thema, die neben ergänzenden inhaltsanalytischen Untersuchungen auch Befragungen der Community beinhalten können (Bowman et al., 2022; Sommer et al., 2025). Die Kommunikationswissenschaft erweist sich als typische, stark empirisch orientierte Sozialwissenschaft mit einer dynamischen Entwicklung ihrer Forschung. Zugleich zeigen die Befunde, dass sich das Fach nicht als Ableger der Soziologie oder anderer angrenzender Wissenschaften betrachten lässt, sondern eigene Forschungsgegenstände, -methoden und -praktiken hat, die in einer genuin kommunikationswissenschaftlichen Weise mit Verdatung und Datafizierung in Wechselwirkung stehen und eigener Forschungsdateninfrastrukturen bedürfen, um die Datafizierung der öffentlichen und medialen Kommunikation angemessen empirisch abbilden zu können.

Literaturverzeichnis

agma (2025a). *Das Allmedia-Dach der agma*. https://www.agma-mmc.de/medien-analyse [28.01.2026].

agma (2025b). *Aufgaben und Ziele der agma*. https://www.agma-mmc.de/ueber/agma/aufgaben-ziele [28.01.2026].

Averbeck-Lietz, S. & Löblich, M. (2017). Kommunikationswissenschaft vergleichend und transnational. Eine Einführung. In S. Averbeck-Lietz (Hrsg.), *Kommunikationswissenschaft im internationalen Vergleich. Transnationale Perspektiven* (S. 1–29). Springer.

Bastos, Marco T. (2025). *So Long Twitter, and Thanks for all the Tweets* (March 05, 2025). http://dx.doi.org/10.2139/ssrn.5206365

Bönisch, J. (2006). *Meinungsführer oder Populärmedium? Das journalistische Profil von Spiegel Online.* Lit-Verlag.

Bosnjak, M., Dannwolf, T., Enderle, T., Schaurer, I., Struminskaya, B., Tanner, A., & Weyandt, K. W. (2017). Establishing an Open Probability-Based Mixed-Mode Panel of the General Population in Germany: The GESIS Panel. *Social Science Computer Review, 36*(1), 103-115. https://doi.org/10.1177/0894439317697949

Bowman, N. D., Rinke, E. M., Lee, E. J., Nabi, R., & de Vreese, C. H. (2022). How Communication Scholars See Open Scholarship. *Annals of the International Communication Association, 46*(3), 205–230. https://doi.org/10.1080/23808985.2022.2108880

Breiter, A., & Hepp, A. (2018). Die Komplexität der Datafizierung: Zur Herausforderung, digitale Spuren in ihrem Kontext zu analysieren. In C. Katzenbach, C. Pentzold, S. Kannengießer, M. Adolf, & M. Taddicken (Hrsg.), *Neue Komplexitäten für Kommunikationsforschung und Medienanalyse: Analytische Zugänge und empirische Studien* (S. 27–48). Berlin. https://doi.org/10.17174/dcr.v4.2

Brentel, I., & Jandura, O. (2018). *Media-Analyse: Radio. Langfristdaten* [Datensatz]. https://doi.org/10.7802/1620

Brosius, H.-B. (1994). Integrations- oder Einheitsfach? Die Publikationsaktivitäten von Autoren der Zeitschriften „Publizistik" und „Rundfunk und Fernsehen" 1983–1992. *Publizistik, 39*, 73–90.

Brosius, H.-B., & Haas, A. (2009). Auf dem Weg zur Normalwissenschaft. *Publizistik, 54*, 168–190.

Bruns, A. (2019). After the 'APIcalypse': Social Media Platforms and their Fight against Critical Scholarly Research. *Information, Communication & Society, 22*(11), 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447

Couldry, N., & Hepp, A. (2017). *The Mediated Construction of Reality*. John Wiley & Sons.

Debatin, B. & Rath, M. (Hrsg.) (2014). Die Blindheit der Gutachterei. Debatte. *Aviso, 59*, 2–9.

Dienlin, T., Johannes, N., Bowman, N. D., Masur, P. K., Engesser, S., Kümpel, A. S., Lukito, J., Bier, L. M., Zhang, R., Johnson, B. K., Huskey, R., Schneider, F. M., Breuer, J., Parry, D. A., Vermeulen, I., Fisher, J. T., Banks, J., Weber, R., Ellis, D. A., Smits, T., Ivory, J. D., Trepte, S., McEwan, B., Rinke, E. M., Neubaum, G., Winter, S., Carpenter, C. J., Krämer, N., Utz, S., Unkel, J., Wang, X., Davidson, B. I., Kim, N., Won, A. S., Domahidi, E., Lewis, N. A., & de Vreese, C. (2021). An Agenda for Open Science in Communication. *Journal of Communication, 71*, 1–26. https://doi.org/10.1093/joc/jqz052

Domahidi, E., Yang, JH, Niemann-Lenz, J. & Reinecke, L. (2019). Outlining the Way Ahead in Computational Communication Science: An Introduction to the IJoC Special Section on "Computational Methods for Communication Science: Toward a Strategic Roadmap". *International Journal of Communication, 13*, 3876–3884.

Donsbach, W., Laub, T., Haas, A. & Brosius, H.-B. (2005). Anpassungsprozesse in der Kommunikationswissenschaft. Themen und Herkunft der Forschung in den Fachzeitschriften „Publizistik" und „Medien & Kommunikationswissenschaft". *Medien & Kommunikationswissenschaft, 53*, 46–72.

Döveling, K., Sommer, D., Robinson, K., & Suter, L. (2024). Mediatization and Entertainment. In N. D. Bowman (Ed.), *Entertainment Media and Communication* (S. 505–520). DeGruyter | Mouton.

Filipović, A. (2015). Die Datafizierung der Welt. Eine ethische Vermessung des digitalen Wandels. *Communicatio Socialis, 48*, 6–15.

gesis.org (2025). *Datenerhebung mit dem GESIS Panel.dbd.* https://www.gesis.org/gesis-panel/datenerhebung/datenerhebung-gesis-paneldbd [08.01.2026].

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C. & Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik, 239*(2), 345–360. https://doi.org/10.1515/jbnst-2018-0022

Haas, A., Karnowski, V., & Peter, C. (2018). Hans-Bernd Brosius und die deutschsprachige Kommunikationswissenschaft. In P. Rössler & C. Rossmann (Hrsg.), *Kumulierte Evidenzen* (S. 225–244). Springer.

Hagenah, J. (2005). Der Media-Analyse-Datenservice des Medienwissenschaftlichen Lehr- und Forschungszentrums (MLFZ), Universität zu Köln. *Publizistik, 50*, 116.

Hagenah, J. (2006a). Möglichkeiten der Nutzung von Media-Analyse Radiodaten für Sekundäranalysen von 1972 bis heute. *Medien & Kommunikationswissenschaft, 54*, 457–485.

Hagenah, J. (2006b). Möglichkeiten der Nutzung von Media-Analyse Fernsehdaten für Sekundäranalysen von 1972 bis heute. *Medien & Kommunikationswissenschaft, 54*, 637–653.

Hagenah, J., & Akinci, H. (2003). Sekundäranalysen der Daten der Media-Analyse mit Hilfe des Serviceangebots des Medienwissenschaftlichen Lehr- und Forschungszentrums (MLFZ). *ZA-Information 53*, 182–190.

Hagenah, J., & Meulemann, H. (2009). The Analytical Potentials of Trend Survey Data from Market Research. The Case of German Media Analysis Data. *Historical Social Research, 34*, 49–61.

Hagenah, J., & Meulemann, H. (Hrsg.) (2006). *Sozialer Wandel und Mediennutzung in der Bundesrepublik Deutschland.* LIT Verlag.

Hagenah, J., & Meulemann, H. (Hrsg.) (2008). *Alte und neue Medien – Zum Wandel der Medienpublika in Deutschland seit den 1950er Jahren.* LIT Verlag.

Hagenah, J., & Meulemann, H. (Hrsg.) (2012). *Mediatisierung der Gesellschaft?* LIT Verlag.

Hagenah, J., Meulemann, H., & Akinci, H. (2006). German Media-Analyse (MA): A Large-Scale Commercial Data Source Available for Secondary Analyses on Media Use and Social Change. *Schmollers Jahrbuch, 126*, 129–137.

Hagenah, J., Niemetz, S. & Sommer, D. (2023). *Initiative zur Gründung eines Forschungsdatenzentrums Medien* [Posterpräsentation] „Stunde der Forschung" zum 30. Jubiläum der Fakultät Verkehr – Sport – Tourismus – Medien, Ostfalia Hochschule für angewandte Wissenschaften, 06.05.2023, Salzgitter.

Haim, M. (2023). *Computational Communication Science. Eine Einführung.* Springer.

Haim, M., & Jungblut M. (2023). How Open is Communication Science? Open-Science Principles in the Field. *Annals of the International Communication Association. 47*(3), 338–357. https://doi.org/10.1080/23808985.2023.2201601

Hartmann, F. G., Kopp, J., & Lois, D. (2022). *Sozialwissenschaftliche Datenanalyse.* Springer.

Hauser, R., Wagner, G. G. & Zimmermann, K. F. (1998). Memorandum: Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung. *IZA Discussion Paper No. 14. In Allgemeines Statistisches Archiv, 82*, 369–379. https://docs.iza.org/dp14.pdf [08.01.2026].

Jandura, O., & Brentel, I. (2018). *Media-Analyse: Pressemedien – Langfristdaten* [Datensatz]. https://doi.org/10.7802/1631

Kämper E., & Brüggemann, A. (2009). *Large Scale Studies in den deutschen Sozialwissenschaften: Stand und Perspektiven. Bericht über einen Workshop der Deutschen Forschungsgemeinschaft* (RatSWD Working Paper Series, 114). Berlin.

Koenen, E. & Sanko, C. (2017). Die Mediengesellschaft und ihre Wissenschaft im Wandel. Disziplinäre Genese und Wandelprozesse der Kommunikationswissenschaft in Deutschland 1945–2015. In S. Averbeck-Lietz (Hrsg.), *Kommunikationswissenschaft im internationalen Vergleich. Transnationale Perspektiven* (S. 113–159). Springer.

konsortswd.de (2025a). *Entwicklung des RatSWD. Die Anfänge des RatSWD reichen in das Jahr 1998 zurück.* https://www.konsortswd.de/ueber-uns/ratswd/entwicklung/ [08.01.2026].

konsortswd.de (2025b). *Übersicht Datenzentren. Die vom RatSWD akkreditierten Forschungsdatenzentren (FDZ) ermöglichen der Wissenschaft einen flexiblen und umfangreichen Zugang zu sensiblen Daten.* https://www.konsortswd.de/angebote/forschende/alle-datenzentren/ [08.01.2026].

Krotz, F. (2017). Mediatisierung: Ein Forschungskonzept. In F. Krotz, C. Despotović, & M. M. Kruse (Hrsg.), *Mediatisierung als Metaprozess. Transformationen, Formen der Entwicklung und die Generierung von Neuem* (S. 13–32). Springer.

Krüger, K. (2021). *Die Ära der Datafizierung. Medieninnovation als Wandel der Medientypologien.* Springer VS. doi.org/10.1007/978-3-658-34019-3

Lauf, E. (2002). Freiheit für die Daten! Sekundäranalysen und Datenbestände in der deutschen Medien- und Kommunikationswissenschaft. *Medien & Kommunikationswissenschaft, 50*, 247–260, https://doi.org/10.5771/1615-634x-2002-2-247

Löblich, M. (2010). Die empirisch-sozialwissenschaftliche Wende: Ein Beitrag zur historischen und kognitiven Identität der Kommunikationswissenschaft. *Medien & Kommunikationswissenschaft, 58*, 544–562. https://doi.org/10.5771/1615-634x-2010-4-544

Matthes, J., Kuhlmann, C., Gehrau, V., Jandura, O., Möhring, W., Vogelgesang, J., & Wünsch, C. (2011). Zur Methodenausbildung in kommunikationswissenschaftlichen Bachelor- und Masterstudiengängen. Empfehlungen einer Kommission im Auftrag der Fachgruppe Methoden der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft. *Publizistik, 56*, 461–481. https://doi.org/10.1007/s11616-011-0133-6

Matthes, J., Kuhlmann, C., Gehrau, V., Jandura, O., Möhring, W., Vogelgesang, J. & Wünsch, C. (2012). Schwerpunkte ermöglichen – eine Antwort auf die Stellungnahme von Krotz, Keppler, Meyen, Neumann-Braun und Wagner. *Publizistik, 57*, 225–227. https://doi.org/10.1007/s11616-012-0149-6

Mauer, R. (2012). Das GESIS Datenarchiv für Sozialwissenschaften. In R. Altenhöner, & C. Oellers (Hrsg.), *Langzeitarchivierung von Forschungsdaten: Standards und disziplinspezifische Lösungen* (S. 197-215). Scivero. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-46476-7

Meulemann, H., & Hagenah, J. (2010). Mass Media Research. In RatSWD (Hrsg.), *Building on Progress: Expanding the Research Infrastructure for the Social, Economic and Behavioral Sciences, Vol. 2* (S. 1173–1196). Budrich.

Mochmann, E. (2022). Quantitative Daten für die Sekundäranalyse. In N. Baur, & J. Blasius (Hrsg), *Handbuch Methoden der empirischen Sozialforschung* (S. 283–294). Springer.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2017). *Reuters Institute Digital News Report 2017.* Reuters Institute for the Study of Journalism.

Pohle, J. (2019). Zu den gesellschaftlichen Auswirkungen zunehmender Verdatung und Automation: Erkenntnisse aus der Frühzeit und Hochphase der Datenschutzdebatte. In N. Burzan (Hrsg.), *Komplexe Dynamiken globaler und lokaler Entwicklungen: Verhandlungen des 39. Kongresses der Deutschen Gesellschaft für Soziologie in Göttingen 2018. Deutsche Gesellschaft für Soziologie (DGS).* https://doi.org/10.21241/ssoar.101126

Prinzing, M., & Herczeg, P. (Hrsg.) (2023). Publikationsstrategien in der KoWi: Wie publizieren? Debattenschwerpunkt. *Aviso, 77*, 3–13.

Quandt, T. (2008). (No) News on the World Wide Web? A comparative content analysis of online news in Europe and the United States. *Journalism Studies, 9*(5), 717–738. doi.org/10.1080/14616700802207664

Riffe, D., & Freitag, A. (1997). A Content Analysis of Content Analyses: Twenty-five Years of Journalism Quarterly. *Journalism and Mass Communication Quarterly, 74*, 515–542. https://doi.org/10.1177/107769909707400306

Schäfer, M.-S., & Rössler, P. (Hrsg.) (2014). Open Access in der Kommunikationswissenschaft. Debatte. *Aviso, 58*, 2–11.

Scharkow, M. (2011). Automatische Inhaltsanalyse und maschinelles Lernen. Dissertation: Universität der Künste. https://opus4.kobv.de/opus4-udk/frontdoor/deliver/index/docId/28/file/dissertation_scharkow_final_udk.pdf [08.01.2026].

Scharkow, M. (2013). Thematic Content Analysis Using Supervised Machine Learning: An Empirical Evaluation Using German Online News. *Quality & Quantity, 47*, 761–773.

Schneider, I., & Otto, I. (Hrsg.) (2007). *Strategien der Verdatung. Formationen der Mediennutzung II.* transcript.

Simon, N., & Bernhardt, N. (2008). *Twitter, mit 140 Zeichen zum Web 2.0.* Open Source Press.

Sommer, D., Hagenah, J., Strippel, C., & Brucks, A.-S. (2025). *Frei und FAIR? Der Umgang mit Forschungsdaten in der deutschsprachigen Kommunikations- und Medienwissenschaft.* [Posterpräsentation] 70. DGPuK-Jahrestagung, 20.03.25, Berlin.

Strippel, C. (2021). Forschungsinfrastrukturen für die Kommunikations- und Medienforschung im deutschsprachigen Raum: Initiativen, Bedarfe und Perspektiven. *Medien & Kommunikationswissenschaft, 69*, 136–157. https://doi.org/10.5771/1615-634X-2021-1-136

Strippel, C., & Domahidi, E. (Hrsg.) (2016). Internationalisierung in der Kommunikationswissenschaft. Debatte. *Aviso, 63*, 2–11.

Strippel, C., & Paasch-Colberg, S. (2020). Diskursarchitekturen deutscher Nachrichtenseiten. In V. Gehrau, A. Waldherr, & A. Scholl (Hrsg.), *Integration durch Kommunikation (in einer digitalen Gesellschaft): Jahrbuch der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft 2019* (S. 153–165). Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft e. V. https://doi.org/10.21241/ssoar.68129

Strippel, C., Bock, A., Katzenbach, C., Mahrt, M., Merten, L., Nuernbergk, C., Pentzold, C., Puschmann, C., & Waldherr, A. (2018). Die Zukunft der Kommunikationswissenschaft ist schon da, sie ist nur ungleich verteilt. *Publizistik, 63*, 11–27.

Trilling, D. (2014). Weg vom manuellen Speichern: RSS-Feeds in der automatisierten Datenerhebung bei Online-Medien. In K. Sommer, M. Wettstein, W. Wirth, & J. Matthes (Hrsg.), *Automatisierung in der Inhaltsanalyse* (S. 73–89). Von Halem.

Volk, S., Weinmann, C., & Vowe, G. (Hrsg.) (2022). Forschung im Team: Was bedeutet die „kollaborative Wende" für die KW? Debatte. *Aviso, 74*, 3–10.

von Oertzen, C. (2017). Die Historizität der Verdatung: Konzepte, Werkzeuge und Praktiken im 19. Jahrhundert. NTM – Zeitschrift für Geschichte der Wissenschaften, *Technik und Medizin, 25*, 407–434. https://doi.org/10.1007/s00048-017-0183-6

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch, 127*(1), 139–169.

Waldherr, A. (2019). Messinstrumente und Sinnkonstruktionen. Methoden als Antreiber und Taktgeber der Kommunikationswissenschaft. *medien & zeit, 34*, 40–47.

Wissenschaftsrat (2021). *Stellungnahme zum Antrag auf strategische Erweiterung von GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim und Köln, großer strategischer Sondertatbestand im Rahmen der Ausführungsvereinbarung WGL.* https://www.wissenschaftsrat.de/download/2021/8820-21.pdf?__blob=publicationFile&v=8#page=13 [08.01.2026].

Anhang

Tabelle 1: Übersicht der von Haas et al. (2018) übernommenen und für die Untersuchung ergänzten Untersuchungsvariablen

Variable von Haas et al. (2018)	**Variable von Sommer et al. (2024)**	**Intercoderreliabilität nach HOLSTI**
Untersuchtes Medium	Art der Zeitschrift	1
	Titel des Aufsatzes	1
Jahr der Veröffentlichung	Jahr der Veröffentlichung	1
	Jahrgang	1
Quartal der Veröffentlichung	Heftnummer	1
Anfangsseite Endseite	Seitenanzahlen	1
Name des/der Autor:in	Name des/der Autor:in	1
Geschlecht der Erstautor:in	Erstautor:in weiblich	1
Anzahl der Autor:innen	Anzahl der Autor:innen	–
	Anzahl der weiblichen Autorinnen	–
	Anzahl der männlichen Autoren	–
Stellung der Autor:innen	Stellung der/des Erstautor:in	0,75
	Anzahl an veröffentlichten Publikationen (bis zum Zeitpunkt des Aufsatzes)	–
	Anzahl insgesamter veröffentlichter Publikationen	–
Institution	Ort der Institution	1
Universität/Hochschule	Fachzugehörigkeit der Institution	1
Thema	Thema: Fachentwicklung	–
	Thema: Mediennutzung	0,5

Variable von Haas et al. (2018)	Variable von Sommer et al. (2024)	Intercoderreliabilität nach HOLSTI
	Thema: Medienwirkung	0,5
	Thema: Methoden	1
	Thema: Mediengeschichte	1
	Thema: Medienpolitik	1
	Thema: Medienökonomie	1
	Thema: Journalismus-/ Kommunikator:innenforschung	1
	Thema: Gender	1
Empirisches Vorgehen	Empirisches Vorgehen	1
	Methode: Befragung	1
	Methode: Gruppendiskussion	1
Methode 1 Methode 2	Methode: Inhaltsanalyse	1
	Methode: Beobachtung	1
	Methode: Experiment	1
	Methode: Computational Methods	1
Art der Daten	Quantitative Forschung	1
	Qualitative Forschung	1
	Kombination aus quantitativer und qualitativer Forschung	1
Anzahl an Quellen insgesamt	Anzahl an Quellen	–
Anzahl Zeitschriftenartikel Englisch Anzahl Monografien Englisch Anzahl Sammelbänder Englisch	Anzahl englischsprachiger Quellen	–
Umfang	Umfang an Seiten	1
	Medium: Presse	1
	Medium: Fernsehen	1
	Medium: Radio	1
	Medium: Audio	1
	Medium: Video	1
	Medium: mediales Internet	0,75
	Medium: nicht-mediales Internet	1
	Medium: Kein Medium	1
	Medium: Allgemein Medien	1

Beyond the Black Box: A Multimodal Approach to Understanding In-App Communication

Patrick Zerrer / Paul Pressmann / Cornelius Puschmann / Philipp Krieter*

Mobile devices are increasingly central as sources of up-to-date information, making the precise recording of information behavior on these devices more relevant for research. Established methods of automated data collection are reaching their limits when capturing in-app communication, such as political content within social media applications. Based on two case studies, we present two approaches that enable the identification of exposure to relevant content and, to some extent, the collection of in-app content. The first case study focuses on identifying relevant exposure across different apps on a mobile device using app tracking and screen recordings. The second case study focuses on linking exposure to seen content and deriving respective content features to obtain an enriched dataset. We discuss the advantages and limitations of both approaches and present conceptual frameworks for processing and analyzing such data.

Key words: mobile tracking, screen recording, mobile sensing, android logs, in-app tracking

1. Introduction

Technological and media changes are contributing to the ongoing fragmentation of media diets (Hasebrink & Popp, 2006). This includes the emergence of new media formats, the growing relevance of social media platforms and messaging services, and the widespread adoption of smartphones (Behre et al., 2024). These developments coincide with the increasing datafication of everyday media and communication behaviors (Breiter & Hepp, 2018). Against this backdrop, automated observation of smartphone use through mobile tracking offers a promising approach for drawing insights into media diets and media-mediated interpersonal communication (Ohme et al., 2024).

Automated mobile tracking offers several advantages as a data collection method. The increasing fragmentation of media and information use makes accurate measurement through self-reporting challenging (Araujo et al., 2017; Parry et al., 2021; Scharkow, 2016). While some researchers have raised justified criticisms of automated tracking (Bosch et al., 2024), others highlight its benefits, noting that it allows for more precise measurement (Stier, Breuer, et al., 2020) and the exploration of duration, frequency, sequence, and content engagement (Adam et al., 2024; Christner et al., 2022). In addition, automated tracking

* Dr. Patrick Zerrer, University of Bremen, ZeMKI Center for Media, Communication and Information Research, Linzer Str. 4,28359 Bremen, Germany, pzerrer@uni-bremen.de, https://orcid.org/0000-0002-8827-1336;
Paul Pressmann, M.A., University of Bremen, ZeMKI Center for Media, Communication and Information Research, Linzer Str. 4, 28359 Bremen, Germany, pressmann@uni-bremen.de, https://orcid.org/0009-0008-1271-4429;
Prof. Dr. Cornelius Puschmann, University of Bremen, ZeMKI Center for Media, Communication and Information Research, Linzer Str. 4, 28359 Bremen, Germany, puschmann@uni-bremen.de, https://orcid.org/0000-0002-3189-0662;
Philipp Krieter, Düsseldorf University of Applied Sciences, Faculty of Media, Münsterstraße 165, 40476 Düsseldorf, Germany, philipp.krieter@hs-duesseldorf.de.

DOI: 10.5771/1615-634X-2026-1-31

provides a comprehensive view of user behavior across the entire smartphone, rather than being limited to individual platforms or apps. However, most current mobile tracking methods cannot capture in-app user behavior, such as watching a short news video on Instagram. Given that a substantial portion of relevant smartphone use occurs within apps, recording these usage episodes represents a central methodological challenge (Muise et al., 2024).

In this paper, we address two key research questions regarding the recording of user behavior within apps. First, how can research-relevant exposure within apps be identified (RQ1)? Second, how can the in-app content viewed by users be recorded (RQ2)? Answering RQ1 requires measuring general smartphone usage behavior and detecting events that are instrumental for addressing the research objectives. Addressing RQ2 involves a more detailed examination of the received and relevant audio-visual content by linking smartphone usage data with extracted content through a combination of scraping and automated content analysis. Additionally, we develop research principles for working with mobile tracking and screen recording data, drawing on both the literature and our case studies. We argue that this approach offers advantages over contemporary data donation strategies, which often place an undue burden on participants and are prone to coverage gaps and omissions.

2. Current state and challenges of automated tracking

The increasing differentiation and digitalization of media consumption have contributed to a growing use of automated tracking methods in communication research (Guess et al., 2020; Jürgens & Stark, 2022; Maier et al., 2025; Stier et al., 2021; Wojcieszak et al., 2023). This trend is largely driven by the recognition that traditional data collection methods, such as surveys and media diaries, often struggle to capture media usage with precision (e.g. Parry et al., 2021).

Many studies in communication science and related disciplines employ mobile or desktop tracking for data collection, each method offering distinct advantages and drawbacks (Adam et al., 2024; Stier et al., 2021; Stier, Kirkizh et al., 2020). All tracking approaches share the primary goal of recording exposure—that is, capturing a 'person's contact with relevant content along with the time at which it occurs. Some methods further combine exposure data with additional sources, using unique idenitifiers to ensure correct allocation. For instance, web content can be collected and linked to an individual participant's exposure. Studies employing different automated tracking methods include proxy-based tracking, browser-based (mobile) tracking, and app tracking (Clemm von Hohenberg et al., 2024).

One method for collecting online exposure data is the proxy approach. In this method, the proxy acts as a digital intermediary that records outgoing and incoming internet connections, providing information about accessed URLs and apps used (Christner et al., 2022). Proxy server-based tools intercept 'participants' requests and store a complete record of the content accessed (Menchen-Trevino & and Karr, 2012). In this way, the proxy functions as an invisible intermediary connecting the 'user's device to the internet, routing all traffic through itself.

Another approach for collecting 'users' exposure is through browser extensions, which record visited URLs—either in full or reduced to the domain—along with the time and often the duration of each visit (Stier, Kirkizh et al., 2020; Wojcieszak et al., 2023). This enables assessments of the number, frequency, variety, and time spent on specific websites, such as news outlets (Wojcieszak et al., 2023), politicians' social media profiles (Stier et al., 2018), or other content types, such as pornography (von Andrian-Werburg et al., 2023). Typically, this approach requires a predefined list of URLs relevant to the research focus

(e.g., news sites), which is then compared to the tracked URLs to extract relevant data. Alternatively, tracked URLs can be coded manually or automatically to identify content relevant to the study. In either case, comparing tracked URLs poses practical challenges, particularly due to the so-called long-tail issue, where a large number of URLs are visited only infrequently.

If URLs are fully tracked, this exposure data can be enriched by scraping the content associated with the recorded URLs and using the URL as a unique identifier to link content to individual users (Clemm von Hohenberg et al., 2024; Kühnemann, 2021; Munzert & Nyhuis, 2019). This means that, for both desktop and mobile browsing, it is already possible to identify exposure and assign the corresponding content to the respective user using the URL as a unique identifier. However, this approach faces limitations in today's increasingly mobile-centric media landscape, as in-app content (e.g., Instagram posts) cannot be easily tracked, restricting data collection to browser-accessed sites. Accordingly, research encounters two main challenges when capturing in-app content: first, correctly identifying relevant exposure, and second, collecting the content that has been accessed and linking it to the recipient's exposure. Assigning content to in-app exposure is particularly because there is no generalizable identifier, such as a URL, available within apps.

The collection of mobile automated tracking data faces several challenges. Smartphones, equipped with numerous sensors, have become an integral part of daily life (Görland, 2020) and collect a wide variety of information through these sensors (Ferreira et al., 2015). Several commercial and open-source applications allow researchers to access this data for research purposes (Christner et al., 2022). These mobile tracking apps record which applications are used on a smartphone and share this information with researchers (Ferreira et al., 2015; Tong et al., 2022; Toth & Trifonova, 2021; Parry & Toth, 2025). This enables the collection of data on app-specific usage duration, such as time spent on Instagram, as well as general smartphone use, including overall screen time (Fan et al., 2021; Tong et al., 2020). Unlike URL-based tracking, this method captures app usage in a mobile context. However, the content within those apps remains inaccessible due to standard iOS and Android security features, making it difficult to infer either exposure to relevant content or the nature of the content viewed based solely on app-tracking data.

Illuminating this methodological "black box" is key advantage of screen capturing and screen recording (Krieter, 2019, 2020; Krieter et al., 2024). In this approach, a tracking application is installed on the device, typically a smartphone, that captures screenshots or records screen videos at regular intervals (Krieter et al., 2024; Reeves et al., 2021; Yee et al., 2023). Screen recording techniques enable automated analysis by generating log files derived from these recordings (Frisson et al., 2016; Krieter, 2020; Yee et al., 2023). This allows researchers to capture of not only general system events, such as opening an app like Instagram (Böhmer et al., 2011; McMillan et al., 2015), but also specific in-app activities, such as viewing or sharing content. These analyses rely on artificial intelligence, computer vision, and machine learning methods to analyze mobile screen recordings (Krieter, 2020; Reeves et al., 2021; Yee et al., 2023).

Depending on the technical setup, screenshots or videos may be transmitted directly to a server operated by the research team, which provides rich data but raises substantial privacy concerns. Alternatively, screen recordings can be 'processed locally on the participant's device, with only derived results uploaded to the 'server. While this latter approach yields less detailed information, it substantially enhances privacy protection (Krieter, 2020; Krieter et al., 2024). Accordingly, screen recording techniques can help identify relevant exposure within apps, although capturing and storing the viewed content itself remains challenging due to privacy-preserving restrictions'.

To summarize the current state of mobile tracking technologies, existing methods primarily enable the recording of web browsing on smartphones by logging visited URLs. When URLs are fully captured, exposure data can be enriched by scraping the corresponding webpages and matching their content to user exposure, using the URL as a unique identifier. However, this procedure is not feasible for in-app environments, such as social media or messaging applications. Consequently, significant methodological gaps remain, both in the identification of relevant in-app content and in the assignment of that content to recorded exposure. This paper addresses these gaps by posing two research questions: the identification of relevant in-app exposure (RQ1) and the linkage of in-app content to exposure (RQ2). Despite the limitations discussed above, automated tracking enables the unobtrusive capture of user behavior and, in some cases, content across multiple platforms, making it particularly well suited to studying the complex contemporary media landscape.

3. Towards a holistic approach to mobile tracking

Most people carry their smartphones with them throughout their everyday lives, and these devices continuously generate data related to daily events, either directly or indirectly. Mobile tracking data exhibit several distinctive characteristics, including high granularity, sequentiality, large volume, and the interweaving of device and software affordances with human actions. These features underscore the need for a comprehensive approach to mobile tracking in research design that extends beyond the technical implementation of data collection. Accordingly, this paper addresses issues of data protection and privacy, socio-technical considerations, and extensions to the operationalization of exposure.

3.1 Data protection and privacy

Due to the fine-grained and sequential nature of data collection on mobile app usage and screen-recorded content, tracking data may contain a wide range of highly sensitive and private information (Krieter, 2020). Such data may include information about health, dating, or finance applications, which can allow inferences about characteristics such gender, sexual orientation, or socioeconomic status. Accordingly, these data require particularly careful, responsible, and conscientious handling by researchers.

In a tracking study, it is essential to inform participants comprehensively and transparently about the functionality of the tracking system, the scope of the data collected, and how these data will be used, and to obtain their informed consent within the framework of the study. Providing a clear and comprehensible explanantion of how app tracking and screen recording operate, as well as how participants' data are handled and stored, is an important way to ensure that participants are adequately informed.

All raw data should be pseudonymized as early as possible in the research process. Any identifiers that could directly or indirectly reveal the identity of participants must be removed or replaced with randomly generated codes. Where feasible, additional anonymization procedures should be applied to ensure that participants cannot be re-identified. Particular attention must be paid to the risk of de-pseudonymization through the combination of study data with third-party information.

Studies have shown that the central storage of data with third parties (e.g. commercial providers of cloud services) conflicts with users' privacy requirements (Hong et al., 2003; Spiekermann & Cranor, 2009). Therefore, it appears more advantageous to use a secure storage location on servers operated by 'one's own institution. Access to stored data must be restricted to authorized members and governed by strict authentication procedures.

3.2 Socio-technical considerations

Mobile automated tracking data constitute a type of digital trace that is often regarded as a more precise—and obstensibly superior—representation of social reality. However, such data primarily provide a partial and surface-level view of human interaction with smartphones. Relying exclusively on these data to reconstruct complex social and societal realities risks oversimplification and misrepresentation of the social world (Breiter & Hepp, 2018; Jungherr & Theocharis, 2017). There is a tendency to privilege what can be quantified through tracking over the inherent complexity of human behavior, with these observations being mistakenly treated as comprehensive and accurate (Jungherr, 2015). To mitigate this risk, a robust strategy that applies the concept of measurement error to digital traces, analogous to other data generating processes in the social sciences is thus called for (Sen et al., 2021).

At the same time, the assumption that mobile automated tracking data possess comprehensive validity often coincides with the belief in their neutrality. Yet, like all digital trace data, they are not neutral; they are produced through technical procedures implemented by powerful institutions (such as corporations and governments) with specific interests (Breiter & Hepp, 2018; Gitelman, 2013). Consequently, analyses of mobile automated tracking data engage with information that is institutionally controlled, generated for particular purposes, and inherently biased. Against this backdrop, it is essential to critically assess digital traces as indicators of human behavior and social reality (Breiter & Hepp, 2018).

3.3 Operationalization of media exposure

Although tracking data almost always include a timestamp for each data point, enabling conclusions about duration and sequence of use, the majority of studies in automated tracking research rely on a relatively superficial operationalization of media exposure. Often, exposure is aggregated across individual media outlets or broad types (e.g. legacy media). Automated tracking data, however, allow media exposure to be operationalized along multiple dimensions, including duration, frequency, order, and the content of visited websites or applications (Christner et al., 2022). Despite this potential, most studies focus primarily on frequency and duration, while neglecting order, sequence, and specific content. As a result, they overlook the complexity and sequential nature of actual media use behavior.

Therefore, usage behavior before, during, and after exposure is captured only inadequately, if at all. Yet, qualitative studies demonstrate that individuals exhibit a wide range of behaviors when reading, watching, or listening to content, which in turn affects reception (Merten, 2021; Wieland, 2023). For many of these behavioral patterns, such as scanning, which refers to the quick and superficial viewing of content (Meijer & Kormelink, 2015), both the temporal sequence and the duration of engagement are crucial for characterizing the behavior.

Incorporating temporal information into the measurement of exposure can provide a more accurate reflection of reality in the data and enhance the explanatory power of exposure in statistical models, as demonstrated by Richter & Stier (2022). Several notable studies have operationalized and analyzed sequential tracking data using information-theoretical approaches (Kulshrestha et al., 2021; Möller et al., 2020). The works of Möller et al. (2020), Jürgens and Stark (2022), and Kulshrestha et al. (2021) showcase the level of depth and detail that can be achieved with automated tracking data when it is properly prepared and operationalized. Nevertheless, this level of granularity is underutilized in most studies, which often resort to simple aggregation of usage times. While aggregation may suffice for

some research questions, it risks overlooking valuable information that has already been collected.

Against this backdrop, mobile tracking data—characterized by its granularity, temporal sequence, and close connection to 'users' everyday experiences—presents specific challenges for research. These challenges include ensuring trustworthy data handling, accurately interpreting individual data points, and and effectively operationalizing the information collected. All of these factors must be carefully considered when designing a study using mobile tracking.

The characteristics of mobile tracking data and the implications for its handling are further illustrated in the first case study.

4. Case study I - Identification of relevant exposure

The first case study addresses the first research question (RQ1), focusing on the identification of research-relevant exposure on mobile devices. It examines how politically active young people engage with political and climate-related information on their smartphones. Within this context, we developed and conceptually assessed the use of screen recordings as a method for measuring media use.

4.1 Data collection and methods

To obtain a detailed picture of information behavior, a total of 25 participants were recruited. Their smartphone use was recorded, with consent, between 14 February 2022 and 6 May 2022. General smartphone usage and the detection of relevant events were measured using a combination of two mobile tracking applications: the app tracking application AWARE and the Keywordlogger developed by Krieter (2020; 2024). This approach resulted in a comprehensive dataset comprising over 9 million data points.

4.2 Mobile automated app tracking

In this case study, we used the AWARE framework as an open-source solution for collecting hardware-, software-, and human-based data from smartphones (Ferreira et al., 2015). The AWARE client was implemented to monitor app usage and overall smartphone activity. Each time a user switched to a different application, a new timestamp was recorded. Additionally, the screen sensor captured the screen's status, such as on/off and lock/unlock events. This setup enabled the collection of information on both the frequency of interactions with the smartphone and their duration.

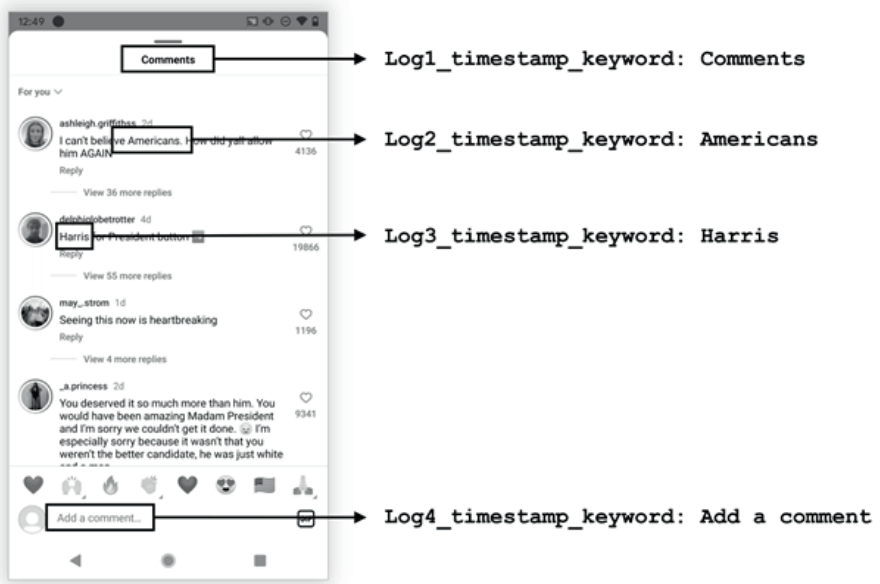
4.3 Screen recording

In addition to monitoring app usage, identifying relevant content within apps presents a central challenge. To address this, we applied screen recording methods adapted for communication science research as part of this case study.

Depending on the research focus, there are two main tracking approaches: (1) monitoring keywords related to smartphone or specific in-app functions (e.g., sharing, forwarding, etc.); (2) using content-related keywords that target specific vocabulary or references pertinent to particular topics (for instance, names of political parties, politicians, or terms describing specific media content such as Instagram accounts, newspapers, and TV channels). It is important that the use of keywords does not compromise privacy protections.

The Keywordlogger analyzes the screen recordings based on these previously defined keywords and records each visible keyword, with the timestamp, for each frame of the video examined. In this way, log files are generated which provide information about the occurrence, duration, and sequence of the defined keywords during the smartphone's usage time.

Figure 1: Example of how the Keywordlogger identifies keywords in the Instagram comments section.



The Keywordlogger essentially takes two approaches to strengthen the privacy of participants. First, the screen videos, which can contain very sensitive data, are only stored temporarily and do not leave the device, as they are only evaluated locally on the smartphone. The evaluation of the videos follows a fixed catalog of predefined events that are to be recognized in the videos. This means that only what is relevant for the research purpose is recognized and saved. The log files created locally on the device are transmitted to the research team in a pseudonymized form via a secure connection.

Second, the participants have the option of stopping the tracking throughout the entire study period by revoking the authorizations for the Keywordlogger & AWARE app on their Android smartphones. In this way, control over the data remains with the participants (Ferreira et al., 2015).

4.4 Operationalization of media exposure

The case study combines different data sources to create a comprehensive dataset that reflects the complex reality of user behavior. Enriching an initial dataset with additional information provided by other data collection methods, such as screen recording or other smartphone sensors, involves several steps.

Table 1: Data structure combined

Participant[1]	**Screen- and app-tracking by AWARE**						**Screen recordings by Keylogger**
	Smartphone display on and off	*Mobile session duration in seconds*	*Mobile application*	*Start of the mobile application*	*End of the mobile application*	*Usage duration in seconds*	*Keywords*
1000	08:33:56–08:35:14	78.05	Clock	08:34:03.988	08:34:06.757	2.77	Eis, Spiegel
1000	08:33:56–08:35:14	78.05	WhatsApp	08:34:06.757	08:34:07.436	0.68	[no keywords recorded]
1000	08:33:56–08:35:14	78.05	Mobile phone keypad[2]	08:34:07.436	08:34:15.012	7.58	Eis, Spiegel, Stadt, Erde
1000	08:33:56–08:35:14	78.05	WhatsApp	08:34:15.012	08:34:53.791	38.78	Wald, Erde, Luft, Wetter, Eis, Jahr, Wald, Erde, Abfall, Luft, Wetter, USA
1000	08:33:56–08:35:14	78.05	Mobile phone keypad[2]	08:34:53.791	08:35:04.034	10.24	Erde, Wetter, Handel, Eis, Erde, Luft, Erde, Luft, Wetter, Handel, Eis
1000	08:33:56–08:35:14	78.05	WhatsApp	08:35:04.034	08:35:12.552	8.52	Stadt, Jahr, Stadt
1000	08:33:56–08:35:14	78.05	WhatsApp	08:35:12.552	08:35:14.605	2.05	[no keywords recorded]

[1] Participant 1000 is the test device.
[2] AWARE records the app that is active in the foreground of the smartphone at that moment. The temporarily prior and subsequent app indicates within which app the keypad was used.

(1) *Data collection* using two different app tracking applications provides an initial raw dataset. We used app tracking and screen recording to capture information about app launches, screen status, and timestamped keywords. This information about smartphone use is constituted in several initial data sets, which must be processed and linked in the following.

(2) The *session concept* is a core conceptual component that enables different data sets to be linked together in a meaningful way. The session concept is based on the work of Peng and Zhu (2020), who understand uninterrupted usage behavior as a session. In this context, thinking in temporal order is a relevant part of the concept. In our understanding, the concept can be applied to different levels, resulting in a matryoshka-like data structure that

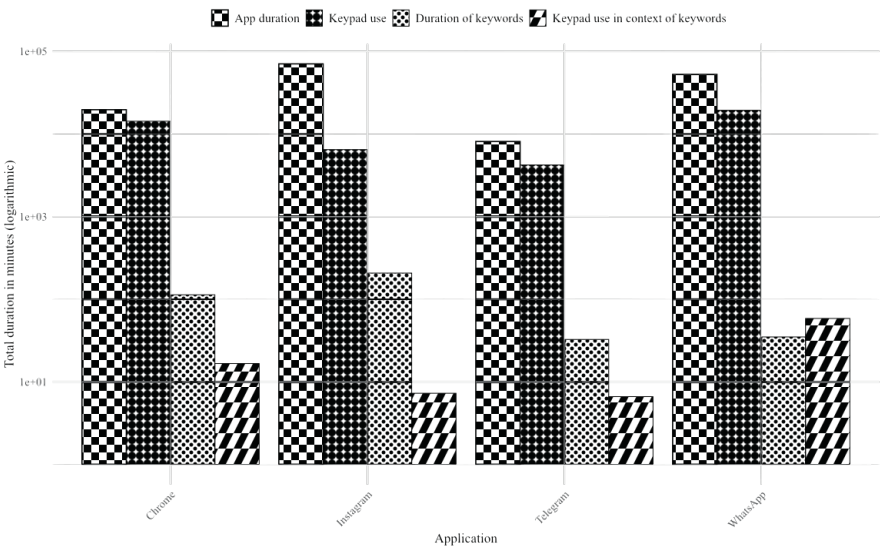
enables flexible handling of multi-level data. In this case study, several session levels were applied, first, *keyword sessions*, which comprise the time span of one or more visible terms; *app sessions*, which represent the uninterrupted usage sequence of an app; and *mobile sessions*, which comprise the aggregated usage behavior between unlocking and locking the smartphone. Among other things, this structure makes it possible to assign terms recorded by the Keywordlogger to the corresponding app session, e.g. Instagram or Whatsapp.

(3) *Contextual enrichment* describes precisely this step of correctly assigning nested session levels. For a better understanding, let us describe the initial situation: the first dataset contains the start and end times of the apps used on the smartphone, while the second dataset is a list of timestamps with the respective recorded keywords. The goal is to correctly assign the terms to the app that matches the time. Therefore, for each keyword it must be checked whether its timestamp falls within the time span of the app usage. The result is a data set that can be aggregated to different session levels without losing contextual information. Furthermore, aggregations at other levels, such as general app or screen usage time, are still possible.

4.5 Results and discussion

The first research question (RQ1) aims to identify research-relevant exposure within apps on smartphones. The combination of app tracking and screen recording enables us, in this case study focusing on the information usage of supporters of climate protests, to make statements about exposure to climate-relevant terms and actors on smartphones and within specific apps.

Figure 2: The visibility of climate-protest terms in apps (N = 25)



Note: An active keypad is labeled as "keypad use" in the context of keywords if at least one keyword is identified during the active and possibly interrupted period of keypad use. This means that the duration of keypad use may exceed the time span in keywords are actually visible.

The dataset created in this way allows for the operationalization of media exposure at different levels (across participants and within participants) and at different points in time (month, week, day, hour, etc.), while capturing multiple dimensions such as binary (non-)exposure, frequency, duration, order, or sequence. For example, the visibility of climate protest-related terms can be analyzed across the entire sample for each app (see Figure 2).

An even more granular analysis consists of calculating the duration of active keypad use with simultaneously visible climate protest terms per app (see Figure 2). Further breakdowns over time or comparisons between individuals can also be conducted.

In this way, the combination of app tracking and screen recording enables the identification of research-relevant constructs, allowing statements about the presence or absence of participants' exposure to specific actors or media-outlets. This approach makes it possible to examine questions regarding the range of information sources within social media apps or messengers, the frequency of exposure, and patterns of use throughout the day to be examined.

5. Case study II – Matching in-app exposure and content

Our second case study demonstrates how the limitations discussed above can be addressed by integrating content-level information with in-app exposure data. Specifically, this case study focuses on answering the second research question (RQ2), which concerns the recording of in-app content. We illustrate this by examining how participants' in-app exposure on TikTok is distributed across various public actors. Additonally, we investigate the share of politically related content on the platform. The case study measures political exposure on TikTok over a one-week period, beginning 1 September 2025.

5.1 Data collection and methods

Using accessibility-based mobile tracking, we recorded the in-app activity of 452 participants', capturing fragments of post descriptions and associated metadata whenever predefined potential political accounts (Zerrer, et al., 2024) or keywords appeared on screen. To reconstruct incomplete textual information, these fragments were matched with a parallel scraping database of TikTok posts from the observed accounts, based on string similarity alignment. The resulting dataset linked participants' exposure events to complete post-level content and corresponding engagement metrics, enabling a distinction between political and non-political content within each actor category. This approach allows for a deeper examination of the content participants received and interacted with, through a combination of automated data extraction, content analysis, and contextual interpretation (Ohme et al., 2024). Moreover, this step proposes a framework for integrating tracking data with context-rich metadata and insights from further content analysis procedures (Freelon et al., 2024; Reeves et al., 2021).

5.2 Capturing In-App Communication: Methodological Pathways

Analyzing in-app communication requires access to specific types of information, most importantly account names and post descriptions. Obtaining such information, however, is not straightforward, as different methodological approaches can vary considerably in coverage, granularity, feasibility, and privacy implications. One option is to adapt the approach outlined in the first case study, where app tracking was combined with keyword logging to approximate exposure to relevant content. This strategy allows for broad coverage across

applications, as it does not require a priori restrictions to specific platforms. However, while this method provides valuable insights into exposure patterns, it remains limited in capturing complete content and its contextual richness. As a result, researchers often gain only partial visibility into what users actually encountered.

A second option is the so-called Screenomics framework (Reeves et al., 2021). Screenomics systematically records a continuous stream of screenshots or screen videos, enabling a highly detailed reconstruction of all visible content, including post descriptions and user interface elements. This method offers a comprehensive perspective on digital communication practices and supports rich contextual analyses. At the same time, it produces massive volumes of data and raises substantial privacy concerns, as continuous recording of all on-screen activities may capture highly sensitive personal information.

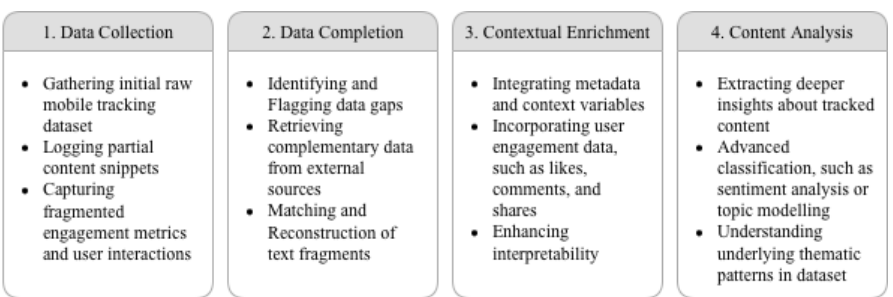
A third approach is to rely on a smartphone's Accessibility Service as a data collection instrument. Originally designed to assist users with disabilities—by enabling enhanced interaction options, such as reading screen content aloud, voice commands or automated actions—the system infrastructure also provides researchers with a powerful tool to study digital behavior on smartphones and obtain a more detailed and continuous record of activity (Andone et al., 2016). By leveraging the Accessibility Service, researchers can track event-based interactions, app usage patterns, displayed content, and user engagement across different platforms, offering more detailed information than screen recordings.

In practice, the Accessibility Service can be configured to capture a wide range of behavioral data directly at the system level. It records which applications are active, the duration and frequency of usage, and interactions within apps, such as clicks and scrolling patterns. Additionally, it can extract text displayed on screen, including advertisements, search queries, and other publicly available content from social media platforms like TikTok, YouTube, and Instagram. If text is not natively accessible through accessibility features, this process can be complemented by established data processing techniques, such as OCR (Optical Character Recognition) algorithms, to extract relevant information. Beyond app interactions, the Accessibility Service allows for the monitoring of visited websites, providing insights into browsing behavior by tracking domain visits and search activities.

5.3 Operationalization of media exposure and in-app content

Building on the methodological approach of the first case study, the second case study applies a structured, multi-stage approach to operationalizing and linking media exposure with in-app content. The main stages of this process are illustrated in Figure 3.

Figure 3: Data enrichment pipeline



Building on the general framework illustrated in Figure 3, the following subsections detail the procedures used to collect, match, and enrich mobile tracking data to reconstruct participants' exposure to political content.

(1.) *Data Collection* captures raw tracking data, using mobile tracking approaches, providing a foundational dataset. However, raw data may be incomplete and fragmented. In this case study, we used the Accessibility Service-based tracking approach to capture posts from predefined accounts of interests and keywords. The raw dataset includes basic smartphone interaction metrics, such as the timestamp a post appeared and the partial description text visible on participants' screens (Table 2, *Raw Tracking Data*).

Table 2: The data enrichment pipeline

1. Raw tracking data			**2. Data matching**	**3. contextual enrichment**		**4. content analysis**
User ID	Timestamp	Raw Description	Full-Text Description	Account	Account Type	Political content?
1000	2025-09-01 11:39:03	Bayer Leverkusen hat sich von seinem Trainer Erik ten Hag getrennt. Das bestätigt der Club via X. Der Niederländer wurde erst zu dieser Saison verpflichtet. Nach nur zwei Spieltagen ist der 55-Jährige fr…	Bayer Leverkusen hat sich von seinem Trainer Erik ten Hag getrennt. Das bestätigt der Club via X. Der Niederländer wurde erst zu dieser Saison verpflichtet. Nach nur zwei Spieltagen ist der 55-Jährige freigestellt worden. Es ist der erste Trainer-Rauswurf in der laufenden Bundesliga-Saison.	RTLWest	Media	No
1000	2025-09-01 21:30:14	Die Fakenews der Woche: Trump ist nicht tot, Es gibt keinen Killerrob...	Die Fakenews der Woche: Trump ist nicht tot, Es gibt keinen Killerroboter namens Noisy, es gibt kein Kopftuchverbot und keine Tankbeschränkung. #1minutejura #nachrichten #lernenmittiktok	Herr Anwalt	Influencer	Yes

Note: Data in the table have been anonymized to protect participants' privacy.

Further, we collected data to create a content library (cache) based on two predefined input sources: an *account list* and a *keyword list*. The account list contained official TikTok profiles of political actors, parties, and institutions, all manually verified as *profiles of interest*. The keyword list included political terms, enabling the detection of additional relevant exposures beyond the predefined accounts.

All profiles from the account list were then systematically scraped to create a content library. While API-based data collection is preferred for its compliance with platform Terms of Service and its ability to systematically gather public metadata, it has limitations such as rate restrictions and unpredictable policy changes (Freelon, 2018). When APIs are insufficient, web scraping can serve as an alternative (Perriam et al., 2020), though it raises ethical and legal concerns related to platform policies and data privacy laws (Trezza, 2023). The content library contains full post-level metadata, including descriptions, publication dates, and engagement metrics such as likes, views, and comments, and serves as a structured reference database for the subsequent enrichment step. The data enrichment process links the raw mobile tracking data with the specific TikTok content that participants were exposed to. The goal is to reconstruct and identify the exact posts that appeared on screen during the observation period, enabling the analysis of content and actor characteristics.

(2.) The *Data Matching* step is required to process the data collected in the initial stage, filling in missing content fragments and refining the dataset's integrity. A key challenge is the fragmentation of textual content captured by methods like OCR. Mobile tracking often produces incomplete text snippets rather than full descriptions or captions. This issue is particularly common on dynamic mobile platforms (e.g., YouTube, TikTok), where text

frequently appears in scrolling interfaces or overlays, resulting in partial captures of the content viewed by users. For exampe, a user might see only a portion of a video description on screen, causing mobile tracking apps to record incomplete segments.

We performed exact and fuzzy string alignment, utilizing the Token-Sort-Ratio for its robustness in handling fragmented texts, to match the recorded text fragments with post descriptions from the scraped dataset. Through this matching process, 1,392 unique exposures were linked to corresponding posts in the content library. The resulting enriched dataset combined participants' individual exposure logs with additional post-level information.

(3.) *Contextual enrichment* integrates metadata and contextual variables, adding depth to the dataset and improving interpretability. In the second case study, this step incorporated additional metadata to expand and clarify the information retrieved through the matching process. After linking each exposure event to its corresponding TikTok post, we systematically added post-level context variables from the parallel scraping dataset. These included quantitative engagement indicators (likes, comments, shares, and view counts) as well as descriptive metadata such as publication date, video duration, and uploader identity. These additions transformed the enriched dataset from a purely behavioral log into a content-rich analytical resource. They allowed us to examine not only which political posts participants encountered, but also how visible and popular those posts were within the wider TikTok environment. For instance, by combining exposure frequency with view and like counts, we could distinguish between highly visible, viral posts and those with limited audience reach. Through this contextual enrichment, the dataset captured both the micro-level experiences of individual users and the broader communicative environment in which political content circulates (see Table 2).

(4.) The final *in-depth content analysis* applies advanced analytical techniques to extract further thematic and semantic insights. This step is designed to extract thematic and semantic insights from the TikTok posts linked to participants' exposure data. While the contextual enrichment stage added engagement metrics and structural information to the raw mobile tracking records, this stage goes further by using LLMs to classify previously matched TikTok video descriptions as either political or non-political, utilizing LLama 70B.

We note that, in this case study, we have by no means exhausted the possibilities of manual or automated methods of analyzing social media content. Further opportunities to expand the dataset include conducting more in-depth content textual analysis, or placing greater focus on image and video data.

Text-Level. The in-depth content analysis step systematically examines these complex aspects by integrating computational text analysis techniques, such as *automated sentiment and stance detection. Sentiment and stance detection* are key methods for understanding the emotional and ideological dimensions of digital content. Sentiment analysis applies natural language processing (NLP) techniques to classify text as positive, negative, or neutral, providing insights into the affective dimensions of media consumption (Liu, 2012). While sentiment analysis captures the emotional tone embedded in content, stance detection identifies ideological biases and patterns of selective exposure (Bestvater & Monroe, 2023; AlDayel & Magdy, 2021).

Image and Video-Analysis. While the present case study focused primarily on textual content, future work could extend the data enrichment pipeline by incorporating additional research tools, such as visual and video content analysis, further broadening the analytical scope. Multimodal content analysis integrates computer vision, speech-to-text processing, and video analysis to assess the thematic and stylistic characteristics of images, videos, and multimedia posts. Techniques such as image classification and object recognition

help categorize visual elements, distinguishing between different content types, including news graphics, political advertisements, and entertainment media. Speech recognition and automated transcription enable textual analysis of spoken content in video-based platforms like YouTube and TikTok, allowing researchers to apply sentiment and stance detection to audiovisual narratives. By combining textual, visual, and auditory cues, multimodal content evaluation provides a more comprehensive understanding of digital media exposure.

5.4 Results and discussion

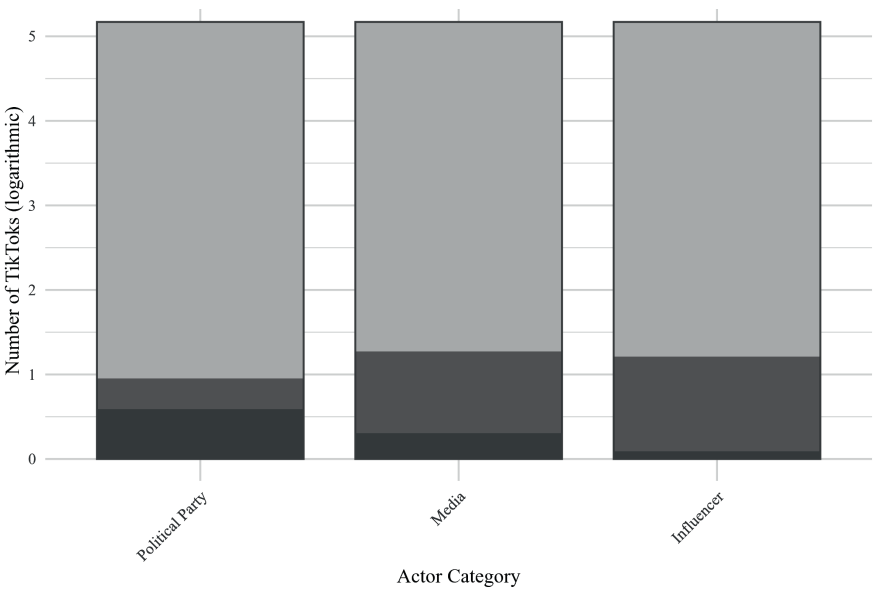
The second research question (RQ2) aims to capture research-relevant exposure within smartphone apps. Enriching the exposure data with a predefined list of relevant TikTok accounts, including media outlets, politicians, and influencers, enables analysis of participants' exposure to posts from these actors.

Figure 3 shows the daily average number of impressions per participant by actor category, with the number of TikToks consumed varying across participants and days.

Both political party and media content are present in the media mix. However, daily exposure to political posts is relatively low, suggesting that these actors adopt a more entertainment-oriented approach on TikTok. The comparatively low proportion of influencer content is also noteworthy. This may be due to limitations in identifying relevant accounts from the available account list, meaning that not all relevant posts were captured.

Nevertheless, this analysis demonstrates the potential of linking exposure with in-app content to gain insights into the individual composition of TikTok feeds within our sample.

Figure 3: Composition on TikTok Exposure per Actor Category (N=452)



Note: light gray = total daily TikToks per person; medium gray = category share; dark grey = share of political posts

In summary, the second case study adopts an integrated approach, transforming fragmented, surface-level tracking data into a structured dataset suitable for in-depth analysis. Each phase addresses specific limitations while building on the output of the previous step. While raw data alone often lacks coherence and context, targeted reconstruction and contextual layering make it possible to interpret not only *what* media content was encountered, but also *how*, *why*, and *with what potential impact*. The strength of this multi-process lies in its cumulative logic: each step enhances the interpretability and analytical depth of the data, enabling researchers to trace connections between exposure patterns, content characteristics, and user engagement. This highly customizable, layered structure opens new possibilities for research questions that cut across different steps, such as linking specific advertising themes to patterns of engagement, or examining content framing varies across sessions and platforms. In doing so, the pipeline not only improves data quality, but also expands the scope of what can be empirically observed using mobile tracking methods.

6. Conceptual Key Principles

When working with (mobile) automated tracking data, a distinctive feature is that data preparation constitutes a significant part of the project, as it can be complex and directly influence subsequent results.

One reason for the need for thorough processing of mobile automated tracking data is that such data was not originally created for scientific purposes, but primarily serves technical or administrative functions (Breiter & Hepp, 2018; Riebling, 2019). Consequently, it must be transformed into a format that is meaningful for research. It is important to recognize that, like all types of digital trace data, automated tracking data is not a completely unaltered raw material (Breiter & Hepp, 2018). In fact, this type of data is shaped by social institutions and technological products or devices they produce (Breiter & Hepp, 2018; Freelon, 2014). Given this context, it can be inferred that social, political, and societal assumptions may be embedded in the data (Breiter & Hepp, 2018).

We align with the perspective of Breiter & Hepp (2018) that tracking data, like other digital traces, reveals its full significance when viewed within a broader context or connected to real-world scenarios (Breiter & Hepp, 2018). Building on this understanding, we propose that the processing of tracking data should focus on five specific dimensions.

(1) The first dimension emphasizes that the privacy of the study participants must be protected, a principle reinforced by the concept of data minimalism. Accordingly, all information that is not relevant to the research interest is removed from the data to be analyzed. During pre-processing, additional measures were implemented to safeguard participant privacy. These included the categorizing apps to reduce the risk that third parties could comprimise the pseudonymization of the participants. Furthermore, sensitive and non-relevant apps—such as those related to health, dating, shopping, games, and finance—were replaced with placeholders.

(2) The second dimension involves the meaningful and targeted processing of automated tracking data into interpretable units of information. In our study, the goal was to obtain as comprehensive and context-rich a picture of information use as possible. To achieve this, we aggregated the data into sessions, defined as sequential, uninterrupted sequences of observed usage behavior (Peng & Zhu, 2020; Zhu et al., 2018). This session-based approach allows additional information about mobile usage to be captured and analyzed, while maintaining a manageable unit of analysis. It also enables the inclusion of previously or subsequently used apps or app categories facilitating the identification of broader usage patterns (Tong et al., 2022).

(3) The third dimension emphasizes the incorporation of a socio-technical perspective, including reflection on structural biases. Tracking data, as a form of digital trace data generated by technical systems for specific purposes, inherently exhibits biases (Breiter & Hepp, 2018). Researchers should take care to ensure that these biases are minimized, or ideally absent, in the research results. To achieve this, we recommend employing mixed-methods designs that incorporate user perspectives through quantitative or qualitative (survey-)methods. Combined with critical reflection on platform logics, this approach provides a foundation for examining methodological choices and analysis strategies.

(4) The fourth dimension involves data enrichment through the accurate linking of information from different sources to create a more comprehensive dataset that supports the analysis of smartphone usage behavior. By integrating recorded terms or content with an application that logs the start and end times of apps usage (Ferreira et al., 2015), it becomes possible to connect tracked events to specific applications. Data collected in this manner provides insights into the content participants accessed (as shown in Table 1 & 2), the channels through which they received the information (e.g., Telegram), and how they subsequently engaged with that content (e.g., continued communication).

(5) The fifth dimension emphasizes replicability by third parties. In this context, precise and comprehensible documentation of the research approach is crucial, as is the avoidance of unnecessarily complex procedures—adhering to the principle of keeping the process as simple as possible.

7. When to use which approach?

While the first approach focuses on identifying relevant exposure, the second goes further by capturing the associated content and linking it to that exposure. Both approaches also allow for the recording of device context—that is, upstream and downstream events on the mobile phone, including previously or subsequently opened apps, as well as exposure to content from different entities. They additionally capture the general duration of smartphone use, both overall and within specific applications. Accordingly, both approaches enable the mapping of a relatively accurate picture of smartphone usage practices. Despite these similarities, the two approaches differ in terms of (1) the type of content data, (2) the scope of data collection, and (3) the accessibility of the required data collection tools.

The identification of exposure enables the recording of research-relevant terms displayed on the smartphone across all apps. Unlike the second approach, there is no need to preselect specific social media platforms, allowing the first approach to adopt a more open and exploratory strategy. However, this openness comes at the cost of depth, as the actual content viewed by participants is not recorded.

The ability to capture content and link it to the corresponding exposure is the primary advantage of the second approach. Although it is less open due to the preselection of one or more social media platforms, it provides both the content viewed as and the associated social media metrics. Recording the actual content also enables further content analysis, which can integrate social media metrics with the individually tracked usage behavior of participants. Accordingly, this approach allows for the creation of a comprehensive picture, although it requires a certain degree of prior knowledge is required, such as lists of relevant accounts and keywords.

The granularity of the data is reflected in the scope of data collection. Collecting data to identify relevant exposure requires relatively fewer steps: first, the mobile tracking sensors (e.g., screen and app usage) are selected; next, the keywords are chosen; a database is set up (e.g., PostgreSQL or MySQL); and the appropriately configured apps are installed on participants' devices. Capturing content data and linking it to exposure, however, involves

additional steps. This includes obtaining a unique identifier for each piece of content, setting up and configuring the scraper, and subsequently scraping the content. Consequently, the collection of content entails a more complex and multi-step data collection process.

Furthermore, the identification of exposure focuses on capturing a comprehensive record of user behavior on the entire smartphone. Due to its high granularity and the ability to aggregate data at different levels, this approach offers considerable flexibility in analytical options. For example, aggregated usage times per app across all participants can be examined alongside the visibility of climate-protest terms, such as *Fridays for Future*, within these apps (see Figure 2). Other possibilities include identifying instances when the smartphone keyboard is used simultaneously with the appearance of protest-relevant terms, which may indicate interpersonal or one-to-many communication in the context of climate protests.

In contrast, collecting content and linking it to exposure allows for more extensive analytical possibilities, including the examination of content characteristics, but it also requires additional prerequisites. Crucially, the content must be accessible, either through scraping or API access, which typically necessitates knowledge of an account name or post ID. Moreover, this approach is mostly limited to social media content and is only partially transferable to other mobile applications, such as messaging platforms. Nevertheless, it supports different levels of data aggregation, enabling further analyses of users' direct interactions with various types of content, particularly through recorded content features and engagement metrics, such as topics, stances, and likes.

8. Limitations of mobile tracking

Mobile automated tracking data presents several challenges and limitations. While it is often regarded as a reliable method for capturing digital information and media usage (Parry et al., 2021), it has notable drawbacks. This type of data provides a high level of detail and numerous measurement points, resulting in a dense and complex dataset.

The tracking app primarily logs apps brought to the foreground, meaning that only visible apps are recorded. As a result, background apps—particularly audio and music applications that continue playing without user interaction—may be underrepresented, potentially leading to an underestimation of audio usage duration.

Furthermore, most tracking apps—such as the AWARE client, Keywordlogger, and the software used in the second case study—are Android-only. This limitation may introduce sample bias, as iOS users are not represented in the study.

Most tracking apps, such as the AWARE client, also capture system apps and other background processes that are irrelevant to the research questions. These are categorized as system apps and excluded from analysis. However, this labeling process can introduce errors, as the sample includes a variety of smartphones with different Android versions, each of which may name system apps and background processes differently.

Additionally, the enrichment of tracking data depends on data quality. Data retrieved from screen recordings is often noisy and fragmented due to inconsistencies, incomplete logging, and platform variations. Similarly, API- and scraping-based data may also be incomplete, as public data is not always consistent, and official APIs impose restrictions or provide only sample datasets, which can make it difficult to accurately assign content to tracked exposure.

Another important consideration is how accurately automated tracking data reflects genuine human behavior. Two key aspects are particularly relevant: distinguishing between data points generated by human behavior and those resulting from technical artifacts. In mobile automated tracking data, it can be challenging to differentiate between actions

stemming from human interaction, such as checking the time, and data produced by technical processes, such as a notification triggering the screen to turn on. Both types of data are recorded in the tracking logs, but only the former represents actual human engagement. Furthermore, tracking data cannot provide definitive information about 'users' attention to the content displayed on their devices. Accordingly, we advocate for a realistic understanding of the capabilities of mobile automated tracking and recommend employing mixed-methods approaches to more effectively capture the intricacies of human behavior.

9. Conclusion

Our paper illustrates the methodological possibilities of in-app tracking using two case studies. The first case study focuses on identifying relevant exposure within smartphone apps, while the second demonstrates how content can be captured, enriched, and linked to exposure. From these studies, we derive five conceptual core principles for working with in-app tracking data: privacy and data minimalism, meaningful preprocessing, incorporation of a socio-technical perspective, data enrichment, and replicability. Additionally, we highlight the strengths and limitations of both approaches.

This work contributes to the development of in-app measurement methods, tools, and standards, as well as to conceptual thinking about data handling and analysis. We hope that these methodological advances will inspire further research to illuminate in-app user behavior.

References

Adam, S., Makhortykh, M., Maier, M., Aigenseer, V., Urman, A., Gil Lopez, T., Christner, C., de León, E., & Ulloa, R. (2024). Improving the quality of individual-level web tracking: Challenges of existing approaches and introduction of a new content and long-tail sensitive academic solution. https://doi.org/10.1177/08944393241287793

AlDayel, A., & Magdy, W. (2021). Stance detection on Social Media: State of the art and trends. Information Processing & Management, 58(4), 102597. https://doi.org/10.1016/j.ipm.2021.102597

Andone, I., Błaszkiewicz, K., Eibes, M., Trendafilov, B., Montag, C., & Markowetz, A. (2016). How age and gender affect smartphone usage. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, 9–12. https://doi.org/10.1145/2968219.2971451

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. Communication Methods and Measures, 11(3), 173–190.

Behre, J., Hölig, S., & Möller, J. (2024). Reuters Institute Digital News Report 2024: Ergebnisse für Deutschland. Verlag Hans-Bredow-Institut. https://www.ssoar.info/ssoar/handle/document/94461

Bestvater, S. E., & Monroe, B. L. (2023). Sentiment is not stance: Target-aware opinion classification for political text analysis. Political Analysis, 31(2), 235–256. https://doi.org/10.1017/pan.2022.10

Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011). Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, 47–56.

Bosch, O. J., Sturgis, P., Kuha, J., & Revilla, M. (2024). Uncovering digital trace data biases: Tracking undercoverage in web tracking data. Communication Methods and Measures, 1–21. https://doi.org/10.1080/19312458.2024.2393165

Breiter, A., & Hepp, A. (2018). The complexity of datafication: Putting digital traces in context. Communicative figurations: Transforming communications in times of deep mediatization, 387–405.

Christner, C., Urman, A., Adam, S., & Maier, M. (2022). Automated tracking approaches for studying online media use: A critical review and recommendations. Communication Methods and Measures, 16(2), 79–95. https://doi.org/10.1080/19312458.2021.1907841

Clemm von Hohenberg, B., Stier, S., Cardenal, A. S., Guess, A. M., Menchen-Trevino, E., & Wojcieszak, M. (2024). Analysis of web browsing data: A guide. Social Science Computer Review, 42(6), 1479–1504. https://doi.org/10.1177/08944393241227868

Fan, Y., Tu, Z., Li, T., Cao, H., Xia, T., Li, Y., Chen, X., & Zhang, L. (2021). Understanding the long-term dynamics of mobile app usage context via graph embedding. IEEE Transactions on Knowledge and Data Engineering, 35(3). https://doi.org/10.1109/TKDE.2021.3110141

Ferreira, D., Kostakos, V., & Dey, A. K. (2015). AWARE: Mobile context instrumentation framework. Frontiers in ICT, 2. https://doi.org/10.3389/fict.2015.00006

Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. Journal of Broadcasting & Electronic Media, 58(1), 59–75. https://doi.org/10.1080/08838151.2013.875018

Freelon, D. (2018). Computational research in the post-API age. Political Communication, 35(4), 665–668. https://doi.org/10.1080/10584609.2018.1477506

Freelon, D., Pruden, M. L., Malmer, D., Wu, Q., Xia, Y., Johnson, D., Chen, E., & Crist, A. (2024). What's in your PIE? Understanding the contents of personalized information environments with PIEGraph. Journal of the Association for Information Science and Technology, 75(10), 1119–1133. https://doi.org/10.1002/asi.24869

Frisson, C., Malacria, S., Bailly, G., & Dutoit, T. (2016). Inspectorwidget: A system to analyze users' behaviors in their applications. 1548–1554. https://doi.org/10.1145/2851581.2892388

Gitelman, L. (Ed.). (2013). "Raw data" Is an oxymoron. MIT Press.

Görland, S. (2020). Medien, Zeit und Beschleunigung: Mobile Mediennutzung in Interimszeiten. Springer.

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. Nature Human Behaviour, 4(5), 472–480. https://doi.org/10.1038/s41562-020-0833-x

Hasebrink, U., & Popp, J. (2006). Media repertoires as a result of selective media use. A conceptual approach to the analysis of patterns of exposure. Communications, 31(3), 369–387. https://doi.org/10.1515/COMMUN.2006.023

Hong, J. I., Boriello, G., Landay, J. A., McDonald, D. W., Schilit, B. N., & Tygar, J. D. (2003). Privacy and security in the location-enhanced world wide web. Proceedings of Fifth International Conference on Ubiquitous Computing: Ubicomp, https://scholar.google.com/citations?view_op=view_citation&hl=en&user=MoFbcc0AAAAJ&cstart=300&pagesize=100&sortby=pubdate&citation_for_view=MoFbcc0AAAAJ:4TOpqqG69KYC.

Jungherr, A. (2015). Analyzing political communication with digital trace data: The role of twitter messages in social science research. Springer.

Jungherr, A., & Theocharis, Y. (2017). The empiricist's challenge: Asking meaningful questions in political science in the age of big data. Journal of Information Technology & Politics, 14(2), 97–109. https://doi.org/10.1080/19331681.2017.1312187

Jürgens, P., & Stark, B. (2022). Mapping exposure diversity: The divergent effects of algorithmic curation on news consumption. Journal of Communication, 72(3), 322–344. https://doi.org/10.1093/joc/jqac009

Krieter, P. (2019). Can I record your screen? Mobile screen recordings as a long-term data source for user studies. 1–10. https://doi.org/10.1145/3365610.3365618

Krieter, P. (2020). Looking inside—mobile screen recordings as a privacy friendly long-term data source to analyze user behavior. Universität Bremen. https://doi.org/10.26092/elib/103

Krieter, P., Zerrer, P., Puschmann, C., & Geise, S. (2024). Following topics across all apps and media formats: Mobile keyword tracking as a privacy-friendly data source in mobile media research. Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 126–131.

Kühnemann, H. (2021). Anwendungen des Web Scraping in der amtlichen Statistik. AStA Wirtschafts- Und Sozialstatistisches Archiv, 15(1), 5–25. https://doi.org/10.1007/s11943-021-00280-5

Kulshrestha, J., Oliveira, M., Karaçalık, O., Bonnay, D., & Wagner, C. (2021). Web routineness and limits of predictability: Investigating demographic and behavioral differences using web tracking data. 15, 327–338. https://doi.org/10.48550/arXiv.2012.15112

Liu, B. (2012). Sentiment analysis and opinion mining. Springer International Publishing. https://doi.org/10.1007/978-3-031-02145-9

Maier, M., Adam, S., Gil Lopez, T., Makhortykh, M., Bromme, L., Christner, C., de León, E., & Urman, A. (2025). Populist radical-right attitudes, political involvement and selective information consumption: Who tunes out and who prefers attitude-consonant information. Mass Communication and Society, 28(1), 101–129. https://doi.org/10.1080/15205436.2024.2310156

McMillan, D., McGregor, M., & Brown, B. (2015). From in the wild to in vivo: Video Analysis of Mobile Device Use. 494–503. https://doi.org/10.1145/2785830.2785883

Meijer, C. I., & Kormelink, G. T. (2015). Checking, sharing, clicking and linking: Changing patterns of news use between 2004 and 2014. Digital Journalism, 3(5), 664–679. https://doi.org/10.1080/21670811.2014.937149

Menchen-Trevino, E., & and Karr, C. (2012). Researching real-world web use with roxy: Collecting observational web data with informed consent. Journal of Information Technology & Politics, 9(3), 254–268. https://doi.org/10.1080/19331681.2012.664966

Merten, L. (2021). Block, hide or follow—Personal news curation practices on social media. Digital Journalism, 9(8), 1018–1039. https://doi.org/10.1080/21670811.2020.1829978

Möller, J., van de Velde, R. N., Merten, L., & Puschmann, C. (2020). Explaining online news engagement based on browsing behavior: Creatures of habit? Social Science Computer Review, 38(5), 616–632. https://doi.org/10.1177/0894439319828012

Muise, D., Markowitz, D., Reeves, B., Ram, N., & Robinson, T. (2024). (Mis) measurement of political content exposure within the smartphone ecosystem: Investigating common assumptions. Journal of Quantitative Description: Digital Media, 4.

Munzert, S., & Nyhuis, D. (2019). Die Nutzung von Webdaten in den Sozialwissenschaften. In C. Wagemann, A. Goerres, & M. Siewert (Eds.), Handbuch Methoden der Politikwissenschaft (pp. 1–25). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-16937-4_22-1

Naab, T. K., Karnowski, V., & Schlütz, D. (2019). Reporting mobile social media use: How survey and experience sampling measures differ. Communication Methods and Measures, 13(2), 126–147. https://doi.org/10.1080/19312458.2018.1555799

Ohme, J., Araujo, T., Boeschoten, L., Freelon, D., Ram, N., Reeves, B. B., & Robinson, T. N. (2024). Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. Communication Methods and Measures, 18(2), 124–141. https://doi.org/10.1080/19312458.2023.2181319

Parry, D., Davidson, B., Sewall, C., Fisher, J., Mieczkowski, H., & Quintana, D. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. Nature Human Behaviour, 5, 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Parry, D., & Toth, R. (2025). Extracting meaningful measures of smartphone usage from Android event log data: A methodological primer. Computational Communication Research, 7(1), 1.

Peng, T.-Q., & Zhu, J. J. H. (2020). Mobile phone use as sequential processes: From discrete behaviors to sessions of behaviors and trajectories of sessions. J. Comput. Mediat. Commun., 25(2), 129–146. https://doi.org/10.1093/jcmc/zmz029

Perriam, J., Birkbak ,Andreas, & and Freeman, A. (2020). Digital methods in a post-API environment. International Journal of Social Research Methodology, 23(3), 277–290. https://doi.org/10.1080/13645579.2019.1682840

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., Chiatti, A., Cho, M., Roehrick, K., & Yang, X. (2021). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. Human–Computer Interaction, 36(2), 150–201. https://doi.org/10.1080/07370024.2019.1578652

Richter, S., & Stier, S. (2022). Learning about the unknown Spitzenkandidaten: The role of media exposure during the 2019 European Parliament elections. European Union Politics, 23(2), 309–329. https://doi.org/10.1177/14651165211051171

Riebling, J. (2019). The medium data problem in social science. In C. M. Stützer, M. Welker, & M. Egger (Eds.), Computational Social Science in the Age of Big Data. Concepts, Methodologies, Tools, and Applications (pp. 77–103). Herbert von Halem Verlag.

Scharkow, M. (2016). The accuracy of self-reported internet use—A validation study using client log data. Communication Methods and Measures, 10(1), 13–27. https://doi.org/10.1080/19312458.2015.1118446

Sen, I., Flöck, F., Weller, K., Weiß, B., & Wagner, C. (2021). A total error framework for digital traces of human behavior on online platforms. Public Opinion Quarterly, 85(S1), 399–422. https://doi.org/10.1093/poq/nfab018

Spiekermann, S., & Cranor, L. F. (2009). Engineering privacy. IEEE transactions on software engineering, 35(1), 67–82. https://doi.org/10.1109/TSE.2008.88

Stier, S., Bleier, A., Bonart, M., Mörsheim, F., Bohlouli, M., Nizhegorodov, M., Posch, L., Maier, J., Rothmund, T., & Staab, S. (2018). Systematically monitoring social media: The case of the German federal election 2017 (GESIS Papers). GESIS - Leibniz-Institut für Sozialwissenschaften. https://doi.org/10.21241/ssoar.56149

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. Social Science Computer Review, 38(5), 503–516. https://doi.org/10.1177/0894439319843669

Stier, S., Kirkizh, N., Froio, C., & Schroeder, R. (2020). Populist attitudes and selective exposure to online news: A cross-country analysis combining web tracking and surveys. The International Journal of Press/Politics, 25(3), 194016122090701. https://doi.org/10.1177/1940161220907018

Stier, S., Mangold, F., Scharkow, M., & Breuer, J. (2021). Post post-broadcast democracy? News exposure in the age of online intermediaries. American Political Science Review, 1–7. https://doi.org/10.1017/s0003055421001222

Tong, L., Mingyang, Z., Hancheng, C., Yong, L., Sasu, T., & Pan, H. (2020). What apps did you use? : Understanding the long-term evolution of mobile app usage. 66–76. https://doi.org/10.1145/3366423.3380095

Tong, L., Tong, X., Huandong, W., Zhen, T., Sasu, T., Zhu, H., & Pan, H. (2022). Smartphone app usage analysis: Datasets, methods, and applications. IEEE communications surveys and tutorials, 24(2), 937–966. https://doi.org/10.1109/comst.2022.3163176

Toth, R., & Trifonova, T. (2021). Somebody's watching me: Smartphone use ttracking and reactivity. Computers in Human Behavior Reports, 4, 100142. https://doi.org/10.1016/j.chbr.2021.100142

Trezza, D. (2023). To scrape or not to scrape, this is dilemma. The post-API scenario and implications on digital research. Frontiers in Sociology, 8. https://doi.org/10.3389/fsoc.2023.1145038

von Andrian-Werburg, M. T., Siegers, P., & Breuer, J. (2023). A Re-evaluation of Online Pornography Use in Germany: A Combination of Web Tracking and Survey Data Analysis. Archives of Sexual Behavior, 1–13. https://doi.org/10.1007/s10508-023-02666-8

Wieland, M. (2023). Informiert oder (doch nur) abgelenkt? Potenziale und Herausforderungen automatisierter Nachrichtenkontakte in mobilen sozialen Medien. Herbert von Halem Verlag.

Wojcieszak, M., de Leeuw, S., Menchen-Trevino, E., Lee, S., Huang-Isherwood, K. M., & Weeks, B. (2023). No polarization from partisan news: Over-time evidence from trace data. The International Journal of Press/Politics, 28(3), 601–626. https://doi.org/10.1177/19401612211047194

Yee, A. Z. H., Yu, R., Lim, S. S., Lim, K. H., Dinh, T. T. A., Loh, L., Hadianto, A., & Quizon, M. (2023). ScreenLife Capture: An open-source and user-friendly framework for collecting screenomes from Android smartphones. Behavior Research Methods, 55(8), 4068–4085. https://doi.org/10.3758/s13428-022-02006-z

Zerrer, P. T. (2024). Political action and news use of the Fridays for Future movement in Germany. https://doi.org/10.26092/elib/3604

Zerrer, P., Puschmann, C., & Pressmann, P. (2025, January 28). German Politics Online: A List of social media accounts. https://doi.org/10.17605/OSF.IO/A7SJD

Zhu, J. J. H., Chen, H., Peng, T.-Q., Liu, X. F., & Dai, H. (2018). How to measure sessions of mobile phone use? Quantification, evaluation, and applications. Mobile Media & Communication, 6(2), 215–232. https://doi.org/10.1177/2050157917748351

Riding the Spider: A Network-Sampling Framework for Multi-Platform Data Collections

Philipp Kessling / Felix Victor Münch*

Research on the digital networked public sphere is not only hindered by challenges in data access but also by a lack of common standards for describing and implementing data collection independently of the form of access or technologies employed. These challenges are particularly pronounced in cross-platform research. In this article, we propose a network-sampling framework to conceptualize, implement, and document explorative data collections in a generalizable, readily operationalizable, and interoperable way. Building on the theoretically established components of the networked public sphere, the concept of multilayer networks, explorative network sampling, and legal and technical realities of cross-platform data access, we segment the data collection process into four modules: a Connector, a Parser, a Filter, and a Sampler. This framework enables researchers not only to describe their data collection in a precise and reproducible way but also to follow guidelines on for developing interoperable software implementations of these modules or to propose new modules themselves.

Key words: network sampling, cross-platform data collection, networked public sphere, multilayer networks, interoperability, replicability

1. Introduction

The study of digital, networked communication is challenged by a threefold, interconnected increase in complexity (Strippel et al. 2018). First, the media and communication landscape has diversified into a multitude of platforms and their associated affordances (see, e.g., Breiter and Hepp 2018). Second, the methods used to study these platforms have become increasingly complex, partly due to the adoption of approaches from the natural and computer sciences (Berry 2011; Lazer et al. 2020). Third, platforms have repeatedly modified their data access policies and interfaces, often restricting or entirely cutting off access to researchers. These changes pose significant obstacles to maintaining consistent and reliable data sources for longitudinal studies and cross-platform comparisons (Bruns 2019; Freelon 2021). Consequently, there is a pressing need for a formal framework to describe data collections that supports the study of larger structures of the public sphere, that is generalizable across studies, and remains reproducible over time.

Meanwhile, pressing societal issues underscore the need for a birds-eye view of digital publics. These include the fragmentation of democratic, liberal societies into smaller, less interconnected communities, which may contribute to polarization (Brüggemann and Meyer 2023; Esau et al. 2024) as well as the strategic use of mis- and disinformation (Rogers 2023; Quandt et al. 2019), which can erode trust in public institutions and established media (Frischlich and Humprecht 2021). Research that investigates the digital networked public and can capture both its short- and long-term structures, from individual to global scales,

* Philipp Kessling, M.A., Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI), Warburgstraße 30b, 20354 Hamburg, Germany, p.kessling@leibniz-hbi.de, https://orcid.org/0000-0002-1739-446X;
Dr. Felix Victor Münch, GESIS Leibniz-Institut für Sozialwissenschaften, Knowledge Technologies for the Social Sciences, Unter Sachsenhausen 6-8, 50667 Köln, Germany, Felix.Muench@gesis.org, https://orcid.org/0000-0001-8808-6790.

 DOI: 10.5771/1615-634X-2026-1-52

is therefore crucial. While studies of short-term effects remain feasible despite the restrictive work-to-rule approach of major platforms toward researcher access (Rau et al., 2025), long-term and large-scale analyses—whether within single platforms (Guan et al. 2022; Münch et al. 2021; Bruns and Moon 2019; Bruns et al. 2017) or across multiple platforms—are severely slowed and rendered prohibitively costly, if not altogether impossible.

This challenge necessitates a shift toward reliable and reproducible data acquisition if research independent and free from conflicts of interest, outside the tightly controlled domains of major platforms—is to remain viable. Current efforts to address this issue, however, largely consist of uncoordinated (though necessary and commendable), single-use attempts to circumvent platform-imposed restrictions. Methods for collecting data at scale, particularly for empirically mapping digital publics both within and across platforms remain fragmented and difficult to reproduce (Wiedemann et al. 2023).

In this article, we propose and exemplify the necessity and feasibility of a generalizable, readily operationalizable framework to conceptualize, implement, and document large-scale data collections that are independent of individual platforms, yet compatible with empirically established theories of the networked public sphere. In the remainder of this article, we discuss theoretical models of networked publics (Bruns 2023; Friemel and Neuberger 2023) and examine their relationship to explorative, network-based sampling of trace data. In doing so, we aim to demonstrate an approach for gathering relevant datasets that

1. integrates a variety of data collection methods in a flexible manner, which is necessary due to the unreliability of platforms, and
2. openly and consistently documents the collection intent and process in a replicable way.

To achieve this, we propose dividing the data collection process into four distinct modules: a Connector, which retrieves research data from platform APIs, scrapers, databases, files, and other sources; a Parser, which extracts meaningful network edges from the retrieved data; a Filter, which selects or rejects edges and nodes based on researcher defined conditions; and a Sampler, which selects edges and nodes for continuation of the sampling process. This modular segmentation allows researchers to describe their data collection approach clearly and reproducibly, while also providing a framework for developing interoperable software implementations of these modules or proposing new ones.

In Section 2 we motivate a multilayer network perspective grounded in theoretical concepts of the (networked) public sphere(s). In Section 3, we introduce approaches to explorative network sampling and assess their suitability for research within this paradigm. Section 4 presents an example case, and in Section 5, we deduce a sampling framework for data collections.

2. Digital, Networked and Public Spheres

Discussing the public sphere today necessarily involves the *Internet*, where most public communication either occurs directly or is documented and reverberated, influencing society. This produces collectable trace data at varying levels of ephemerality, making communication more observable and therefore measurable—a development that contributed to the computational turn in the social sciences (Lazer et al., 2009). Even on early Internet message boards, and increasingly with the emergence of the Blogosphere and the early stages of Twitter in the mid- to late 2000s, previously silent users of media with limited audiences became what Bruns (2008) calls 'produsers'—media users who not only consume media as passive recipients but also produce their own content for audiences across the growing World Wide Web. Combined with the interconnected technical nature of the web, this led to a media landscape shifting from centralized, broadcast-oriented media toward

networked, distributed, fine-grained contributions from anyone willing to participate. As a result, members of the general public and non-traditional content producers now share the arena with established media outlets, rendering old theoretical frameworks for understanding the public sphere outdated (Schmidt et al., 2017). In particular, the notion of a virtual stage dominated by normatively guided deliberation processes (Habermas, 1962) is challenged, prompting the need for new conceptual building blocks to describe networked public spheres. For instance, Bruns (2023) and Friemel and Neuberger (2023) define the basic entities in the digital public as follows:

- *Actors*: Humans and/or aggregations of humans in organizations, as well as automated computer systems that interact with humans or with other automated systems, such as chatbots or social bots[1].
- *Content*: Actor-generated content hosted on one or more platforms and shaped by the platforms' affordances.
- *Platforms*: As discussed in Section 4, platforms and their affordances often create barriers that prevent actors and content from leaving them. Nevertheless, substantial cross-platform diffusion occurs (see below), and different platforms play distinctive roles in the public sphere. Historically, the choice of platform for research was largely dictated by data availability, but this approach is far from ideal.

Further, Friemel and Neuberger (2023) theorize link types between actors and content as *production, reception*, or *curation*. Depending on a platform's affordances, these types can be mapped to platform-specific connections. The affordances of a *single* platform already produce a complex interactive system with multiple modalities, interaction types, and entity classes (boyd and Ellison 2007). For example, Facebook supports different types of actors, such as user profiles and organizational pages, as well as various social ties, including 'friendships,' 'follows,' and 'likes.' However, given the observable shift from a few dominant social media platforms (Facebook, Twitter, Instagram) toward a more fragmented landscape—including platforms such as TikTok, Telegram, Bluesky, Mastodon, Threads, and Truth Social—single-platform research can no longer sustain the empirical foundations needed to advance our understanding of the public sphere. This underscores the necessity for a data collection and sampling framework that provides a generalizable, cross-platform, and researcher-definable model of connections between actors and content within their respective public spherules. From communication processes generating the above links, higher order constructs emerge:

Personal Publics: Enabled by connections within and across platforms, communication on social media in general and former Twitter in particular *"is happening in networked, distributed conversations: single tweets forming the basic units […] are bundled […] in the constant stream of information within a personal timeline, filtered via social connections made explicit"* (Schmidt 2014).

Issue Publics: These often materialize as *hashtag publics*. More broadly, they can be understood as transitive intersections of multiple personal publics centered around a specific issue or hashtag (Bruns and Burgess 2011).

Network of Public "Spherules": Personal and issue publics often give rise to longer-lasting figurations (Hasebrink and Hepp 2017; Hepp and Hasebrink 2014), that is, repeating and, over the mid- to long-term, stable meso-scale patterns. These can be regarded as small public spherules in their own right, such as right-wing counter-publics on Telegram or journalistic networks on Twitter.

1 We use this term broadly to acknowledge that automation exists in social networks, whether benign or malicious.

Combined with the need to account for the temporality of all these constructs, the (digital) public must be understood as a complex system characterized by emergent and dynamic phenomena (Waldherr 2017). These phenomena within networked publics arise from technical and socio-technical interactions, fundamentally based on the diffusion of a single content piece or digital resource over one or multiple steps. Due to the networked public sphere's centralization and structuring by large platforms, many digital resources can be considered to be members of these platforms. For objects residing on a platform, interactions between them can be captured as trace data. However, empirically investigating effects and phenomena in the networked public sphere requires a concrete operationalization of the linkage types under study. Accordingly, Friemel and Neuberger's categorization of *connections* (see above) can be interpreted as abstract roles to which specific linkage types within and across platforms can be mapped (2023). Useful trace data can include, for example:

1. Account-to-account interactions within a given platform (*reception*),
2. Account-to-content links, such as authorship indicators (*production*), and
3. In-platform links or hyperlinks that may direct to content or accounts on other platforms (*curation*).

To model a subset of the networked public sphere, various network models can be applied to facilitate analysis. Within a single platform, when considering only one type of linkage, a simple directed or undirected network may suffice. Extending the model to multiple platforms, and assuming that large groups of nodes represent objects of a single type—such as users of on a single platform—enables the networked public sphere to be represented as a layered network, with one node type per layer (see Kivelä et al., 2014, for an in-depth discussion of different formalizations). Allowing both multiple platforms and different linkage and node types within a single layer results in a network that is both multilayer and multiplex. In some cases, additional actor-to-account matching may be desirable to meet the requirements of multilayer network modeling.

3. Explorative Network Sampling and Web Crawling

To gather empirical insights within these models, networked data is required. Random sampling of content or actors typically produces networks that are too sparse to support inferences about global structures, making network sampling methods necessary. Network sampling involves drawing a sample based on network-like relations and can be applied either for network down-sampling (reducing the size and order of a network) or in an exploratory manner. Exploration usually begins at specific nodes, with their neighborhoods mapped by following edges (Hu and Lau 2013).

The networked public sphere model described above is largely unknown empirically. Random node samples, the network analogue of conventional quantitative social science sampling, are insufficient for revealing structural patterns, as they tend to be too sparse. Instead, subsets of this network *can* be explored from a limited set of digital resources by following, for example, links from content objects to other content or actors. For example, one might examine the timelines or profile pages of a given actor and the content embedded within them. Each content object contains outgoing links to both actors and other content objects—for instance, mentions of other actors or hyperlinks to additional content. By following these outgoing links, the neighborhood of the initial nodes can be discovered. Jost et al. (2023) apply this approach to map actor structures on Telegram.

The most widely known, large-scale samples of digital resource-networks with URL-encoded relations are referred to as crawls, which have long underpinned Internet search

engines. Today, several openly accessible crawls exist, such as CommonCrawl ("Common Crawl - Overview" n.d.) which provides regular crawls of the public Internet, with a single crawl encompassing multiple terabytes of website data. Historically, crawlers have been implemented for Internet research, for example, IssueCrawler was used to map pre-platformized versions of the Internet, such as the Blogosphere (Rogers 1996; Rogers 2010).

Although the basic technical approach is similar, the collection strategies used by search engines differ from those required for researching digital platforms and public networked spheres. For instance, search engines regularly revisit digital resources (Wolf et al. 2002) but often omit social media posts. Moreover, websites frequently optimize their content to be discoverable and highly indexed by search engines. In contrast, studying the structure of networked public spheres requires uncovering the internal structures of platforms to effectively navigate the barriers—"fences"—built by these platforms.

The literature documents a wide range of algorithms for exploratory network sampling beyond the exhaustive link-following typically employed by crawlers. Although many strategies are available, each algorithm introduces distinct biases and affords different properties in the resulting sampled network (Leskovec and Faloutsos 2006). For example, classical social network sampling procedures such as snowball sampling, can be adapted for use in URL-networks; however, depending on the breadth of the search (i.e., number of links followed per step) these methods may either become highly dependent on the initial seed set or, with greater breadth, remain overly local (Goodman 1961).

Ricaud et al. (2020) present a generalization of many network sampling algorithms that enables configurable, probability-based sampling incorporating both edge and node metadata. This approach is particularly well-suited for research on digital platforms and, by extension, the networked public sphere. Depending on the research objective, alternative methods, such as breadth first, depth-first tree extraction, random-walk sampling, or their derivatives, may be applied.

However, the theoretically optimal algorithm is not necessarily viable in practice. In the context of sampling networks from platforms, the choice of strategy is constrained by economic viability: some approaches are more costly than others in terms of API calls or the volume of scraping traffic required. Consequently, strategies that yield informative networks with the fewest possible requests are often preferable (Coscia and Rossi 2018). Taken together, selecting an appropriate algorithm and carefully parameterizing it allows for fine-grained control of how far and broadly the amplifying process traverses the overall, yet unknown, network.

4. Cross-Platform Data Modeling and Acquisition in Practice

Research on interplatform phenomena faces substantial challenges in data logistics and the operationalization of units of analysis, as differences in platform affordances hinder the identification of functional equivalences. As Heft et al. (2024) note, *"[…] studies across platforms and communication venues, thorough insights into platforms' general architectures (Bossetta, 2019) and their ways of structuring content and enabling access through various features are paramount (Pearce et al., 2020), as these fundamentally shape data collection possibilities and limitations.*"

At a technical level of abstraction, however, the Internet can be understood as a collection of interlinked digital resources, with platforms representing large aggregations of such resources, particularly when considering the definition of Very Large Online Platforms

(VLOPS)[2]. All of these resources are identified by a Universal Resource Identifier (*URI*), which on the Web typically takes the form of a Universal Resource Locator (*URL*) (Berners-Lee 1994). URLs enable Hyperlinks (i.e., (hyper)text containing embedded addresses), and can be regarded as the *connective tissue* of the Web, linking resources to one another and data exchange via APIs (Nielsen et al., 1996).

Hyperlinks create network structures between resources that were early on crawled and analyzed by researchers and search engines using network-analytic methods and metrics—such as Pagerank (Brin et al. 1998)—to understand the structure of the Web and rank websites by relevance. The scale of web crawling required for global analyses and rankings is immense; for example, the openly accessible CommonCrawl hosts multiple terabytes of data for a single crawl ("Common Crawl – Overview" n.d.).

Nevertheless, even such efforts fail to capture a rapidly growing segment of the Internet dominated by URLs primarily intended for use via APIs. Unlike the 'traditional' web of websites, APIs are interfaces that enable the fine-grained control over access and thus lend themselves to commodifying that access. The commodification of user generated content on *platforms*, and the resulting shift toward data connections between websites via APIs, was central to early discussions of platformization (Helmond 2015).

These walled gardens expand, as platforms like Instagram or Twitter/X increasingly restrict public data access, for example, by requiring logins for previously public content like tweet replies or Instagram stories. Yet even these environments remain collections of digital resources that largely possess URLs, despite platforms' efforts to channel users toward proprietary apps—thereby inserting additional layers of machines, behavioral analysis, and access control between users and content. This persistence of URLs reflects platforms' own interests in linking to content and steering users from the open Web into their controlled ecosystems.

A technical approach to addressing this power imbalance lies in diversifying data access risks by developing and supporting multiple ways of collecting platform data. Beyond official APIs, mechanisms such as data access requests and web scraping can provide researchers with at least a resource-constrained access. Applying (network) sampling strategies to the URL-based networks described above further mitigates the need for complete data, given the inherently skewed distributions of social media data (Barabasi and Albert 1999). Consequently, this approach enables the interchangeable use of diverse data sources, including APIs, web-scraping, data donations, tracking data, data repositories, and DSA-mandated data access (Ohme et al. 2023). Suitable sampling strategies also allow these sources to be used while reducing the likelihood of triggering automated rate limits and similar restrictions.

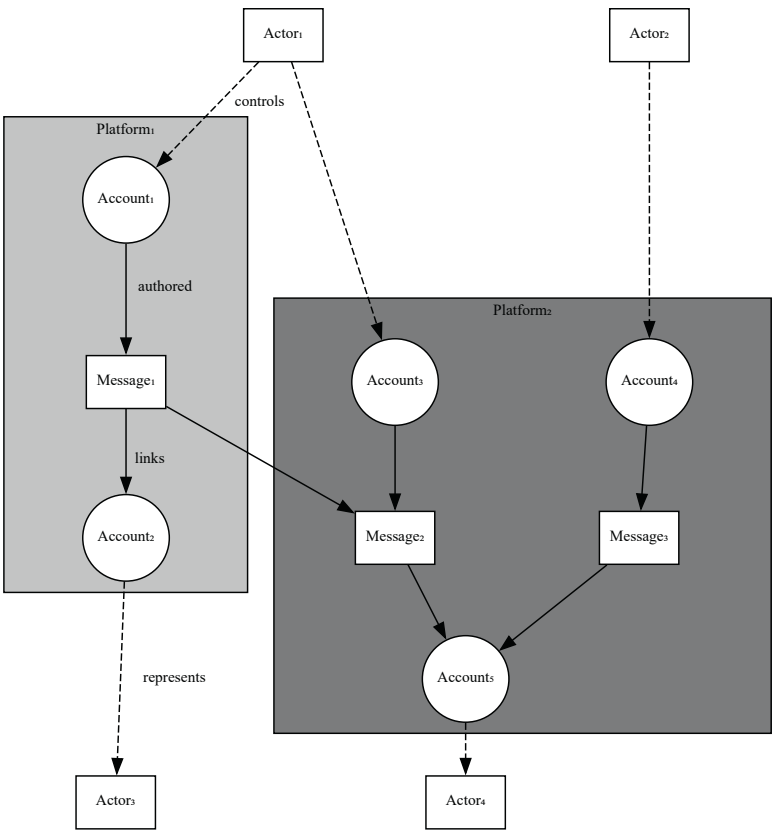
At the same time, this approach also introduces numerous additional degrees of freedom that threaten research reproducibility. It aggravates a foundational problem that hampers the integration of heterogeneous data sources: the imprecise description of data collection processes in much of the existing literature. In many cases—without intending to "name and blame," though examples are readily found—such descriptions are limited to broad statements about which data were collected from which platform and over which time frame, while crucial details such as the exact timing of collection are omitted. When sampling methods such as "snowball" or "random" sampling are mentioned, the specific variant, implementation, and parameterization are often left unspecified. Given this widespread lack of precision, and the substantial downstream effects that even minor changes in data collec-

2 **V**ery **L**arge **O**nline **P**latform**s** are a legal category for enterprises introduced by the European Union's Digital Services Act (DSA).

tion strategies can produce (Olteanu et al. 2019; Sen et al. 2021), we argue that standards for specifying social media and web data collections are both lacking but urgently needed to improve repeatability and reproducibility. More formal specifications of data collection processes would significantly support researchers in replicating and validating results, as well as in repeating data collections within longitudinal or multi-phase research projects.

Taking an abstract view, the platform-specific connections discussed above can be represented as actor-message-platform networks that rely on URL-based pointers for at least the inter-platform connection$_1$. Consider Figure 1, which illustrates a minimal actor-message network with two edge types spanning two platforms. Beginning the exploration with two known actors—actor$_1$ and actor$_2$—we retrieve the messages authored by these actors via their platform-specific accounts: account$_1$ and account$_3$ for actor$_1$ and account$_4$ for actor$_2$. Actor$_1$ has posted messages on two platforms, platform$_1$ and platform$_2$, whereas actor$_2$ has posted only on platform$_2$. Based on the message texts, we identify two additional actors—actor$_3$ and actor$_4$—who are interacted with through these messages.

Figure 1: Example of an actor-message network in which actors interact with other actors via platform-specific accounts. Messages mediate these interactions and may link to messages on other platforms



Further, the simplified actor-message network illustrates how a network-based sampling of the networked public sphere can be conducted. Given a seed set—the actors near the top of Figure 1 as inputs to the process—it is necessary to specify accounts on a per platform basis. Using a rule set that defines how to interpret message content and metadata, these instructions are executed for each observed message associated with each account. In this way, relationships between messages and additional entities within the networked public sphere are identified—applied to our example, this yields two further accounts for which the same process can be repeated.

To illustrate our argument, we outline a research scenario that highlights both the challenges and potential solutions. For the sake of conciseness, we focus on possible linkages between two platforms; however, the process described here can be extended to additional platforms. We use Telegram and YouTube as example cases, as both are highly relevant platforms for political communication but have not yet been examined as extensively in terms of their structural characteristics as platforms such as Twitter.

Telegram's publicly viewable entities include channels, which can function either as one-to-many broadcasting venues or many-to-many forums with potentially thousands of participants. Replies to individual posts, reactions, and other interactive features can be enabled by the channel's owners or moderators[3]. From this, two relevant and distinct entity types emerge: channels and messages. Channels are controlled by one or more users, and when multiple moderators are active, each message is signed with the corresponding username. The message's text can contain a variety of marked-up references, such as hashtags or hyperlinks. Information—in the form of messages—can be easily amplified within Telegram, as channels can repost messages from other channels. This reposting process replicates the original message within the new channel's context while providing a backlink to the original message in the source channel.

Within YouTube, entities exhibit similar interactions, with channels able to post videos, though they lack the ability to repost content. Between the two platforms, directed connections exist from Telegram to YouTube, as hyperlinks embedded in Telegram messages point to YouTube resources. These connections can reference either a specific video or a channel (see Figure 2 (a)).

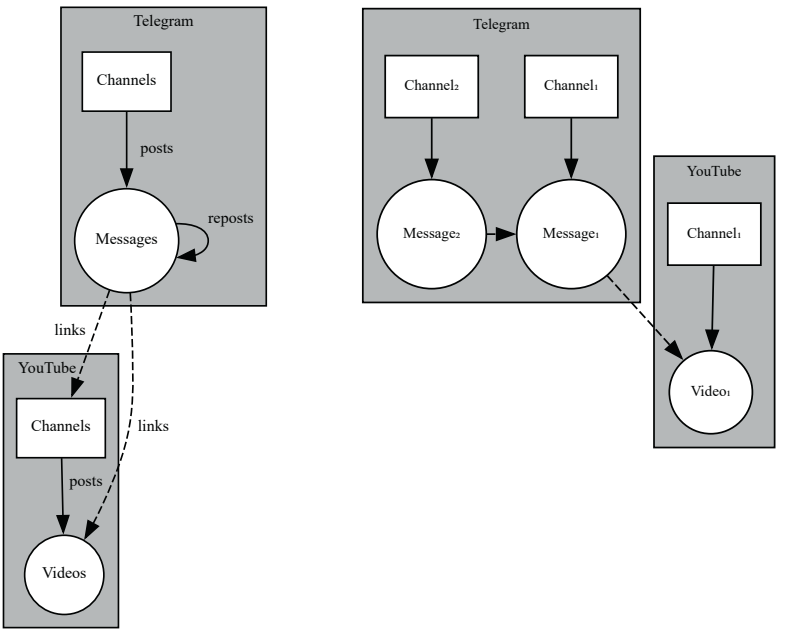
For linkages tracing the diffusion of information within Telegram, we use resharing connections captured in the trace data. For example, as shown in Figure 2 (b): a Telegram channel_1 posts a message_1 which is subsequently amplified by channel_2. Channel_2 thereby generates a message_2 which links back to message_1. Message_1 also contains a reference to a YouTube video_1 hosted on YouTube channel_1. In this way, Telegram channel_1 amplifies information from YouTube channel_1 and is, in turn, amplified by channel_2. Zooming out from this minimal example, since channels have more than one message and introducing more channels into our considerations, reveals a latent network that connects actors, accounts or channels on specific platforms and content objects which allow for a near infinite amount of connections.

3 For more information on Telegram channels, see https://core.telegram.org/api/channel.

Figure 2: Structural diagrams of possible scenarios of information diffusion within and between the two platforms, Telegram and YouTube, allow for a virtually infinite number of connections

(a) Schematic representation of possible interactions between object types on both Telegram and YouTube.

(b) Example of linkages occurring within and between the two platforms.



5. The Network Sampling Framework

Integrating networked public sphere theory (Section 2), explorative network sampling approaches (Section 3), and the practical realities of cross-platform data acquisition (Section 4), the proposed framework represents *actors* as nodes connected through their messages—following Friemel and Neuberger (2023)—via productive, receptive, or curative links. Nodes are necessarily URL-addressable objects, such as profile pages or a blog roll associated with an actor, and are connected to other nodes that also represent actors. Connections are established through features of communication outputs, including feeds, messages, postings, videos, and other metadata extracted from the respective platforms. The interpretation of these objects is flexible and can be adapted to the researcher's specific use case; for example, one could extract links from a message by identifying mentioned accounts or by querying the platform's followers API endpoint. Accordingly, the framework is designed to be modular and adaptable to different combinations of platforms and their intrinsic connection and object types.

Given this model, the realities of cross-platform data acquisition and the complex yet sparse nature of social networks require a network generation process in which edges and

nodes are selected exploratorily, sampling the neighborhoods of the original seed nodes while remaining as agnostic as possible about the data source. This process can be generalized into the following sequence of steps, executed by four modules (see Table 1 for an overview of the generalized steps):

1. A sampling round (or hop) begins by identifying the seed nodes. For the first hop these nodes must be provided by the researcher. The seed nodes are assigned to a network layer—for example, accounts on BlueSky or channels on YouTube.
2. Each layer is associated with a *connector* that retrieves the desired information for its nodes.
3. Retrieved information is stored in a database and then processed by the *parser* for the layer. Parsing generates edges, which may connect nodes within the same layer or across different layers.
4. Edges are filtered according to researcher-defined criteria, such as including only messages containing a specific keyword or written in a particular language.
5. A new set of seed nodes is generated by evaluating the sampled network using a *sampler* assigned to each layer.

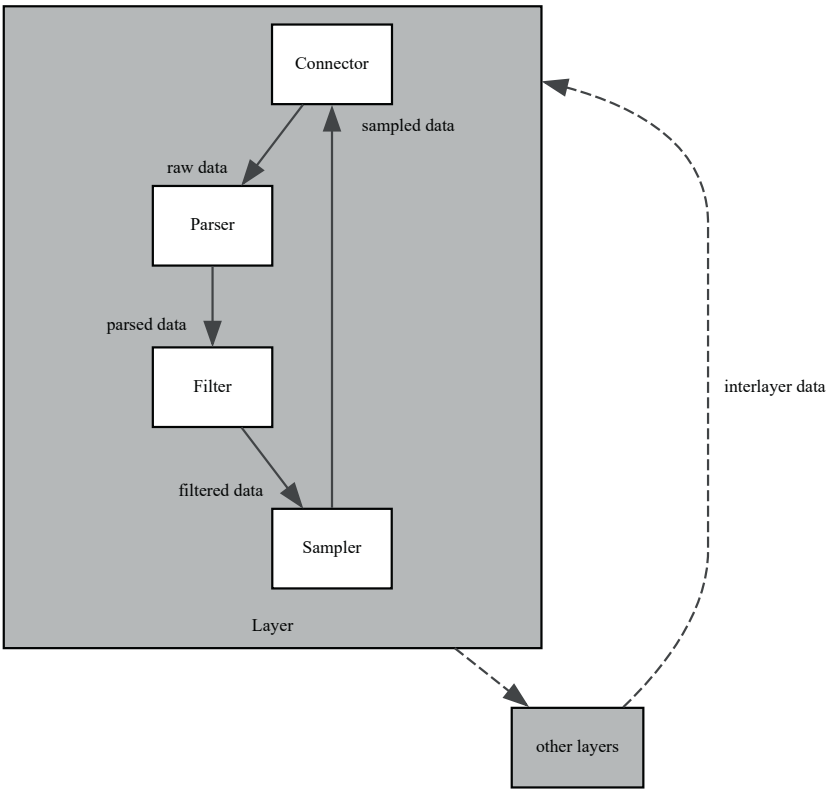
This process is repeated until a stopping condition is met, such as achieving a predetermined sample size for each layer.

Table 1: At the top-level, the sampling framework consists of the following modules

Module	Attributes	Description
Connector	Data connection specification	Describes a program or template for retrieving data for a resource: It enables the collection of information on both actors and content, for example, by scraping public Telegram channels, accessing BlueSky's API, or loading local data files.
Parser	Edge Rules Node Rules	The data returned by the connector and stored in the network layer is processed using a specified rule set, generating nodes and edges from the raw data.
Filter	Edge Rules Node Rules	Rejects specific edges or nodes prior to sampling based on defined conditions. Filtering can be applied to the network's topology, metadata of actors or content, or the content itself—for example, by language.
Sampler	Sampling Algorithm Specification	For each inter- and intralayer—or for the network as a whole—an algorithm is specified to select nodes and edges for sampling. Examples include snowball, rank-degree, forest-fire, or probability-based approaches, depending on metadata or structural properties of the network (cf. Spikyball sampling).

Seeds are the set of nodes from which a network exploration process begins. For each layer, the desired node identifiers are provided to initiate explorative sampling. The format of these identifiers depends on the type of *connector* used; for example, exploring a follower network or account timelines requires account handles or IDs, whereas other connectors may accept content IDs or hashtags. Accordingly, each entry in the database must be mapped to an account handle or ID, as specified in the retrieval configuration (cf. Listing 1).

Figure 3: Exemplary sampling process for a single layer



Listing 1: Example of a multi-layer seed definition containing accounts from two German broadcast programs

```
seeds:
  tiktok:
    - 'DW News': dwnews
    - 'ZDF Heute': zdfheute
  twitter:
    - 'DW News': dwnews
    - 'ZDF Heute':ZDFheute
  facebook:
    - 'DW News': deutschewellenews
    - 'ZDF Heute':ZDFheute
  instagram:
    - 'DW News': dwnews
    - 'ZDF Heute': zdfheute
```

5.1 Connector

We refer to logical data sources as connectors, which encapsulate API endpoints, web-scraping routines, or direct access to databases or files. The input to a connector is always a node identifier, and its output is processed by the *parser* to generate edges, which are then added to the corresponding network layer along node metadata.

Introducing this abstraction as a common, unified interface enhances the stability of data collection processes. Since APIs evolve over time, the *abstraction* allows the underlying implementation to adapt without affecting the interface. The same principle applies to web scraping, where code often breaks due to website changes. In some cases, fallback implementations may be desirable—for example, switching from an API to scraping if needed. Similarly, databases or data files can be queried directly, facilitating the integration of existing datasets into ongoing retrieval tasks.

Formulating relevant connections is inherently platform-dependent. For example, in a database of public speakers, account information may exist for four platforms: Facebook, Instagram, Twitter/X, and TikTok. A connector to the TikTok Research API could access the endpoint that reports which other channels a given channels follows, and return both account information (such as subscriber- and like-count) and connection information—in this case, a list of channels followed by the given channel. Another connector could retrieve the network in the opposite direction, reporting followers, while a third could collect the account's post timeline. Similarly, other connectors can wrap API endpoints from additional platforms, databases or files[4].

Table 2: Input and output listing of the connector module

Input	Output
Identifier: A valid identifier for the node to be retrieved, such as a username or numeric ID, depending on the necessities of the implementation. **Table specifications**: A named set of process specifications, where each specification defines a process that accepts the identifier and returns a list of **records**. These processes can be scripts, programs, or other executable routines. **Common attributes**: Additional parameters applied to all processes listed in the table specifications. Examples include a function to retrieve an OAuth token or commonly used parameters such as timeouts, page sizes, or hard limits.	A keyed data collection, where the keys correspond to the table names, and values contain the retrieved records.

4 As an example, consider an implementation that wraps around a Python package for the TikTok Research API https://github.com/Leibniz-HBI/spiderexpress-tiktok.

Listing 2: Example of a connector defined for the TikTok Research API. It accesses the follower endpoint, retrieving both channel information and account connections, and returns these data for further processing

```
layer:
  tiktok:
    connector:
      infos:
        type: request
        endpoint: https://open.tiktokapis.com/v2/research/user/info/
        request_body_template: '{"user_name": "$node_name"}'
        method: POST
      connections:
        type: request
          endpoint: https://open.tiktokapis.com/v2/research/user/fol-
lowing/
        request_body_template: '{"user_name": "$node_name"}'
        method: POST
```

5.2 Parser

As noted above, URLs themselves do not carry intrinsic meaning—hyperlinks have no inherent semantics. However, because the networked public sphere is highly centralized from a platformization perspective, platform-specific rules can be useful. For example, posts on X.com follow a predictable URL schema: https://x.com/$username/status/$post_id. Similar rules can be defined for other platforms, as their technical systems generally produce predictable hyperlink structures.

Leveraging this predictability, relevant identifiers of posts or accounts can be extracted from URLs that reference these resources. A minimal syntax for extracting the necessary identifiers from hyperlinks can be derived from the following steps: select a field from the datum's content or metadata, access the specified field, optionally apply a regular expression with a capture group, and assign the resulting edge the appropriate edge type. Using the identifier of the processed resource and the extracted reference(s), edges can then be formed between the corresponding nodes.

Similarly, node metadata can be extracted from trace data. For example, a *connector* (cf. Section 5.1) that retrieves both account information and the account's social connections provides two types of data. The account information typically contains fields that can be directly mapped within the framework. Social connections can be extracted and mapped in a comparable manner, allowing them to be represented as edges in the network.

Table 3: Input and output listing of the parser module

Inputs	Output
Data: A keyed dictionary in which each key corresponds to a list of records. **Edge rule set**: A list of edge extraction specifications. Each specification defines which objects and fields to access for both the source and the target of an edge, as well as any additional metadata to extract. Every specification must	A keyed data collection in which the keys correspond to table names and the values contain the retrieved records.

Inputs	Output
specify an edge type and the layers in which the source and target nodes reside. **Node rule set**: A list of node extraction specifications, where each specification maps a field from the data object to the node metadata within the network.	

Listing 3: Example of parser rules that process the data gathered by the connector described in the previous example, extracting both edge and node information for inclusion in the network

```
parser:
  edges:
    source: $node_name
    target:
      type: follow
      field: connections.data.[].name
      regex: null
    type: repost
  node:
    display_name:
      field: infos.display_name
      regex: null
    subscriber_count:
      field: infos.suscriber_count
      regex: null
```

5.3 Filter

Depending on the use case, it may be necessary to reject portions of the network parsed in the previous step. For example, certain actors could be excluded for data protection or privacy reasons, such as by setting a minimum subscriber or follower count for inclusion. It may also be useful to filter posts containing specific keywords to maintain topical relevance, or to select content in a particular language to focus on a single "language sphere."

Table 4: Inputs and outputs of the filter module

Inputs	Outputs
Network: Node and edge data as produced by the **parser**. **Edge/node rule set**: Expressions evaluated using the network as an input.	The network with edges and nodes removed according to the evaluation of the specified rule sets.

5.4 Sampler

This module processes the *filtered* network output and determines the seed set for the next sampling hop. It can also use the previous sampler state as input when employing a stateful strategy, for example, to avoid revisiting nodes or edges that have already been sampled.

As discussed in Section 3, the choice of sampling strategy affects exploration patterns, and different strategies are suited to different use cases (see Hu and Lau 2013 for an overview). Examples include Snowball sampling (Goodman 1961), Forest Fire (Leskovec and Faloutsos 2006), Rank-degree sampling, which ranks nodes based on known degrees in the original network (Voudigari et al. 2016; Münch et al. 2021), and Spikyball, a generalization of multiple strategies that leverages various platform metadata for probability-based sampling (Ricaud, Aspert, and Miz 2020).

Table 5: Input and output listing of the sampler module

Inputs	Outputs
Network: Node and edge data from the relevant layer(s). **Sampler specification**: Defines the strategy to use and, optionally, any parameters required by the strategy. **Sampler state:** Optional state information, used by stateful strategies to avoid revisiting nodes or to manage multi-hop walks.	A new seed set, as well as the sampled network, if the sampling algorithm is stateful, the updated sampler state.

5.5 Limitations

The proposed framework assumes a multilayer network structure, which requires matching nodes across layers. Consequently, it operates at the actor-level, where each actor may control multiple accounts across multiple platforms. This approach is feasible when using a known set of actors with a pre-established mapping to their accounts. However, when combined with explorative network sampling, it becomes necessary to (automatically) match accounts both within a single platform and across platforms, assigning them to individual actors. Research methods for this task are still in early stages, such as mining and clustering accounts based on behavioral characteristics (Bruns et al. 2025).

A further limitation is that using URL-based network sampling of digital resources inherently creates a strong dependence on the network structure of the underlying spherules. If certain digital resources are not connected to the portion of the hidden network where the sampling begins, they cannot be discovered, meaning some remain entirely undetected. Nevertheless, content that reaches a sufficient level of virality is likely to be captured, and for most relevant networks, the majority of nodes belong to a so-called giant component.

Another limitation of focusing on URLs is that information diffusion can bypass these unique identifiers. For example, users often share screenshots of posts or repost content, breaking the chain of observable links. This limitation adds to the inherent multimodality of contemporary social media and web platforms, which cannot be captured by URL-based networks alone. However, it is possible to enrich the discovered networks post-hoc with semantic interlinking, such as tracing paraphrased texts or tracking the diffusion of memes.

On the technical side, the proposed approach—and web scraping in general—can be undermined if platforms stop assigning URLs to content or limit their usefulness. Without accessible URLs, profiles and content objects cannot be reached from the open Internet. For example, Meta allows WhatsApp channels to have an invite link (a URL), but the channel contents are not publicly accessible. Nevertheless, not exposing content or user profiles via URLs is a significant drawback for platforms, as it prevents important content from being referenced externally, whether in media publications or by search engines.

Lastly, the current diversification of platforms, particularly the Fediverse, presents unique challenges. Content can flow between platforms without relying on URLs to reference other posts. However, this content is not uniformly distributed across all instances, as some instances can block others. This situation complicates the assessment of platform affordances, since content can exist independently of any single platform, and individual platforms may implement widely varying policies and technical approaches.

6. Conclusions and Outlook

In this article, we proposed a networked-based, exploratory sampling framework for analyzing networked public spheres. Building on theoretical foundations from media and communication studies, as well as legal and technical considerations, we developed a process that enables researchers to capture and sample data from multiple sources in a well-documented, interoperable, and repeatable manner.

Our main contribution lies enhancing documentability, (inter-)operationalizability, and repeatability. By structuring the data collection process into the four proposed modules (Connector, Parser, Filter, and Sampler, see Table 1) and their associated rulesets, researchers can not only describe their data collection precisely and reproducibly but also obtain guidance for building their own module implementations. These implementations can interface seamlessly with others, reducing dependence on a single data source. We developed this framework alongside an experimental example implementation as a Python command-line tool, which is publicly available and actively maintained[5].

The proposed framework can be further extended to address the limitations discussed in Section 5.5. For example, a new module could be added to automatically match accounts to actors based on topological, metadata, or behavioral features. Likewise, additional modules and extensions are both possible and desirable to enhance the framework's functionality and adaptability.

A question that remains open—and that will inform both substantive research and the further development of this framework—is under which circumstances content posted on one platform crosses over to other platforms or sites, becoming linked from external sources. This process is naturally influenced by the affordances of both the source and target platforms: for example, YouTube videos are frequently linked from other platforms, reflecting YouTube's role as a content search engine rather than solely a social network. In contrast, linking Instagram posts or videos has only recently begun to gain traction. Additionally, cultural and language-specific practices shape platform affordances, resulting in differences in how platforms are used across national and linguistic contexts.

Finally, the sampling algorithms and their suitability for specific research tasks must be carefully evaluated. For instance, some research questions may require more localized data, where smaller nodes are prioritized before bridging communities through well-connected, larger nodes, necessitating different algorithms than those used to study the macrostructure of platform figurations. Another promising avenue for multidisciplinary research is the development of sampling strategies for multilayer networks, which consider not only individual layers or inter-layer connections but also the multilayer network as a whole from a global perspective.

With these questions in mind, we hope this contribution will support empirical research on digital publics and their interconnectedness, enhancing scalability, scope, repeatability, and longitudinal observability. The distribution and diffusion of content across multiple

5 https://github.com/Leibniz-HBI/spiderexpress.

platforms is a research area of growing importance today and is likely to become even more critical in the near future.

References

Barabasi, Albert-László, and Reka Albert. 1999. Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Berners-Lee, Tim. 1994. "Universal Resource Identifiers in WWW: A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network as Used in the World-Wide Web." Request for {{Comments}} RFC 1630. Internet Engineering Task Force. https://doi.org/10.17487/RFC1630

Berry, David. 2011. The computational turn: Thinking about the digital humanities. *CULTURE MACHINE*, *12*.

boyd, danah m., and Nicole B. Ellison. 2007. "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer-Mediated Communication* 13 (1): 210–30. https://doi.org/10.1111/j.1083-6101.2007.00393.x.

Bossetta, Michael. 2019. *The Digital Architectures of Social Media: Comparing Political Campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 U.S. Election* (No. arXiv:1904.07333). arXiv. https://doi.org/10.48550/arXiv.1904.07333

Breiter, Andreas, and Andreas Hepp. 2018. "The Complexity of Datafication: Putting Digital Traces in Context." In *Communicative Figurations: Transforming Communications in Times of Deep Mediatization*, edited by Andreas Hepp, Andreas Breiter, and Uwe Hasebrink, 387–405. Transforming Communications – Studies in Cross-Media Research. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-65584-0_16

Brin, Sergey, Rajeev Motwani, Lawrence Page, and Terry Winograd. 1998. "What Can You Do with a Web in Your Pocket?" *IEEE Data Eng. Bull.* 21 (2): 37–47.

Brüggemann, Michael, and Hendrik Meyer. 2023. "When Debates Break Apart: Discursive Polarization as a Multi-Dimensional Divergence Emerging in and Through Communication." *Communication Theory* 33 (2–3): 132–42. https://doi.org/10.1093/ct/qtad012

Bruns, Axel. 2008. "Life Beyond the Public Sphere: Towards a Networked Model for Political Deliberation." *Information Polity* 13: 71–85.

Bruns, Axel. 2019. After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, *22*(11), 1544–1566. https://doi.org/10.1080/1369118X.2019.1637447

Bruns, Axel. 2023. "From 'the' Public Sphere to a Network of Publics: Towards an Empirically Founded Model of Contemporary Public Communication Spaces." *Communication Theory* 33 (2–3): 70–81. https://doi.org/10.1093/ct/qtad007

Bruns, Axel, and Jean Burgess. 2011. "The Use of Twitter Hashtags in the Formation of Ad Hoc Publics." In *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*, edited by A. Bruns and P. De Wilde, 1–9. United Kingdom: The European Consortium for Political Research (ECPR).

Bruns, Axel, Kateryna Kasianenko, Vish Padinjaredath Suresh, Ehsan Dehghan, and Laura Vodden. 2025. Untangling the Furball: A Practice Mapping Approach to the Analysis of Multimodal Interactions in Social Networks. *Social Media + Society*, *11*(2). https://doi.org/10.1177/20563051251331748

Bruns, Axel, and Brenda Moon. 2019. "One Day in the Life of a National Twittersphere." *Nordicom Review* 40 (s1): 11–30. https://doi.org/10.2478/nor-2019-0011

Bruns, Axel, Brenda Moon, Felix Victor Münch, and Troy Sadkowsky. 2017. "The Australian Twittersphere in 2016: Mapping the Follower/Followee Network." *Social Media + Society, 3*(4). https://doi.org/10.1177/2056305117748162.

"Common Crawl – Overview." n.d. https://commoncrawl.org/overview. Accessed January 28, 2026.

Coscia, Michele, and Luca Rossi. 2018. "Benchmarking API Costs of Network Sampling Strategies." In *2018 IEEE International Conference on Big Data (Big Data)*, 663–72. https://doi.org/10.1109/BigData.2018.8622486

Esau, Katharina, Tariq Choucair, Samantha Vilkins, Sebastian F. K. Svegaard, Axel Bruns, Kate S. O'Connor-Farfan, and Carly Lubicz-Zaorski. 2024. "Destructive Polarization in Digital Communication Contexts: A Critical Review and Conceptual Framework." *Information, Communication & Society* 0 (0): 1–22. https://doi.org/10.1080/1369118X.2024.2413127

Freelon, Deen. 2021. "The Post-API Age Reconsidered: Web Science in the '20s and Beyond." In *13th ACM Web Science Conference 2021*, 3–3. Virtual Event United Kingdom: ACM. https://doi.org/10.1145/3447535.3466177.

Friemel, Thomas N, and Christoph Neuberger. 2023. "The Public Sphere as a Dynamic Network." *Communication Theory* 33 (2–3): 92–101. https://doi.org/10.1093/ct/qtad003

Frischlich, Lena, & Edda Humprecht (2021). *Trust, Democratic Resilience, and the Infodemic*. https://doi.org/10.5167/UZH-202660

Goodman, Leo A. 1961. "Snowball Sampling." *The Annals of Mathematical Statistics* 32 (1): 148–70. https://doi.org/10.1214/aoms/1177705148

Guan, Lu, Xiao Fan Liu, Wujiu Sun, Hai Liang, and Jonathan Zhu. 2022. "Census of Twitter Users: Scraping and Describing the National Network of South Korea." *PLOS ONE* 17 (November): e0277549. https://doi.org/10.1371/journal.pone.0277549

Habermas, Jürgen. 1962. *Strukturwandel der Öffentlichkeit – Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*. 1990th ed. Suhrkamp.

Hasebrink, Uwe, and Andreas Hepp. 2017. "How to Research Cross-Media Practices? Investigating Media Repertoires and Media Ensembles." *Convergence, 23*(4): 362–77. https://doi.org/10.1177/1354856517700384

Heft, Annett, Kilian Buehling, Xixuan Zhang, Juni Schindler, and Miriam Milzner. 2024. Challenges of and Approaches to Data Collection across Platforms and Time: Conspiracy-Related Digital Traces as Examples of Political Contention. *Journal of Information Technology & Politics*, *21*(3), 323–339. https://doi.org/10.1080/19331681.2023.2250779

Helmond, Anne. 2015. "The Platformization of the Web: Making Web Data Platform Ready." *Social Media + Society* 1 (2): 205630511560308. https://doi.org/10.1177/2056305115603080

Hepp, Andreas, and Uwe Hasebrink. 2014. Kommunikative Figurationen – ein Ansatz zur Analyse der Transformation mediatisierter Gesellschaften und Kulturen. In *Von der Gutenberg-Galaxis zur Google-Galaxis: Alte und neue Grenzvermessungen nach 50 Jahren DGPuK* (pp. 343–360). UVK Verlagsgesellschaft.

Hu, Pili, and Wing Cheong Lau. 2013. "A Survey and Taxonomy of Graph Sampling." arXiv. https://doi.org/10.48550/arXiv.1308.5865

Jost, Pablo, Annett Heft, Kilian Buehling, Maximilian Zehring, Heidi Schulze, Hendrik Bitzmann, and Emese Domahidi. 2023. "Mapping a Dark Space: Challenges in Sampling and Classifying Non-Institutionalized Actors on Telegram." *M&K Medien & Kommunikationswissenschaft* 71(3–4): 212–29. https://doi.org/10.5771/1615-634X-2023-3-4-212

Kivelä, Mikko, Alex Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. 2014. "Multilayer Networks." *Journal of Complex Networks*, 2(3): 203–71. https://doi.org/10.1093/comnet/cnu016

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy and Marshall van Alstyne. 2009. *Computational Social Science. Science, 323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Lazer, David, Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, et al. 2020. "Computational Social Science: Obstacles and Opportunities." *Science* 369 (6507): 1060–62. https://doi.org/10.1126/science.aaz8170

Leskovec, Jure, and Christos Faloutsos. 2006. "Sampling from Large Graphs." In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 631–36. Philadelphia PA USA: ACM. https://doi.org/10.1145/1150402.1150479

Münch, Felix Victor, Ben Thies, Cornelius Puschmann, and Axel Bruns. 2021. "Walking Through Twitter: Sampling a Language-Based Follow Network of Influential Twitter Accounts." *Social Media + Society, 7*(1): 2056305120984475. https://doi.org/10.1177/2056305120984475

Nielsen, Hendrik, Roy T. Fielding, and Tim Berners-Lee (1996). Hypertext Transfer Protocol – HTTP/1.0 (Request for Comments No. RFC 1945). Internet Engineering Task Force. https://doi.org/10.17487/RFC1945

Ohme, Jakob, Theo Araujo, Laura Boeschoten, Deen Freelon, Nilam Ram, Byron B. Reeves, and Thomas N. Robinson. 2023. "Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking." *Communication Methods and Measures, 0* (0): 1–18. https://doi.org/10.1080/19312458.2023.2181319

Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Emre Kıcıman. 2019. "Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries." *Frontiers in Big Data* 2.

Pearce, Wareen, Suay M. Özkula, Amanda K. Greene, Lauren Teeling, Jennifer S. Bansard, Janna Joceli Omena, and Elaine Teixeira Rabello. 2020. Visual Cross-Platform Analysis: Digital Methods to Research Social Media Images. *Information, Communication & Society*, *23*(2), 161–180. https://doi.org/10.1080/1369118X.2018.1486871

Quandt, Thorsten, Lena Frischlich, Svenja Boberg, and Tim Schatto-Eckrodt. 2019. Fake News. In *The International Encyclopedia of Journalism Studies* (pp. 1–6). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118841570.iejs0128

Rau, Jan Philipp, Philipp Kessling, Gregor Wiedemann, and Felix Victor Münch. 2025. "Research Data Access in the Context of Art. 40 DSA for the German Federal Election: A Mixed Experience at Best." Frankfurt, Leizig. https://doi.org/10.58079/13ZE0

Ricaud, Benjamin, Nicolas Aspert, and Volodymyr Miz. 2020. "Spikyball Sampling: Exploring Large Networks via an Inhomogeneous Filtered Diffusion." arXiv. https://doi.org/10.48550/arXiv.2010.11786.

Rogers, Richard 1996. "The Future of Science and Technology Studies on the Web." *EASST Review* 15, 25–27.

Rogers, Richard. 2010. "Mapping Public Web Space with the Issuecrawler." In *Digital Cognitive Technologies: Epistemology and Knowledge Society*, edited by Claire Brossard and Bernard Rebers. London, England: Wiley.

Rogers, Richard. 2023. "'Serious Queries' and 'Editorial Epistemologies'." In *The Propagation of Misinformation in Social Media: A Cross-platform Analysis*. Amsterdam University Press. https://doi.org/10.5117/9789463720762

Schmidt, Jan-Hinrik. 2014. "Twitter and the Rise of Personal Publics." In *Twitter and Society*, 3–14. New York, Washington, D.C., Bern.

Schmidt, Jan-Hinrik, Lisa Merten, Uwe Hasebrink, Isabelle Petrich, and Amelie Rolfs (2017). Zur Relevanz von Online-Intermediären für die Meinungsbildung. Arbeitspapiere des Hans-Bredow-Instituts, 40, 107 S. https://doi.org/10.21241/SSOAR.71784

Sen, Indira, Fabian Flöck, Katrin Weller, Bernd Weiß, and Claudia Wagner. 2021. "A Total Error Framework for Digital Traces of Human Behavior on Online Platforms." *Public Opinion Quarterly* 85 (S1): 399–422. https://doi.org/10.1093/poq/nfab018

Strippel, Christian, Annekatrin Bock, Christian Katzenbach, Merja Mahrt, Lisa Merten, Christian Nuernbergk, Christian Pentzold, Cornelius Puschmann, and Annie Waldherr. 2018. "Die Zukunft der Kommunikationswissenschaft ist schon da, sie ist nur ungleich verteilt: Eine Kollektivreplik auf Beiträge im „Forum" (Publizistik, Heft 3 und 4, 2016)." *Publizistik* 63, 11–27 (Januar). https://doi.org/10.1007/s11616-017-0398-5.

Voudigari, Elli, Nikos Salamanos, Theodore Papageorgiou, and Emmanuel J. Yannakoudakis. 2016. "Rank Degree: An Efficient Algorithm for Graph Sampling." *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, 120–29. https://doi.org/10.1109/ASONAM.2016.7752223

Waldherr, Annie. 2017. "Öffentlichkeit als komplexes System. Theoretischer Entwurf und methodische Konsequenzen." *M&K Medien & Kommunikationswissenschaft,* 65(3): 534–49. https://doi.org/10.5771/1615-634X-2017-3-534

Wiedemann, Gregor, Felix Victor Münch, Jan Philipp Rau, Phillip Kessling, and Jan-Hinrik Schmidt. 2023. "Concept and Challenges of a Social Media Observatory as a DIY Research Infrastructure." *Publizistik*, 201–223 August. https://doi.org/10.1007/s11616-023-00807-6.

Wolf, J. L., Squillante, M. S., Yu, P. S., Sethuraman, J., & Ozsen, L. (2002, May 7). *Optimal Crawling Strategies for Web Search Engines*. WWW2002.

Unpacking Translation Effects

Influences of Target Language Choice on Topic Modeling in Multilingual Environments

Nadezhda Ozornina / Mario Haim*

Machine translation is widely used in communication science to consolidate texts, not least for exploratory clustering approaches such as multilingual topic modeling. However, the impact of target language choice on topic modeling results remains unclear. This study examines these effects by (a) consolidating texts into one of the original document languages and (b) translating texts into an intermediary language not present in the dataset under study. To assess the effects, we use a corpus of parallel United Nations texts in Russian and German (N = 3,760). We compare the results of structural topic modeling after translating Russian texts into German, chosen as the original language, with consolidating the entire corpus into English as an intermediary language. The translation approaches are compared based on feature overlap, topical prevalence, and topical content. The findings show that intermediary-language translation yields a more symmetrical topic distribution and higher overlap in top words, but significantly reduces vocabulary size compared to consolidation into the original language. The results are replicated using a second bilingual journalistic corpus (N = 434) and validated across different numbers of topics. Finally, we discuss best practices for target language selection in multilingual topic modeling and situate them within the context of recent developments in computational communication science.

Key words: topic modeling, machine translation, multilingual text analysis, German, Russian, methods, computational communication science, computational social science

1. Introduction

With the increasing volume of journalistic and social media data, exploratory automated text analysis methods such as topic modeling, have gained prominence in communication science (Günther, 2022; van Atteveldt et al., 2022). Unsupervised approaches to clustering textual content have become especially valuable for studying multilingual contexts, as discussions on digital platforms increasingly take place in multiple languages simultaneously (Hase et al., 2021; Maier et al., 2022). In such cases, the ability to identify cross-case topics is crucial for gaining a comprehensive understanding of the issues under study and for facilitating quantitative comparison between cases (Lind et al., 2022). However, because topic modeling algorithms are typically designed for monolingual text collections, they often struggle to detect similar meanings across languages due to vocabulary misalignment (Chan et al., 2020). This poses a significant challenge for comparative analyses of large-scale textual data and cross-case comparability.

To enable the application of topic modeling in multilingual contexts, machine translation techniques combined with probabilistic topic modeling are commonly employed. Previous studies in communication science have demonstrated the ability of this approach to

* Nadezhda Ozornina, M.A., Ludwig-Maximilians-Universität München, Department of Media and Communication, Akademiestr. 7, 80799 München, Germany, Nadezhda.Ozornina@ifkw.lmu.de, https://orcid.org/0000-0002-8637-9394
Prof. Dr. Mario Haim, Ludwig-Maximilians-Universität München, Department of Media and Communication, Akademiestr. 7, 80799 München, Germany, haim@ifkw.lmu.de, https://orcid.org/0000-0002-0643-2299.

DOI: 10.5771/1615-634X-2026-1-71

produce valid outputs for exploratory analyses when applied to texts in multiple languages (de Vries et al., 2018) and when using different translation tools (Reber, 2019; Licht et al., 2024). However, it remains unclear whether the choice of target language for machine translation impacts the outcomes of multilingual topic modeling and to what extent such effects are robust across different types of texts (Lucas et al., 2015; Maier et al., 2022). Understanding these potential effects is crucial for assessing the potential and limitations of machine translation for topic modeling and for developing best practices to ensure validity in multilingual exploratory text analysis.

This study investigates the impact of target language choice on topic modeling outcomes in multilingual text collections by comparing two strategies for target language selection. Specifically, it examines (a) consolidation into one of the original languages already present in the dataset and (b) translation into an intermediary language not included in the original text corpus. In doing so, our findings aim to advance methodological discussions on validation in automated text analysis (Baden et al., 2022) and to offer recommendations in the context of recent developments in machine translation and computational methods.

To provide a case study based on different language groups and media systems, the research employs Russian and German versions of official United Nations documents (N = 3,760). We apply structural topic modeling (STM, Roberts et al., 2019) after consolidating the texts into German, chosen as the original language, and compare the results with topic models based on translations into English as intermediary language. Drawing on previous evaluation strategies (de Vries et al., 2018; Reber, 2019), we assess the impact of translation on topic modeling results in three ways: feature overlap in document-feature matrices (DFMs), topical prevalence, and topical content. To ensure the broader applicability of our findings to communication science and to examine their robustness in more linguistically diverse settings, we replicate the analysis using a second, substantially different, bilingual journalistic corpus (N = 434). For both text types, the results are further validated across different numbers of topics.

2. Theoretical Background

This section provides background on multilingual topic modeling and the role of machine translation in enhancing its applicability across languages. Building on the discussion of how target language choice shapes topic modeling outcomes, we then develop the research questions.

2.1 Multilingual Topic Modeling

As an approach to unsupervised text analysis, topic modeling aims to uncover structures in large text collections by analyzing patterns of word co-occurrences (Blei, 2012; Jacobi et al., 2016). In standard probabilistic approaches, a generative model uses the words of a textual corpus to group documents into a predefined number of clusters while maximizing inter-cluster differentiation. These clusters have been shown to resemble latent topics consisting of words and documents. This probabilistic clustering is optimized using two matrices: one representing the distribution of topics across individual documents, indicating each topic's prevalence within a document, and the other representing the distribution of words across topics, indicating each word's prevalence within a topic (e.g., Haim, 2023; van Atteveldt et al., 2022). More recently, embedding-based approaches using models explicitly trained on multilingual corpora, such as BERTopic (Grootendorst, 2022), have also been employed to explore textual data by clustering feature-vector representations of multilingual content.

In communication science, topic modeling's ability to represent documents as mixtures of latent topics and to characterize these topics through words makes it a useful tool for exploring concepts such as frames, issues, and writing styles in large-scale textual data (Günther, 2022). Given the growing volume of multilingual content in digital environments, applying topic modeling algorithms to linguistically diverse contexts is becoming increasingly important. In such analyses, identifying cross-case topics provides a deeper understanding of the concepts under study and facilitates quantitative comparisons across different countries and research settings (Lucas et al., 2015; Hase et al., 2021).

Despite the growing relevance of case comparisons in multilingual text analysis (Lind et al., 2022), it remains unclear how current approaches to topic modeling can be validly applied to linguistically diverse settings. This challenge is particularly pronounced for probabilistic models, in which the identification of cross-case topics is hindered by the so-called "Babel problem" (Chan et al., 2020). This problem refers to the fact that similar concepts are represented by different words in different languages, which bag-of-words algorithms originally designed for monolingual data are unable to reconcile. Consequently, when probabilistic topic modeling is applied to multilingual data without prior consolidation, documents are likely to be grouped by language rather than by content, hindering the identification of topics across cases in multilingual text collections (for an example, see Lind et al., 2022). Applying monolingual embeddings to multilingual data poses related challenges, as vector representations trained on a single language cannot be reliably transferred to other linguistic contexts.

2.2 Mitigating Bias Through Machine Translation

One common approach to facilitate cross-case comparison for multilingual data is the use of machine translation algorithms (Lucas et al., 2015). In this process, the collected text corpus is first divided into subcorpora corresponding to the original languages, and each subcorpus is then consolidated into the chosen target language before further analysis. Machine translation is typically performed using tools such as DeepL or Google Translate (Reber, 2019), and the resulting translations are subsequently used as input for probabilistic topic modeling.

Nowadays, machine translation is a widely applied strategy that offers both advantages and disadvantages compared to other consolidation approaches, such as manual translation, multilingual dictionaries, or pre-training based on multilingual documents (see overviews in Lind et al., 2022 and Maier et al., 2022). In particular, it enables more comprehensive analyses than techniques based on multilingual dictionaries, which may be limited in capturing the full range of relevant meanings (Maier et al., 2022). Additionally, machine translation provides a faster and more cost-effective solution, requiring fewer resources than custom training of multilingual models based on parallel data (Lind et al., 2022). However, the reproducibility of machine translation is constrained by the "black-box" nature of existing algorithms, which are constantly updated (Chan et al., 2020), making researchers dependent on third-party tools and platforms. Moreover, translation costs can increase significantly with the volume of text, making it an expensive option for consolidating large multilingual corpora (Lucas et al., 2015; Reber, 2019). One recent approach to overcoming these limitations is the use of open-source translation tools such as OPUS-MT, which offer a more affordable and reproducible alternative to commercial providers (Licht et al., 2024).

With the advent of large language models, clustering based on multilingual embeddings has been explored as an alternative to combining machine translation with probabilistic topic modeling. Recent studies (e.g., Licht, 2023; Licht & Lind, 2023) demonstrate the capability of embedding models such as multilingual BERT to produce high-quality cross-

lingual outputs, making them an attractive strategy for aligning multilingual data without additional translation efforts. In this approach, the translation effort can be effectively frontloaded and integrated into large language models that are explicitly trained on multiple languages. However, unlike probabilistic topic modeling, multilingual embeddings may rely on partially transparent and comprehensible models, thus limiting researcher control, interpretability, and validation of produced outputs (Licht & Lind, 2023; Rinke et al., 2022).

In our study, we focus on the broader applicability of machine translation for probabilistic topic modeling. Here, key concerns relate to the quality of machine translation as input for automated text analysis. First, results may contain systematic errors in the translation of frequently occurring words, which can alter the representation of topics in the original data and assign the same concepts from different languages to separate topics (Lucas et al., 2015). Despite recent improvements in translation system quality (Chan et al., 2020), such errors still warrant consideration in automated analyses. Second, research in computational linguistics shows that, as languages vary in vocabulary richness and linguistic structure, machine translation may oversimplify the meaning of the original texts, leading to vocabulary loss, overgeneralization, and the omission of language-specific meanings and concepts across different language pairs (Kotait, 2024; Vanmassenhove et al., 2019). Together with systematic errors, these issues risk of the intended topical structure being "lost in translation" (de Vries et al., 2018), limiting the extent to which translated texts reflect the distribution and content of topics in the original corpus. Further investigation is needed to determine how prevalent these issues are when applying current translation systems for text consolidation.

2.3 Influences of Target Language Choice

Previous studies have examined the applicability of machine translation in various research settings, comparing texts written in different languages (de Vries et al., 2018), evaluating different translation tools (Reber, 2019; Licht et al., 2024), and assessing translation outcomes at the level of full texts versus DFMs (e.g., Lucas et al., 2015). Overall, findings suggest that machine translation provides valid inputs for multilingual topic modeling, with only minor translation errors and slight changes in topic prevalence and top words (for an overview, see Appendix A in supplementary materials at https://osf.io/qfch3). However, studies also indicate that translation quality may vary depending on the strategies employed and may perform differently across datasets and research settings (Maier et al., 2022; Reber, 2019), an aspect that has not been thoroughly explored.

Despite the general effectiveness of machine translation for topic modeling, there is a suggestion that translation quality issues may vary depending on the selected target language for text consolidation (Wang et al., 2022). Therefore, the choice of language into which texts are translated may influence the topic modeling outcomes (Lucas et al., 2015). However, the potential impact of target language choice has not yet been empirically evaluated, and English is commonly used for translation without consideration of possible effects on the meanings encoded in the original texts. Given English's status as a "lingua franca" in communication science (Baden et al., 2022; Lind et al., 2022), it is crucial to understand the limitations its usage may impose, particularly when consolidating texts from different languages. Moreover, the consequences of selecting alternative target languages for machine translation remain unclear in current research (Licht et al., 2024).

Building on previous discussions of target language choice (Koltsova & Pashakhin, 2020; Lucas et al., 2015), this study proposes two approaches to machine translation in multilingual text analysis, based on its relationship between the target language and the

languages present in the analyzed corpus. These strategies are illustrated using corpora consisting of texts in two languages.

The first strategy involves consolidating the texts into one of the two languages already present in the corpus, referred to as the *original language*. For example, if the dataset includes Russian and Ukrainian publications, researchers may select one of the corpus languages as the basis for text consolidation and translate the other language into it—for instance, translating the Ukrainian texts into Russian (Koltsova & Pashakhin, 2020). In this approach, translation is applied to only one part of the corpus, conserving translation resources and minimizing the loss of the original vocabulary. At the same time, this strategy introduces varying degrees of transformation across subcorpora, potentially compromising methodological equivalence between cases at the input stage of analysis (Baden et al., 2022; Licht & Lind, 2023). Because some portions of the corpus are translated while others remain unchanged, there is a risk of systematic bias favoring the untranslated texts in the resulting topic model (see Lucas et al., 2015). However, the effect of such "unequal" data consolidation on topic modeling outcomes has not yet been empirically evaluated.

An alternative strategy involves translating all documents into an *intermediary language* that is not present in the original dataset and serves as an external consolidation basis for multilingual text analysis. An example of this approach is provided by Lucas et al. (2015), who translated social media posts from Arabic and Chinese into English to facilitate a "symmetrical" transformation of both subcorpora. This consolidation strategy ensures that both sets of texts undergo an equal degree of machine translation, enabling cross-case equivalence at the input stage for topic modeling. While this approach mitigates the scenario in which one subcorpus remains untranslated, it requires more translation resources and results in a greater loss of the original text vocabulary. In addition, the outcomes of this strategy for different language pairs may systematically differ in quality due to varying degrees of linguistic similarity among the languages involved (Wang et al., 2022).

For well-resourced language pairs, such as English and German, machine translation quality may be higher than for less closely related languages, where original texts may undergo greater simplification and translations may contain more errors due to difficulties in conveying meaning across differences in morphology, scripts, and sentence structure. Recent experiments in computational linguistics illustrate this loss of linguistic richness, for example when translating Arabic texts into English (Kotait, 2024). Consequently, the need to force original texts "into the corset of English-like language structure" (Baden et al., 2022) highlights the limitations of using English as the target language in comparative settings and underscores the importance of a more detailed examination of the influence of target language choice on topic modeling.

Overall, the impact of target language choice on topic modeling results remains unclear. Although some considerations in communication science suggest that using an intermediary language may provide a more symmetrical transformation, whereas consolidating into one of the original languages may better preserve vocabulary, these assumptions have not been sufficiently examined. Therefore, we pose the following research question:

RQ1: How do topic modeling results of a multilingual corpus differ when (a) consolidating into one of the original languages versus (b) translating into an intermediary language?

2.4 Replicability Across Text Types

In addition to examining different translation strategies, there is a need to evaluate results across various settings, an aspect that has been rarely explored in previous studies on the

impact of machine translation on topic modeling. A particularly important consideration is the replicability of results across text types with varying linguistic variability, which can significantly influence translation quality and, consequently, topic modeling outcomes (Maier et al., 2022). Current methodological research on multilingual topic modeling is often conducted using official documents (e.g., de Vries et al., 2018), leaving it unclear whether these findings generalize to other text types commonly used in communication science. One such type is journalistic publications, which tend to be more diverse and metaphorical in language compared to official documents, making them more susceptible to translation errors (Maier et al., 2022). This leads to the second research question:

RQ2: How does machine translation impact topic modeling results when applied to official documents versus journalistic publications?

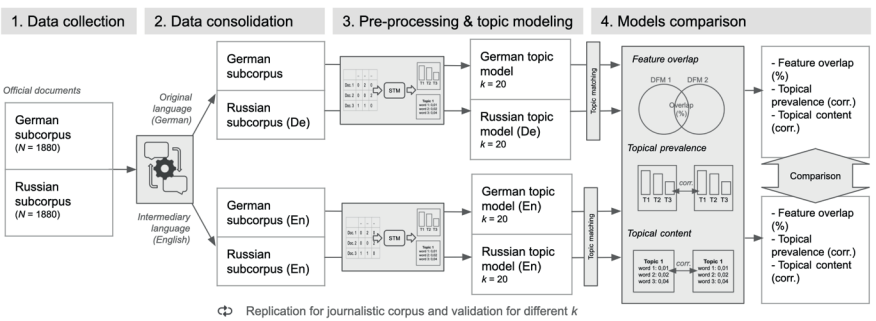
3. Methodology

In the following section, we describe the study design and methodological procedures applied. Reproducible scripts for each step are available in the supplementary materials of the publication.

3.1 Study Design and Case Selection

To investigate the impact of machine translation, we calculate and compare topic models after translating texts into different target languages across multiple text types (see Figure 1). For this study, we focus on official documents and journalistic publications written in Russian and German, allowing us to examine potential translation issues across different language branches (Fortson, 2011). This language selection also ensures sufficient data availability for both text types and serves as a prototypical case study for comparative research, representing differences in journalistic cultures and media systems (Hallin & Mancini, 2011; Hanitzsch et al., 2019). We illustrate consolidation into an original language by translating Russian texts into German and implement the intermediary language strategy by consolidating all texts into English.

Figure 1: Study Design



3.2 Data Collection

To assess the influence of target language choice on topic modeling results, we use a collection of openly available parallel texts from United Nations official documents (Eisele & Chen, 2010) in both Russian and German (N = 3,760), published as the part of the OPUS project (Tiedemann, 2012). Although no longer fully up to date, this corpus contains a sufficient number of UN resolutions in multiple languages and has proven its usefulness in methodological studies on multilingual text analysis (e.g., Windsor et al., 2019).

To enable comparison across text types, we replicate the UN corpus analysis using a corpus of journalistic publications (N = 434). These texts were obtained through web scraping of content on culture, lifestyle, and politics from various Russian media sources, supplemented with professional translations into German (for replication results, see Appendix C). Unlike UN publications, these journalistic texts are less standardized, incorporate more language-specific terms, and include much more recent data. For each text type, the corpus of parallel publications is divided into subcorpora written in Russian and German.

3.3 Machine Translation

For UN publications, text consolidation was performed using Google Translate, as it required fewer resources while providing quality comparable to other translation tools (Reber, 2019). The translation was carried out in May 2023 using the browser-based document translator. To implement consolidation into the original language, Russian UN documents were translated into German, while the German texts remained unchanged. For intermediary language strategy, both Russian and German texts were translated into English. The same procedure was applied to the second corpus of journalistic articles.

3.4 Pre-processing

As a result of applying both translation strategies to the UN publications, we obtained four subcorpora: original German, Russian translated into German, German translated into English, and Russian translated into English. Each subcorpus contained 1,880 documents and was preprocessed by removing URLs, numbers, punctuation marks, symbols, and language-specific stopwords with *spacyr* (Benoit & Matsuo, 2017). These steps are standard practice in topic modeling pre-processing and help to focus on meaningful features while removing noise for further analysis (Maier et al., 2018; van Atteveldt et al., 2022).

Next, we performed lemmatization using language-specific models chosen for their quality and computational efficiency (*en_core_web_sm* and *de_core_news_sm*). The outcomes of this step were evaluated using the type–token ratio (*TTR*) before and after lemmatization. *TTR* is a widely used linguistic metric for measuring lexical diversity, calculated as the ratio of unique tokens to the total number of tokens, with values closer to one indicating greater language variability (see Kettunen, 2014). We observed a similar reduction in *TTR* across the subcorpora, with German showing higher mean *TTR* values than English (after lemmatization: 0.549 vs. 0.444), consistent with findings from previous studies (Bentz & Kiela, 2014). These results indicate that the lemmatization process was appropriate for both languages, producing neither too many nor too few tokens relative to the original content.

Then, we applied vocabulary pruning of lemmatized data using common thresholds, removing terms that appeared in less than 0.5 % or more than 99 % of documents (Reber, 2019). At this stage, the four resulting subcorpora (original German, Russian translated into German, German translated into English, and Russian translated into English) were prepared for analysis and converted into separate DFMs containing between 4,373 and 8,397 features.

The same pre-processing steps were applied to the journalistic publications, resulting in four DFMs, each containing 217 documents (N features: 6,611–9,989). We also performed the same lemmatization checks using *TTR*, which produced results similar to those observed for the UN documents (mean *TTR* after lemmatization: 0.737 for German and 0.635 for English). These results further confirmed the higher linguistic variability of journalistic data compared to UN publications.

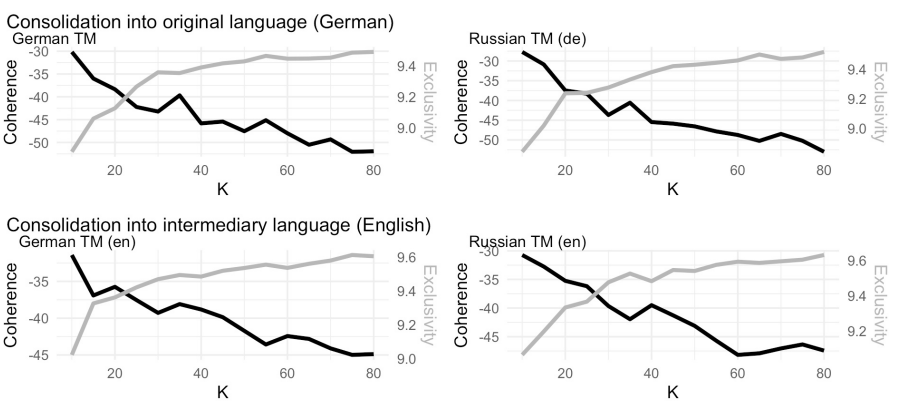
3.5 Topic Modeling

In the next step, we applied Structural Topic Modeling (STM; Roberts et al., 2019). STM has become a widely used technique for topic modeling in the social sciences, as it allows for the inclusion of covariates and is readily accessible through an *R* package. For each DFM derived from the UN corpus, we computed topic models with topic numbers ranging from 10 to 80 with incremens of 5. For each topic solution under each consolidation strategy, we automatically matched topics from the texts originally written in German and Russian using the method proposed by de Vries et al. (2018). In this process, each word was considered individually, and the topics from both models with the highest loading for that word were paired. Matching topics were then identified based on the frequency of such pairings.

To enable a more comprehensive analysis and labeling of topic models derived from the UN corpus, we focused on the model with $k = 20$ topics. This number was chosen based on statistical evaluations of coherence and exclusivity (Figure 2), which are widely applied in current studies (Bernhard-Harrer et al., 2025). Additionally, we manually examined and labeled the resulting topic matches based on their top words, which are provided in Appendix B. To ensure that the chosen topic solution was not unique, the model comparisons described in the following steps were conducted for all previously computed topic solutions, allowing the findings to be validated across different topic numbers.

For the journalistic corpus, we computed topic models with 5 to 40 topics to validate the results across different topic configurations. For detailed analysis, we focused on the solution with $k = 15$, reflecting the smaller size of the dataset and based on coherence and exclusivity metrics. The corresponding figure with statistical metrics and the main results for the journalistic publications is provided in Appendix C.

Figure 2: Coherence and Exclusivity for Different k (UN-Corpus)



3.6 Model Comparison

Following de Vries et al. (2018), differences between the translation strategies were evaluated using three metrics: feature overlap between DFMs, topical prevalence, and topical content. These metrics were calculated separately for each translation strategy, and the results were then compared to assess which target language produced more similar topic models.

Feature overlap assesses whether the translated texts share similar vocabulary by measuring the proportion of overlapping lemmas in the DFMs before the application of topic modeling. Higher overlap indicates greater similarity in the vocabularies of the consolidated texts. We also calculated the *TTR* to examine whether the observed differences reflect variations in vocabulary richness.

Topical prevalence measures the correlation of topic distributions across documents in the translated subcorpora originally written in German and Russian. This metric helps determine whether matched topics maintain similar proportions across the datasets. We additionally examined document-level topical prevalence, following previous studies (de Vries et al., 2018; Reber, 2019). These results were consistent with the corpus-level findings and are therefore not described in detail in the main text. While examining these outcomes (see Appendix D) could help identify documents driving translation misalignment, this aspect is beyond the scope of the current study, which focuses on general quantitative evaluation.

Finally, topical content quantifies the correlation of word distributions across topics from different subcorpora. This metric allows us to determine whether matched topics use the same words with similar frequency.

4. Results

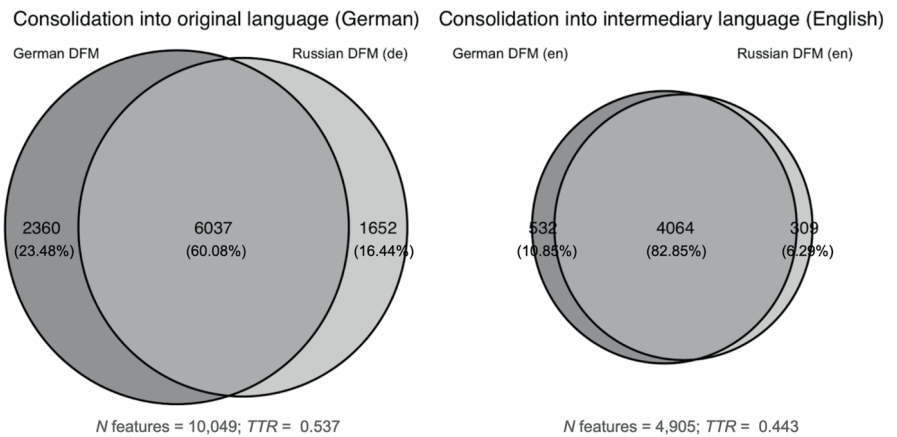
This section first presents the results for the corpus of official UN documents (k = 20). Subsequently, the results of the replication using journalistic publications are discussed in more detail (4.4).

4.1 Feature Overlap

Overall, consolidating UN publications into an intermediary language results in a higher degree of vocabulary overlap between the resulting subcorpora. When German and Russian texts are translated into English, more than 80 % of all DFM features are shared between the two translated versions (Figure 3, right-hand side). In contrast, when German texts remain unchanged and Russian texts are translated into German, just over half of the features (60.08 %) are shared between the resulting DFMs after pre-processing (Figure 3, left-hand side).

Another notable observation is that consolidating texts into the original language results in a considerably larger vocabulary in the data used for topic modeling. When texts are consolidated into German, the combined DFMs contain N = 10,049 unique lemmas—more than twice as the number found when both subcorpora are translated into English (N = 4,905), despite originating from the same collection of documents. The *TTR* values for models based on consolidation into German are also generally higher (0.54) compared to English translations (0.45). These results indicate that, in our case, translation into an intermediary language reduces vocabulary diversity more than the strategy of retaining part of the corpus in its original language for consolidation. Possible consequences and the generalizability of this issue are presented in the discussion section.

Figure 3: Unique DFM-Features and Their Overlap (UN-Corpus)



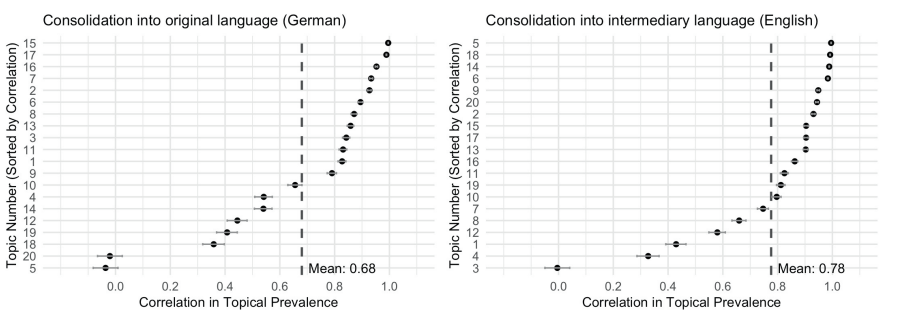
Note. The percentage is calculated based on the total number of features in the respective consolidation strategy.

4.2 Topical Prevalence

To further examine the outcomes at the level of topic models, we compared the translation strategies based on topical prevalence. For the UN documents, translating into an intermediary language produced a more similar distribution of topics across subcorpora than consolidation into an original language.

In particular, topics show a higher correlation in distribution across the translated Russian and German subcorpora when the texts are consolidated into English (M = 0.78, Figure 4, right-hand side) compared to consolidation into German (M = 0.68, Figure 4, left-hand side). Furthermore, in the English versions of the topic models, more than half of the matched topics have a correlation greater than 0.9, whereas in the German models, this is true for only five of twenty topic matches. Because we are working with parallel data, these results suggest that the intermediary language strategy is more likely to align similar topics in their distribution than consolidation into the original language.

Figure 4: Correlation in Topical Prevalence (UN-Corpus)

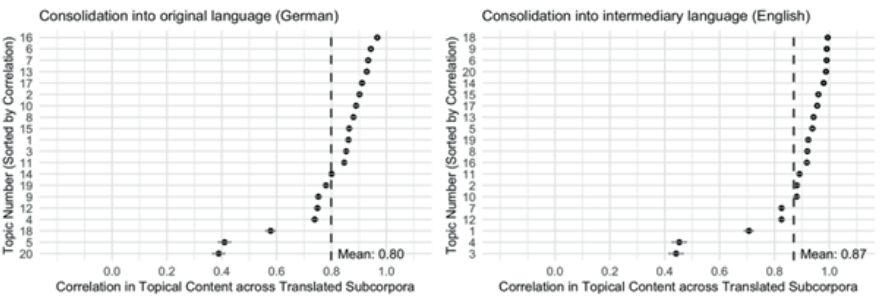


4.3 Topical Content

The results for topical content are consistent with the previous findings. We observe a higher level of similarity between the translated subcorpora in representing matched topics through words when consolidating the texts into the intermediary language.

For models based on translations into English (Figure 5, right-hand side), the correlations of word distributions for most matched topics range from 0.83 and 0.99, with a mean value of 0.87. For consolidation into German (Figure 5, left-hand side), the correlations are slightly lower (M = 0.80), though still indicating strong alignment. This suggests that translation into an intermediary language tends to produce topics using the same words with more similar frequencies compared to consolidation based on one of the original languages. Nevertheless, for both strategies, consolidation led to the generation of largely similar representations of parallel corpora in terms of topical content.

Figure 5: Correlation in Topical Content (UN-Corpus)



4.4 Replication Across Text Types

Comparing the results for UN publications with those for journalistic data, we find that the trends observed in official documents are also present in the journalistic corpus for k = 15 topics (see Table 1). First, for journalistic articles, intermediary-language translation results in greater DFM overlap (64.7 %) compared to consolidation into an original language (49.4 %), while reducing the number of features and *TTR* values used for topic modeling. Second, English-language translations exhibit a more symmetrical topic distribution (M = 0.42) and higher overlap in top words (M = 0.75), compared to consolidation into German (M = 0.25 for topical prevalence; M = 0.67 for topical content).

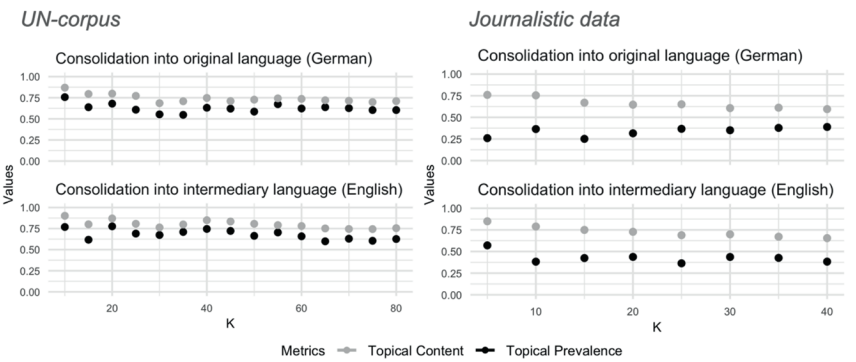
It is also evident that, across all three metrics, the similarities between models are lower for journalistic publications than for official documents. This is particularly noticeable in the mean correlation coefficients for topical prevalence, which do not exceed 0.5 for the journalistic corpus but range between 0.68 and 0.78 for UN publications. This suggests that, on average, matched topics are more similarly distributed in topic models of official documents, reflecting a more coherent vocabulary in the political discourse of UN resolutions. In contrast, the journalistic corpus contains a larger number of features and exhibits higher *TTR* values, indicating greater linguistic variability in journalistic articles despite the smaller corpus size.

Table 1: Results of Comparison Across Text Types

	Official documents *(k = 20)*		**Journalistic publications *(k = 15)***	
	Consolidation into original language (German)	Consolidation into intermediary language (English)	Consolidation into original language (German)	Consolidation into intermediary language (English)
Feature overlap	Overlap = 60.1% *N* Feat. = 10,049 *TTR* = 0.537	Overlap = 82.9% *N* Feat. = 4,905 *TTR* = 0.448	Overlap = 49.4% *N* Feat. = 12,967 *TTR* = 0.715	Overlap = 64.7% *N* Feat. = 8,211 *TTR* = 0.629
Topical Prevalence	Mean Corr. = 0.68	Mean Corr. = 0.78	Mean Corr. = 0.25	Mean Corr. = 0.42
Topical Content	Mean Corr. = 0.80	Mean Corr. = 0.87	Mean Corr. = 0.67	Mean Corr. = 0.75

These outcomes are consistent across different numbers of topics (*k*) for both text types, further highlighting the robustness of the findings (Figure 6). As shown, topical content and topical prevalence remain at similar levels within each text type, although the values are consistently lower for journalistic publications.

Figure 6: Mean Correlation Coefficients across Different Number of Topics



5. Discussion and Best Practices

For RQ1, we find that intermediary-language translation results in greater DFM feature overlap, a more symmetrical topic distribution, and higher overlap in top words compared to consolidation into an original language. This aligns with previous research (Lucas et al., 2015) and confirms that translating the entire corpus into an intermediary facilitates a more balanced representation of the analyzed data, as all subcorpora undergo the same level of content transformation.

However, translating texts into an intermediary language (English) results in a less diverse representation of features compared to direct consolidation into an original language (German). This reflects a noticeable vocabulary simplification in texts processed through

machine translation and supports previous findings (Kotait, 2024; Vanmassenhove et al., 2019), extending them to the new language pairs containing Russian, English, and German. The reason for this overgeneralization may be that, during the training of neural translation models such as Google Translate, frequent translations are reinforced while rare translations get suppressed, particularly when translating into a language with lower linguistic variability (Kettunen, 2014). Consequently, outputs from neural translation systems tend to be less diverse, potentially losing meaningful aspects and culture-specific items in subsequent analysis.

In the context of multilingual topic modeling, our statistical evaluation suggests that this simplification does not significantly impact model outcomes, likely because this method of automated text analysis is inherently focused on providing generalized representations of textual data. However, such simplification could be a critical concern for other types of automated analyses where individual word translations are important—for instance, in studies examining emotions across languages or categorizing text based on specific linguistic features and word components (see Windsor et al., 2019).

Regarding RQ2, we find that the impact of machine translation on journalistic publications follows patterns similar to those observed for official UN documents. For both text types, translation into an intermediary language leads to higher vocabulary overlap, greater similarity in topic prevalence, and more consistent topical content. These results demonstrate the applicability of our findings across different text types and provide additional replication for studies on multilingual analysis in communication science.

Nevertheless, it is important to note that similarities between the models from the Russian and German corpora are lower across all three metrics in the journalistic data. This may be due to both the smaller corpus size and the higher linguistic variability of journalistic texts (Maier et al., 2022). To investigate which features in the journalistic corpus might contribute to these discrepancies, we went through the top words for topics with prevalence and content correlations below the mean and examined their contextual usage.

The journalistic articles used in our study contain a broader range of lexical features than UN documents and include more context-specific, error-prone terms that can affect translation accuracy and lead to inconsistencies. For example, proper names and locations are often transcribed differently during corpus consolidation (e.g., "*Kirgistan*" in an original German text versus "*Kirgisistan*" in the translation from Russian into German), reducing vocabulary overlap between models. Vocabulary simplification is also evident, as context-specific, traditional, or slang terms may lack direct equivalents and are therefore replaced with simpler expressions. For instance, the Russian word *Khorovod*, referring to a traditional circle dance, is translated merely as "dance around" in English, losing its cultural and traditional significance. However, upon manual inspection of the top words, such issues were found to be relatively rare.

For future research, the findings of this study suggest that both target-language strategies can provide valid inputs for multilingual topic modeling. However, one strategy may be more or less suitable depending on the specific case, research objectives, and the languages present in the analyzed corpora.

Consolidation into the original language can be a useful strategy when studying cases in which the languages are closely related or belong to the same branch, such as Slavic or Nordic languages (Koltsova & Pashakhin, 2020; Licht et al., 2024). In such instances, translating into one of the languages already present in the corpus allows researchers to save translation resources while preserving the vocabulary richness of the data. This approach enables a more detailed examination and comprehensive representation of the analyzed

texts, including language-specific characteristics and original meanings that could be lost or blurred when translating into an intermediary language.

If the languages belong to different language groups, such as Chinese and Arabic (Lucas et al., 2015), consolidation into English as an intermediary language for automated analysis may be more appropriate. Here, English would serve as a common bridge between languages due to extensive training resources, making it a practical choice for intermediary translation and enabling a more symmetrical transformation of textual data. This strategy also allows researchers to achieve more aligned representations in terms of DFMs, topical prevalence, and topical content. However, it is important to keep in mind that translating into an intermediary language can reduce vocabulary richness and may omit certain details or nuances of meaning from the original data, which should be considered when interpreting the results of automated analysis.

Regardless of the strategy employed, it is important to ensure the quality of machine translation and to validate its outcomes. In research, this could be achieved through various approaches. For instance, one could manually examine translated documents and the top words in the topic models (see Lucas et al., 2015) to identify potential translation errors that could affect topic representations. Particular attention should be given to topics that appear prevalent only in one of the compared subcorpora. In such cases, it is essential to determine whether the topic is genuinely unique to that subcorpus or whether translation issues contributed to this outcome. Such input validation is especially important when analyzing journalistic publications and other text types relevant to communication science (e.g., social media posts; Maier et al., 2022), which often contain context-specific information and figurative language.

Beyond machine translation, the results of other text transformation steps applied prior to analysis should also be validated, as some words may be processed differently by standard pre-processing algorithms in different languages. For example, in our study, the name of the Russian social media platform *VKontakte* was lemmatized according to German verb rules, producing the non-existent word "*vkontaken*" and causing vocabulary misalignment in our topic models. However, we have not examined this issue systematically and relied only on the top words from topic models, so we can only suggest that such issues are rare. Nevertheless, we encourage researchers to consider the implications of target language choice and to address the potential limitations of machine translation in future studies.

6. Conclusion

Overall, the study demonstrates that both translation strategies examined provide a valid basis for multilingual topic modeling. Machine translation is therefore particularly valuable for analyzing large-scale data in the digital environments of communication science, enabling the identification of cross-case topics and supporting comparative analyses (Lind et al., 2022). Our comparison of strategies indicates that intermediary-language translation leads to greater overlap in features, topic distributions, and top words, but substantially reduces vocabulary diversity compared to consolidation into an original language. This pattern is also observed in journalistic publications, which exhibit higher linguistic diversity and underscore the need for careful validation of machine translations.

This study extends existing research on the impact of machine translation on topic modeling outcomes (de Vries et al., 2018; Reber, 2019; Maier et al., 2022) by examining the role of different translation strategies in automated text analysis. It also provides best practices for selecting an appropriate consolidation approach for large-scale text collections and validating the results of machine translation. More broadly, these findings highlight that thorough evaluation of analytical steps related to language alignment or cross-lingual con-

cept identification is crucial for establishing the validity of applied methods. For instance, this is particularly relevant for approaches based on outputs from generative models, which may overgeneralize interpretations of culture-specific phenomena across languages during training, forcing them into English-like patterns (Yoo et al., 2025).

The study has several limitations. First, only two languages (German and Russian) were analyzed, and in the original-language strategy, translation was conducted only into German. Expanding the study to include additional languages would help further assess the robustness of our findings. Second, the analysis is based on two text types derived from relatively small and partly outdated corpora. Future research could enhance generalizability by incorporating social media data and using larger, more diverse content. Finally, the study does not include detailed qualitative examination of resulting topics, which could provide deeper insights for future analyses and applies a limited, partly imprecise procedure of topic matching (de Vries et al., 2018). More comprehensive qualitative analysis and systematic examination of documents with low topic alignment could illuminate patterns in topic prevalence and clarify potential sources of translation-related issues.

Looking ahead, several aspects warrant further investigation. In particular, more modern approaches to automated analysis, such as word embeddings and transformer models (see Licht et al., 2024), should be explored in greater detail. These methods show promise due to their ability to incorporate contextual information from the textual data. However, they also present limitations, including the need for technical expertise, the potential for biases, higher computational demands, and challenges in interpretability due to the indirect relationship between input data and model outputs (Licht & Lind, 2023). Furthermore, while our study found that machine translation had minimal impact on topic modeling results, it would be valuable to examine its effects on other text analysis methods in greater depth (Windsor et al., 2019). Finally, as online environments become increasingly multimodal, future research could investigate text-image approaches to topic modeling (e.g. Torres, 2024) to assess their applicability in comparative settings.

References

Baden, C., Pipal, C., Schoonvelde, M., & Van Der Velden, M. A. C. G. (2022). Three Gaps in Computational Text Analysis Methods for Social Sciences: A Research Agenda. *Communication Methods and Measures*, *16*(1), 1–18. https://doi.org/10.1080/19312458.2021.2015574

Benoit, K., & Matsuo, A. (2017). *spacyr: Wrapper to the "spacy" "NLP" library* (1.2.1) [Software]. https://CRAN.R-project.org/package=spacyr [28.01.2026].

Bentz, C., & Kiela, D. (2014). *Measuring and Modelling Lexical Diversity across Languages* [Conference presentation]. 5th UK Cognitive Linguistics Conference, Lancaster, UK.

Bernhard-Harrer, J., Ashour, R., Eberl, J. M., Tolochko, P., & Boomgaarden, H. (2025). Beyond Standardization: A Comprehensive Review of Topic Modeling Validation Methods for Computational Social Science Research. *Political Science Research and Methods*, 1–19. https://doi.org/10.1017/psrm.2025.10008

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Chan, C.-H., Zeng, J., Wessler, H., Jungblut, M., Welbers, K., Bajjalieh, J. W., van Atteveldt, W., & Althaus, S. L. (2020). Reproducible Extraction of Cross-lingual Topics (rectr). *Communication Methods and Measures*, *14*(4), 285–305. https://doi.org/10.1080/19312458.2020.1812555

de Vries, E., Schoonvelde, M., & Schumacher, G. (2018). No Longer Lost in Translation: Evidence that Google Translate Works for Comparative Bag-of-Words Text Applications. *Political Analysis*, *26*(4), 417–430. https://doi.org/10.1017/pan.2018.26

Eisele, A., & Chen, Y. (2010). MultiUN: A Multilingual Corpus from United Nation Documents. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, 2868–2872.

Fortson, B. W. IV. (2011). *Indo-European Language and Culture: An Introduction* (2nd ed.). John Wiley & Sons. https://www.wiley.com/en-us/Indo-European+Language+and+Culture%3A+An+Introduction%2C+2nd+Edition-p-9781405188968 [28.01.2026].

Grootendorst, M. (2022). BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. arXiv. https://arxiv.org/pdf/2203.05794

Günther, E. (2022). *Topic Modeling: Algorithmische Themenkonzepte in Gegenstand und Methodik der Kommunikationswissenschaft.* [Topic Modeling: Algorithmic Topic Concepts in the Subject and Methodology of Communication Science.] Herbert von Halem Verlag. https://www.halem-verlag.de/produkt/topic-modeling/

Hallin, D. C., & Mancini, P. (Eds.). (2011). *Comparing Media Systems Beyond the Western World* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139005098

Haim, M. (2023). *Computational Communication Science: Eine Einführung.* [Computational Communication Science: Introduction.] Springer VS. https://link.springer.com/book/10.1007/978-3-658-40171-9

Hanitzsch, T., Hanusch, F., Ramaprasad, J., & De Beer, A. S. (Eds.). (2019). *Worlds of Journalism: Journalistic Cultures Around the Globe*. Columbia University Press. https://doi.org/10.7312/hani18642

Hase, V., Mahl, D., Schäfer, M. S., & Keller, T. R. (2021). Climate Change in News Media across the Globe: An Automated Analysis of Issue Attention and Themes in Climate Change Coverage in 10 Countries (2006–2018). *Global Environmental Change*, *70*, 102353. https://doi.org/10.1016/j.gloenvcha.2021.102353

Jacobi, C., van Atteveldt, W., & Welbers, K. (2016). Quantitative Analysis of Large Amounts of Journalistic Texts Using Topic Modelling. *Digital Journalism*, *4*(1), 89–106. https://doi.org/10.1080/21670811.2015.1093271

Kettunen, K. (2014). Can Type-Token Ratio be Used to Show Morphological Complexity of Languages? *Journal of Quantitative Linguistics*, *21*(3), 223–245. https://doi.org/10.1080/09296174.2014.911506

Koltsova, O., & Pashakhin, S. (2020). Agenda Divergence in a Developing conflict: Quantitative Evidence from Ukrainian and Russian TV Newsfeeds. *Media, War & Conflict*, *13*(3), 237–257. https://doi.org/10.1177/1750635219829876

Kotait, R. (2024). Richness Lost in Machine Translationese. *The Egyptian Journal of Language Engineering*, *11*(1), 66–85. https://doi.org/10.21608/EJLE.2024.267336.1064

Licht, H. (2023). Cross-Lingual Classification of Political Texts Using Multilingual Sentence Embeddings. *Political Analysis*, *31*(3), 366–379. https://doi.org/10.1017/pan.2022.29

Licht, H. & Lind, F. (2023). Going Cross-Lingual: A Guide to Multilingual Text Analysis. *Computational Communication Research*, *5*(2), 1–31. https://doi.org/10.5117/CCR2023.2.2.LICH

Licht, H., Sczepanski, R., Laurer, M., & Bekmuratovna, A. (2024). *No More Cost in Translation: Validating Open-Source Machine Translation for Quantitative Text Analysis.* OSF. https://doi.org/10.31219/osf.io/9trjs

Lind, F., Eberl, J., Eisele, O., Heidenreich, T., Galyga, S., & Boomgaarden, H. G. (2022). Building the Bridge: Topic Modeling for Comparative Research. *Communication Methods and Measures*, *16*(2), 96–114. https://doi.org/10.1080/19312458.2021.1965973

Lucas, C. G., Nielsen, R. A., Roberts, M. E., Stewart, B., Alex, S., & Tingley, D. (2015). Computer-Assisted Text Analysis for Comparative Politics. *Political Analysis*, *23*(2), 254–277. https://doi.org/10.1093/pan/mpu019

Maier, D., Baden, C., Stoltenberg, D., de Vries-Kedem, M., & Waldherr, A. (2022). Machine Translation vs. Multilingual Dictionaries Assessing Two Strategies for the Topic Modeling of Multilingual Text Collections. *Communication Methods and Measures*, *16*(1), 19–38. https://doi.org/10.1080/19312458.2021.1955845

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Reber, U. (2019). Overcoming Language Barriers: Assessing the Potential of Machine Translation and Topic Modeling for the Comparative Analysis of Multilingual Text Corpora. *Communication Methods and Measures*, *13*(2), 102–125. https://doi.org/10.1080/19312458.2018.1555798

Rinke, E. M., Dobbrick, T., Löb, C., Zirn, C., & Wessler, H. (2022). Expert-Informed Topic Models for Document Set Discovery. *Communication Methods and Measures*, *16*(1), 39–58. https://doi.org/10.1080/19312458.2021.1920008

Roberts, M. E., Stewart, B., & Tingley, D. (2019). stm: An R Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2), 1–40. https://doi.org/10.18637/jss.v091.i02

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 2214–2218. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf [28.01.2026].

Torres, M. (2024). A Framework for the Unsupervised and Semi-Supervised Analysis of Visual Frames. *Political Analysis*, *32*(2), 199–220. https://doi.org/10.1017/pan.2023.32

van Atteveldt, W., Trilling, D., & Arcila, C. (2022). *Computational Analysis of Communication*. Wiley Blackwell. https://cssbook.net/ [28.01.2026].

Vanmassenhove, E., Shterionov, D., & Way, A. (2019). Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation. *Proceedings of Machine Translation Summit XVII: Research Track*, 222–232. https://aclanthology.org/W19-6622/ [28.01.2026].

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in Machine Translation. *Engineering*, *18*, 143–153. https://doi.org/10.1016/j.eng.2021.03.023

Windsor, L. C., Cupit, J. G., & Windsor, A. J. (2019). Automated Content Analysis across Six Languages. *PLOS One*, *14*(11), e0224425. https://doi.org/10.1371/journal.pone.0224425

Yoo, M. H., Kim, J., & Song, S. (2025). Multilingual Capabilities of GPT: A Study of Structural Ambiguity. *PloS one, 20*(7), e0326943. https://doi.org/10.1371/journal.pone.0326943

Made, Contested, Invisible – A Critical Data Studies Perspective on Reimagining Methodology in a Datafied Communication Science

Annekatrin Bock*

This paper explores how Critical Data Studies (CDS) and communication studies can enrich each other, particularly in the context of the current uptake of methodological approaches supported by large-scale data analytics and "AI." First, the paper discusses the entanglement of Science and Technology Studies (STS) and CDS as a theoretical and methodological framework. Second, it highlights key assumptions of CDS: (1) the "made-ness" of data, which emphasizes that data are not neutral, but shaped by algorithms and practices; (2) the "interpretative sovereignty over data," which addresses power relations and access barriers in data analysis; and (3) the "(in-)visibility of data," which concerns the de- and recontextualization of data and their (in)visibility in research processes. Third, two methodological examples—data journeys and the walkthrough approach—illustrate how CDS approaches are applied in communication studies. The paper concludes by emphasizing the importance of cooperative collaboration between different scientific disciplines and paradigms to develop innovative and critical methodological approaches capable of addressing the challenges of future-oriented communication studies in a data-infused, algorithmically orchestrated society.

Key words: Datafication, Science and Technology Studies, Critical Data Studies, data journeys, walkthrough approach, socio-technical systems, Retrieval-Augmented Generation (RAG)

1. Introduction

Digital technologies, algorithmic systems, and data-driven practices are no longer peripheral to social life—they are constitutive of it (Esposito, 2024; Mayer-Schönberger & Cukier, 2013). From social media platforms to AI-powered content generation, data are produced, collected, and used in ways that reshape how individuals communicate, how institutions operate, and how knowledge is produced. This transformation is not merely technical; it is deeply social, political, and epistemological. As communication research grapples with these changes, a growing body of scholarship calls for a critical re-engagement with data—not as neutral evidence, but as something *made*, *contested*, and *inherently relational* (boyd & Crawford, 2012; van Dijck, 2014; Zakharova, 2022).

Against this background, communication science research is currently called upon not to replace its established methods, but to (re-)position itself methodologically by critically reflecting on its existing strengths and exploring how they can be reimagined in response to a datafied world. Contemporary research, for example, discusses the application of large language models (LLM) in qualitative research (e.g. Franken, 2022; Wiedemann, 2013; Wollin-Giering et al., 2024). Yet, while many scholars highlight the potential of emerging technologies—such as LLMs, and while the implications of retrieval-augmented generation (RAG) remain under debate—to revolutionize data processing and analysis, this enthusiasm often rests on a reductive view of both technology and method (e.g. Binz et al., 2025; Peters

* Prof. Dr. Annekatrin Bock, Universität Vechta, Medienkompetenzzentrum Vechta, Kolpingstraße 25, 49377 Vechta, Germany, annekatrin.bock@uni-vechta.de, https://orcid.org/0009-0006-9951-4846.

 DOI: 10.5771/1615-634X-2026-1-88

& Chin-Yee, 2025). The assumption that AI tools can "automate" communication research, for instance, overlooks the fact that data are not raw material but a product of specific practices, power relations, and interpretive choices (Crawford et al., 2014; Pink et al., 2022).

This paper argues that a fundamental challenge for communication science lies not merely in adopting new tools, but in rethinking the very foundations of data and methodology, a task for which Critical Data Studies (CDS) and Science and Technology Studies (STS) are well equipped. Both offer promising starting points for empirically grasping, analyzing, and theorizing the (not always entirely new) developments associated with the increasing datafication of communication and communication science (e.g. Dieter et al., 2019; Light et al., 2018; Troeger & Bock, 2022). Informed by STS, which in turn combine insights from information science and cultural studies, CDS have long been concerned with the datafication of societies driven by the availability and processability of digital data (e.g. Lippert & Mewes, 2021; M. T. Schäfer & Van Es, 2017; Schneider et al., 2021). For example, CDS can be applied to analyze data traces (Breiter & Hepp, 2018), data practices (Decuypere, 2021), and data flows (Jarke et al., 2023) in communication science research.

Drawing on the core tenets of CDS—the *made-ness of data*, *interpretative sovereignty over data*, and the *invisibility of data flows*—this contribution reframes the current methodological debate not as a crisis of obsolescence, but as an opportunity for critical innovation. Data is not a given; it is constructed through decisions about what to collect, how to process it, and which interpretations to privilege. Access to data is unevenly distributed, reinforcing existing power imbalances between corporations, institutions, and independent researchers. Crucially, much of data's journey—from collection to analysis to publication—is obscured by what scholars call "blackboxing" (Hillman, 2023; Rogers, 2017).

Against this background, the following argumentation is divided into three parts. First, drawing on an understanding of datafication as a social "meta-trend in the context of digitality" (Breiter & Bock, 2023), the entanglement of datafication, communication, and society are problematized. Second, STS and the core assumptions of contemporary CDS are introduced, and their points of reference for communication studies are outlined. Third, two methodological examples are used to reflect on the meaningful interconnection between CDS and research questions in communication science.

The article concludes with an outlook on implications of a CDS perspective for communication science concerned with datafied communication and its socio-political implications. It frames this perspective as an invitation for researchers to follow data not as a linear path, but as a network of decisions, negotiations, and exclusions—revealing what is made visible, what is ignored, and who holds the power to decide. Far from advocating a retreat from technology, this contribution urges communication science to embrace complexity, uncertainty, and reflexivity. In a world where data is made, access is unequal, and processes are often invisible, the researcher's role is not to automate, but to *interrogate*. Consequently, the future of communication science methodology lies not in faster tools, but in deeper questions.

2. Datafied Society | Communication | Research

"Smart", "personalizable", "data-driven" technologies, along with the interwoven practices of collecting, producing, distributing, and using data, are widely considered to be fundamentally transforming society (Breiter & Bock, 2023; Couldry, 2017; Decuypere et al., 2021; Hintz et al., 2019; Mayer-Schönberger & Cukier, 2013). The collection and utilization of data has a long tradition not only in datafied societies but, in particular, in empirical research (M. T. Schäfer & Van Es, 2017). What is remarkable about contemporary datafication, however, is that data are not only analogue but also digital, available in unprecedented

quantities and varieties (Crawford et al., 2014). As a result, data can be generated and processed in real time and flexibly expanded and scaled across heterogeneous data sets, enabling exhaustive compilation and fine-grained analysis (Kitchin, 2014).

Datafication, understood as a "social meta-trend in the context of digitalization", is therefore a "highly complex social process in which different stakeholders, actor constellations, practices and power structures interact" (Breiter & Bock, 2023). Data-enabled communication has consequently altered the ways in which societies communicate through technology. On the one hand this development has encouraged communication research (e.g. Katzenbach, 2022; Keller & Klinger, 2019; Livingston, 2018; van Es & de Lange, 2020) that examines what it means to engage communicatively with machines that, while lacking genuine understanding, nevertheless act as conversational partners (Pütz & Esposito, 2024). This discourse is further enriched by studies on technology-driven social changes that conceptualizes technologies as "transformers" reshaping human experience (Allert, 2020; Dander et al., 2020).

On the other hand, these developments also entail methodological implications. Currently, a rapid evolution can be observed from large language models (LLMs) toward Retrieval Augmented Generation (RAG), which combines traditional language models with sophisticated information retrieval processes (Sahin, 2025). This integration enables information from extensive text corpora to be accessed and incorporated prior to answer generation, thereby significantly enhancing the contextual awareness and informational depth of generative socio-technical systems compared to earlier models. As such advanced models become increasingly embedded in interpersonal communication, they function not merely as tools but as co-creators of individual interactions, social dynamics, and broader societal norms (Esposito, 2024; Pütz & Esposito, 2024). In an increasingly digital and networked world, shaped by technologies that frame digital autonomy through algorithmic processes, scientific interest in these dynamics of change continues to grow, even as their societal effects remain opaque.

There is already a substantial body of research addressing the datafication of communication and society (e.g. Houben & Prietel, 2018; M. T. Schäfer & van Es, 2017). For instance, Houben and Prietel (2018) introduce the concept of a "data society" to describe a society that continuously produces data, describes itself through data, and depends on data in many respects. Schäfer and van Es also problematize the changing role of data: "Just as electricity changed industrial processes and domestic practices in the nineteenth century, a data-driven paradigm will constitute the core of twenty-first-century processes and practices" (Schäfer und van Es, 2017, S. 11).

For the authors of these contributions, it is therefore consequential that data increasingly move to the center of media and communication research. They conclude that in a society in which many aspects of language, discourse, and culture are datafied, research must also critically examine data itself as an object of inquiry. Scholarship concerned with the datafication of society (Bock et al., 2023; e.g. Houben & Prietel, 2018; M. T. Schäfer & van Es, 2017) shares, among other insights, the view that datafication processes are essential for contemporary social transformations and that the role of data for the reproduction and control of society requires continuous critical reflection.

The following considerations therefore rest on the assumption that communication in a datafied society, in a world shaped by algorithmic decision-making and interwoven with socio-technical systems, is undergoing significant change, and that communication science research must adapt accordingly in order to continue to adequately address emerging objects of study and the associated research questions.

Consequently, communication science has long shared an interest in data and data-based communication processes with other research disciplines (e.g. Breiter & Hepp, 2018; Gentzel, 2017). Currently, however, against the backdrop of ongoing datafication processes and stimulated by the possibilities of socio-technical systems, often simplistically referred to as "AI", a new entanglement of methodological approaches in (communication) science can be observed. At the same time, many established methods appear to require adaptation, if not to have become partially outdated, in light of technologies based on machine learning, deep learning, or language modeling. For example, who will continue to rely on manual transcriptions of guided interviews or complex text reduction and categorization procedures informed by content analysis when LLMs can process interview data within seconds?

In view of these developments, communication research is increasingly called upon to (re-)position itself methodologically, or rather to reflect more systematically on its existing strengths and integrate them into the available systems. This requires collaboration across different schools of thought and research paradigms in order to find answers for the question of which alternative or innovative approaches might replace established methods or meaningfully complement them. Irina Zakharova points out, that social science research on datafication is "performative" (Zakharova, 2022), meaning that research on datafication can actively shape the phenomena it investigates. She therefore argues for greater reflexivity regarding methodological and conceptual approaches in data research and, following Law (2021) and Law and Lin (2020), calls for "more careful data studies, attuned to empirical, conceptual, and ontological multiplicities (Mol, 2002) of datafication processes" that are simultaneously open to reflection and "(self-)critique" (Zakharova, 2022, p. 20).

As illustrated so far, an emerging body of research promotes a closer interweaving of datafication research and qualitative research approaches (e.g. Wiedemann, 2013; Zakharova, 2022). One field that already brings together technological implications with critical and reflexive inquiry are STS, which have gained considerable prominence in Anglo-American research contexts and have been substantially informed by media and communication research. STS, together with CDS, therefore serve as a theoretical frame of reference for the following considerations.

3. Science and Technology Studies as a theoretical and methodological reference frame

As an interdisciplinary field, STS explore the entanglement of science, technology, and society, seeking to understand how scientific knowledge and technological systems are developed, how they influence society, and how society, in turn, influences them (Sismondo, 2010). STS conceive of science and technology as "thoroughly social activities", since "scientists and engineers are always members of communities, trained into the practices of those communities and necessarily working within them" (Sismondo, 2010, p. 24).

STS combine insights and methods from various disciplines, including sociology, anthropology, history, philosophy, and political science, to provide valuable perspectives on how scientific knowledge and technological systems emerge and operate within broader social and cultural contexts. The field has developed and become institutionalized since the 1970s (Gentzel, 2017, Lengersdorf & Wieser, 2014). Today, STS encompass a range of interdisciplinary research approaches, including critical data, code, and algorithm studies, which address, for example, the power of algorithmically structured orders and the emergence of big data within data assemblages.

Interdependencies between STS and communication studies have been discussed, for instance, by Carrozza & Pereira, who explore "whether Digital Methods could thus be an area where STS and Communication Studies meet" (Carrozza & Pereira, 2016, p. 231). Based on their literature review, the authors are able to show that "it is not only Communication

Studies and STS that might connect through digital methods; digital methods also emerges as a mode of communication of different scientific practices" (Carrozza & Pereira, 2016, p. 232).

To interrogate the entanglement of communication studies and STS, this article focuses on CDS as a research strand closely linked to STS. Three core assumptions of CDS serve as a starting point for further reflections on intersections and cultivations relevant to communication studies: the "made-ness of data", "interpretative sovereignty over data", and the "(in-)visibility of data".

3.1 The made-ness of data

Research that approaches data from a CDS perspective assumes that data are neither neutral nor raw nor value-free; rather, they are always "cooked" and aggregated through algorithms imbued with specific interests and connotations (boyd & Crawford, 2012, p. 667). Data surround us and are generated through our actions and our digital tools. They are collected, evaluated, monetized and, in part, made publicly available (Crawford et al., 2014; Mayer-Schönberger & Cukier, 2013; van Dijck, 2014). While communication studies often treat data as a product or units of calculation, CDS, with its emphasis on the made nature of data, invites a shift in perspective toward data practices—that is, toward *doing data*.

Sergio Sismondo, for example, highlights that conversations about data interpretation in the laboratory rarely become part of the research publication. "Other mechanisms must replace conversations in order to turn data into convincing evidence" (Sismondo, 2010, p. 127). Similarly, in contemporary communication science, data are predominantly treated as proof or fact ("data as evidence"). The negotiation processes before and during data interpretation, before data become data and evidence, are usually absent from published papers. In fact, some journals even require that data analysis be documented in a linear, clear, and comprehensible manner. While data are "made" in the sense understood by STS and CDS and emerge in contexts beyond scientific inquiry, "at the same time digital methods rely heavily on these data produced for other purposes than research in multiple ways" (Zakharova, 2022, p. 59).

For instance, during data collection, researchers may employ surveys, interviews, or automated data collection tools such as web scrapers, each representing a distinct form of data production. Surveys and interviews require active participation from respondents, whose answers are influenced by their interpretations and biases. Web scrapers, on the other hand, collect data passively from online sources, which themselves are shaped by the algorithms and structures of the platforms being scraped.

During data analysis, researchers may clean, transform, and aggregate data, processes that also constitute forms of data production and can introduce additional biases and assumptions. Even in the publication phase, the way data are presented—in charts, tables, or narratives—reflects choices that influence how the data are perceived and interpreted.

For example, consider a study on social media usage patterns. In the data collection phase, researchers might use Application Programming Interfaces (APIs) to gather tweets, making decisions about which keywords to search for and which time periods to include. During data analysis, they might apply sentiment analysis algorithms to categorize tweets as positive, negative, or neutral, embedding assumptions about language and emotion. In the publication phase, they might present their findings in a bar chart, which involve choices about grouping and visual displaying the data. Each of these steps involves a form of data production, introducing biases and assumptions that shape the final results.

The example illustrates the importance of researchers remaining particularly attentive to the constructed nature of data at every stage of the research process. Reflections on the "made-ness of data" give rise to important follow-up questions for negotiating new methodologies or further developing existing ones in communication studies: How is data produced throughout the research process? How do data sets and their informative value change when certain data gaps are addressed through methods such as "data implementation", while others remain unfulfilled? Which evaluation logics and procedures are taken for granted as standard and unquestioned, and which must still demonstrate their empirical "usefulness" or "correctness"?

3.2 The interpretative sovereignty over data

Not only are data „made", but their analysis and subsequent use are also subject to significant restrictions. For example, the resources required to access and make data available are unevenly distributed. Some can collect digital traces of human interaction "from a multitude of technical systems, such as websites, social media platforms, smartphone apps, or sensors" (Stier et al., 2020, p. 504), yet they are constrained by what information is technically accessible, crawlable, or otherwise collectable. Research that relies on non-public, proprietary data can be prohibitively expensive. Certain data owners may charge substantial fees or permit access only after appropriate legitimization or disclosure of the research design, which can limit academic freedom. At the same time, independent research, such as that conducted in communication science institutes at universities, often lacks the necessary financial resources.

Opportunities to utilize data for decision-making or to participate in the design of datasets and surveys are similarly unevenly distributed (cf. eg. Bock et al., 2023; Eubanks, 2019; Hepp et al., 2022; Kennedy, 2018; Warschauer, 2003). STS-informed work demonstrates that data are not uncontroversial (van Dijck, 2014; Zuboff, 2019), yet they underpin far-reaching decision-making processes across modern societies (e.g. Couldry, 2017), particularly in domains where communication plays a central role. This includes, for example, journalistic newsrooms increasingly publishing AI-generated content or curating content algorithmically for specific audiences, as well as social media platforms that create algorithmically driven filter bubbles, influencing information behavior and communication to varying degrees (Mahrt, 2019).

Questions about the role of bots or hate speech in election campaign communication further illustrate the stakes, particularly in political communication research (Keller & Klinger, 2019; Pentzold et al., 2019). The concept of interpretative sovereignty over data raises critical concerns regarding who holds the power to control, interpret, and leverage data (e.g. von Scherenberg et al., 2024).

Consider, for example, social media platforms such as Facebook, TikTok, or LinkedIn. The platform providers collect vast amounts of user data, which they use to target advertisements and influence user behavior. Access to this data for academic research is often limited, either due to high costs imposed by the platform providers as data owners or because of strict data-sharing policies designed to protect what is considered proprietary data. These restrictions constrain researchers' ability to study the societal impact of social media on society and to critically examine the algorithms and data practices of these platforms. Consequently, interpretative sovereignty over data largely remains in the hands of corporations, which can shape public discourse and policy decisions according to their own interests.

This example underscores the need to recognize the limitations of data resulting from unequal access and (un-)availability. For communication science research informed by CDS, insights into the interpretative sovereignty over data raise several important questions:

Who possesses the skills or knowledge to locate and analyze openly available research data? And how can research on datafied communication critically engage with its subject matter if access to the necessary data remains restricted or entirely blocked?

3.3 The (in-)visibility of data

Data and their entanglements are not always visible or equally empirically tangible. Data are often "decontextualized" and must subsequently be "recontextualized" (Zakharova, 2022; Zakharova & Bock, 2023). In other words, data, figures, and datasets are "made" from actions. These data artifacts then "end-up" in the hands of individuals who recontextualize them—initiating actions, issuing instructions, drawing conclusions, and making decisions based on the available data. The underlying processes and data flows are not always transparent and are often partially obscured, with conscious or unconscious "blackboxing" complicating data analysis (Hillman, 2023; Jarke et al., 2023; Rogers, 2017; Zakharova & Bock, 2023).

The insight from CDS that data can be made visible or invisible generates important methodological considerations for communication science research (Stehle et al., 2024). For example, when using data for research purposes, it is essential to reflect on which data are actually tangible, visible, and observable during collection and analysis. Qualitative and ethnographic research have long established strategies for visualizing the researcher's own situatedness and the limitations of the field of observation (e.g. situatedness, dealing with going native, follow the data, etc. (e.g. Agar, 2006; Quinlan, 2008)). In these approaches, the impossibility of accessing invisible data is not viewed as a deficiency in scientific quality; rather, it is productively transformed into socio-critical questions about access: Who determines which data are made freely available, and what are the societal implications if some data are emphasized while others are rendered invisible?

Consider, for example, algorithmic decision-making in political communication during election campaigns. Political parties and candidates often use data analytics to target specific voter demographics with tailored messages. However, the data and algorithms underpinning these targeted communications are frequently invisible to both the public and researchers. This invisibility carries significant implications for democratic processes, as it can lead to the manipulation of public opinion and contribute to the spread of misinformation.

For communication science research, several critical questions arise from the considerations on the (in)visibility of data: Which data are invisible, or actively made invisible? Which data collection and analysis methods can be employed to uncover such invisible data and render it researchable? Which decision-making and analytical logics determine which data are visualized and which are overlooked in the process? How do these practices of (in)visualisation affect the interpretation and meaningfulness of the research?

4. Discussing methodological approaches

The assumptions of CDS discussed here serve as a theoretical foundation for considering the practical implications of datafication for communication science research, particularly regarding the applicability of CDS approaches in empirical practice. CDS encompass a diverse methodological spectrum aimed at uncovering the social, political, and ethical dimensions of data practices (Decuypere, 2021).

These methods include data ethnographies, which involve immersive observation and participation in data-driven environments; data activism, which focuses, for example, on advocating for transparency and accountability in data use; and data feminism, which examines issues such as gendered aspects of data production and interpretation (e.g. Charles

& Gherman, 2019; Knox & Nafus, 2018; Pink et al., 2018). Additionally, CDS employ approaches such as data visualization critiques, which analyze biases and assumptions embedded in visual representations of data, and data justice frameworks, which assess the fairness and equity of data-driven decision-making processes (e.g. Gerlitz, 2017; Williamson, 2016). Among these methods, data journeys and walkthrough approaches are particularly illustrative and will be discussed in more detail in the following sections.

4.1 Data journeys

Juliane Jarke and colleagues (2023), for example, use the visualization of invisible practices as a starting point for their conceptual analysis, employing data journeys as a method to investigate data-driven educational management and administration. Drawing inspiration from Eleftheriou et al. (2018) and the work of Jo Bates and colleagues on data journeys (Bates et al., 2016), they focus on visualizing data flows—tracing where data travel, where interruptions occur, where data leave an institution or silo, and what happens in subsequent decision-making steps.

Jarke and her colleagues pay particular attention to the challenges and breaks in visualization, as well as the lengthy negotiation process required to produce a shared representation of the data flow. What is typically invisible in publications and presentations—the ambivalences, ruptures, and complexities of datafication in administrative organizations—is central to their analysis, making these phenomena empirically tangible.

The authors highlight that data flows in communication processes cannot be neatly visualized; instead, they are messy and fragile. Much of what could provide especially valuable insights into decision-making in management and administrative communication proves impossible to visualize or observe and therefore cannot be captured by traditional methods such as guided interviews or surveys. The example of fragile data journeys thus underscores the limitations of "established" methods, which largely rely on participants' ability to verbalize their impressions.

4.2 The walkthrough approach

Another example of an STS- respectively CDS-informed method applied in communication research is the walkthrough approach, which draws analytical attention to the affordances of socio-technical systems and the practices interwoven with them (Light et al., 2018; M. S. Schäfer & Wessler, 2020; Troeger & Bock, 2022). The socio-technical walkthrough allows researchers to analyze algorithmic personalization of socio-technical systems such as platforms or apps, making system affordances visible and open to reflection (Troeger & Bock, 2022).

Walkthroughs, that utilize a "research persona", include a systematic exploration of the respective platform, whereby interactions and feedback received by the research persona are documented for each step, each time the platform is approached. "Research personas" and "othering" techniques are employed to visualize, better understand, and critically reflect on individual usage paths and algorithmically personalized feedback. As Light et al. (2018) note, the walkthrough approach is particularly suited "to explore interfaces, including how they 'script' the user, by systematically documenting and abstracting interface features from their normative infrastructural settings."

The socio-technical walkthrough functions as a "border crosser" (Troeger & Bock, 2022, p. 51), combining qualitative and quantitative aspects of research, making it well-suited for multi-method designs and interdisciplinary studies. For communication science, this approach is particularly valuable because it integrates media-analytical and reception-ana-

lytical perspectives, addressing two central questions of the field: What do people do with technology, and what does technology do with people?

In sum, data journeys allow to trace the lifecycle of data from its creation to its use and eventual disposal, revealing the various actors, technologies, and contexts that shape data at each stage (Bates et al., 2016; Eleftheriou et al., 2018; Jarke et al., 2023). This method is particularly valuable for understanding the complex and often hidden processes through which data are produced, circulated, and interpreted. Socio-technical walkthrough approaches, in contrast, systematically explore and document the algorithmic personalization of data-driven platforms and technologies. By simulating user interactions, these walkthroughs can uncover design choices, biases, and power dynamics embedded in data-driven interfaces (e.g. Dieter et al., 2019; Light et al., 2018; M. S. Schäfer & Wessler, 2020; Troeger & Bock, 2022). Focusing on data journeys and walkthroughs in this contribution highlights the dynamic and relational nature of data, as well as the ways in which data practices are embedded in broader social and technological systems. These approaches offer a comprehensive and nuanced understanding of the made-ness, interpretative sovereignty, and (in-)visibility of data, making them particularly relevant for communication studies.

5. Concluding thoughts

This contribution advances a threefold argument: First, it underscores the need for communication science to adapt its research paradigms in response to the datafication of society, emphasizing interdisciplinary collaboration to develop innovative and reflexive methods (e.g. Houben & Prietel, 2018; M. T. Schäfer & van Es, 2017). Second, it demonstrates how STS, and particularly a CDS perspective, provides a productive theoretical lens for understanding data as made, interpreted, and invisibly entangled within socio-technical systems. Third, it highlights methodological approaches such as data journeys (Bates et al., 2016; Jarke et al., 2023) and socio-technical walkthroughs (Troeger & Bock, 2022) as tools that expose the constructed, contingent, and relational nature of data—making them especially suited for critical communication research.

CDS add a valuable perspective to communication science by emphasizing the social, political, and ethical dimensions of data practices (e.g. boyd & Crawford, 2012; Decuypere, 2021; van Dijck, 2014). By adopting a CDS lens, communication scientists can gain a deeper understanding of how data are produced, interpreted, and utilized across different communication contexts. This approach encourages researchers to critically examine the biases, assumptions, and power dynamics embedded in data-driven communication processes, thereby enhancing the rigor and relevance of their analyses (e.g. Gerlitz, 2017; Rogers, 2017; Zakharova & Bock, 2023).

Moreover, CDS methods—such as data journeys and socio-technical walkthroughs—offer practical tools for uncovering hidden aspects of data, making them visible and open to critical scrutiny. Without broader adoption of CDS, communication science risks overlooking the complexities and nuances of data practices, potentially producing oversimplified or biased conclusions. By integrating CDS, communication science can enrich its methodological toolkit, foster interdisciplinary collaboration, and contribute to more informed, reflective, and responsible approaches to data-driven communication practices.

There is already an ongoing discussion within the discipline regarding the futures of research in a datafied society. The computational turn, according to van Es and colleagues (2021), provides tools that greatly expand the possibilities for conducting research and allow researchers to engage directly with the activities that are shaping our digital society. In the context of an "algorithmic turn" (Katzenbach, 2022) or a "computational turn"(van Es et al., 2021, p. 61) which introduce new uncertainties for communication science research, van

Es and colleagues advocate for the development of best practices for tool application. These include standards for the use of data visualizations in publications and training colleagues and students in both the use of the tools and in critical tool literacy. Tool critique is a key component of responsible research, ensuring that computational tools are applied in a reflective and informed manner (van Es et al., 2021).

This contribution builds on these discussions but takes a slightly different direction. These developments shed new light on the so-called "science war" (Parsons, 2003) and invite a closer interlinking of the logics of qualitative and quantitative research—or at least a stronger dialogue between the two approaches. What are the implications if standardized research continues to follow the "drive to replication" (van Es et al., 2021, p. 21) in an era of algorithmic data production and analysis? What "rules" and research criteria should the discipline adopt when reproducibility or traceability cannot be guaranteed due to algorithmic methods?

One possible avenue is to draw on the quality criteria and research logic of qualitative or ethnographically oriented research, which has long emphasized, for example, the explication of the constructedness of data and the situatedness of researchers (e.g. Breidenstein et al., 2020; Hirschauer, 2014). Communication science research already supports this perspective: "In other words, responsible knowledge claims not only make transparent how they have been constructed, they also provide a justification and account as to why such claims were constructed in the way they have" (van Es et al., 2021, p. 56).

In the current landscape of datafied communication and datafied research, there is not yet a satisfactory, fixed "method compendium" with standardized rules (van Es et al., 2021, p. 61), as has long existed for traditional approaches such as observation, interviews, or content analysis. It is also uncertain whether such a standardized set of methodological tools will ever be available in an era of algorithmic data collection and analysis—particularly given the continuing (and potentially exponential) growth in the capabilities of socio-technical systems, exemplified by developments such as Retrieval-Augmented Generation (RAG) in generative AI.

This raises a critical question for the discipline: in light of the rapid acceleration of change in communication science research objects, should the field focus on standardizing individual methods, or—as this article argues—invest more energy in methodological innovation that foregrounds the "made-ness", "interpretive processes", and "invisibility" of data?

As an outlook for future communication science interested in datafied communication and its socio-political implications, three key questions emerge, highlighting challenges for further research.

First, if data is "cooked", "biased", or otherwise "problematic", how can researchers critically engage with its use in politics, science, and society? From a CDS perspective, researchers are encouraged to adopt a stance of critical data literacy. This involves interrogating the provenance of data, the methods used to collect it, and its inherent biases. Reflexive engagement with data requires acknowledging its socio-political context and the power dynamics that shape it (e.g. Eubanks, 2019; Zakharova & Bock, 2023). Collaborative efforts with data scientists, ethicists, and domain experts can foster a more nuanced understanding and responsible understanding of data, while promoting transparency and accountability through open data practices and robust documentation can help mitigate these challenges.

Second, if the availability of data for research is highly regulated and concentrated among powerful "*data haves*", how can researchers who are "*data non-haves*" contribute to shape infrastructures for research? Navigating the power imbalances requires proactive engagement in the development of research infrastructures. Researchers who are "data

non-haves" can advocate for open data policies, data sharing agreements, and collaborative research frameworks that democratize data access.

Engaging with policymakers, funding bodies, and institutional stakeholders can help ensure that infrastructures prioritize equitable data access. Additionally, building alliances and networks with other "data non-haves" can amplify voices and influence and foster more inclusive data infrastructures. Initiatives such as the Communicative AI Research group—with its ComAI research space and a focus on governance at the ZEMKI in Bremen—offer promising platforms to explore and advance these goals.

Third, if data is "travelling", "invisible", and "intertwined", what methodological (re-)adjustments should, can, or must researchers make? Scholars are encouraged to embrace interdisciplinary approaches that integrate qualitative and quantitative methods to trace data journeys and uncover hidden dynamics. Adopting a relational perspective that situates data within its broader ecosystem can illuminate how data are intertwined with people, technologies, and institutions (e.g. Wiedemann, 2013).

Methods such as data ethnography, data provenance analysis, and network analysis can provide insights into the complex lifecycles and entanglements of data. Additionally, fostering a culture of continuous learning and methodological adaptation enables researchers to remain responsive to the evolving landscape of data and its societal implications.

Literature

Agar, M. (2006). An ethnography by any other name ... *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *Vol 7*, No 4 (2006): Qualitative Research in Ibero America. https://doi.org/10.17169/fqs-7.4.177

Allert, H. (2020). Plattformökonomie und Entstaatlichung: Familienorientiert, ortsunabhängig und #freilernend. In V. Dander, P. Bettinger, E. Ferraro, & K. Rummler (Eds), *Digitalisierung – Subjekt – Bildung. Kritische Betrachtungen der digitalen Transformation* (pp. 183–212). Barbara Budrich. https://doi.org/10.2307/j.ctvvb7n3h.13

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, *3*(2), 2053951716654502. https://doi.org/10.1177/2053951716654502

Binz, M., Alaniz, S., Roskies, A., Aczel, B., Bergstrom, C. T., Allen, C., Schad, D., Wulff, D., West, J. D., Zhang, Q., Shiffrin, R. M., Gershman, S. J., Popov, V., Bender, E. M., Marelli, M., Botvinick, M. M., Akata, Z., & Schulz, E. (2025). How should the advancement of large language models affect the practice of science? *Proceedings of the National Academy of Sciences*, *122*(5), e2401227121. https://doi.org/10.1073/pnas.2401227121

Bock, A., Breiter, A., Hartong, S., Jarke, J., Jornitz, S., Lange, A., & Macgilchrist, F. (Eds). (2023). *Die datafizierte Schule* (1. Auflage 2023). Springer VS. https://doi.org/10.1007/978-3-658-38651-1

boyd, d., & Crawford, K. (2012). Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), Article 5. https://doi.org/10.1080/1369118X.2012.678878

Breidenstein, G., Hirschauer, S., Kalthoff, H., & Nieswand, B. (2020). *Ethnografie: Die Praxis der Feldforschung* (3., überarbeitete Auflage). UVK Verlag.

Breiter, A., & Bock, A. (2023). Datafizierte Gesellschaft | Bildung | Schule. In A. Bock, A. Breiter, S. Hartong, J. Jarke, S. Jornitz, A. Lange, & F. Macgilchrist (Eds), *Die datafizierte Schule* (pp. 1–35). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-38651-1_1

Breiter, A., & Hepp, A. (2018). The complexity of datafication: Putting digital traces in context. In A. Hepp, A. Breiter, & U. Hasebrink (Hrsg.), *Communicative Figurations* (pp. 387–405). Springer International Publishing. https://doi.org/10.1007/978-3-319-65584-0_16

Carrozza, Tiago Santos, & Pereira. Chiarra (2016). Mapping digital methods: Where science and technology studies and communication studies meet? In A. Gradim & C. Moura (Hrsg.), *Communicating and Evaluating Science* (pp. 203–236). Beira Interior University.

Charles, V., & Gherman, T. (2019). Big data analytics and ethnography: Together for the greater good. In A. Emrouznejad & V. Charles (Eds), *Big Data for the greater good* (Vol. 42, pp. 19–33). Springer International Publishing. https://doi.org/10.1007/978-3-319-93061-9_2

Couldry, N. (2017). The Myth of Big Data. In M. T. Schäfer & K. van Es (Eds), *The datafied society: Studying culture through data* (pp. 235–239). Amsterdam University Press.

Crawford, K., Gray, M. L., & Miltner, K. (2014). Big data| Critiquing big data: Politics, ethics, epistemology | Special Section Introduction. *International Journal of Communication*, *8*(0), Article 0. https://ijoc.org/index.php/ijoc/article/view/2167

Dander, V., Bettinger, P., Ferraro, E., Leineweber, C., & Rummler, K. (Eds). (2020). *Digitalisierung – Subjekt – Bildung: Kritische Betrachtungen der digitalen Transformation* (1st edn). Verlag Barbara Budrich. https://doi.org/10.2307/j.ctvvb7n3h

Decuypere, M. (2021). The Topologies of Data Practices: A Methodological Introduction. *Journal of New Approaches in Educational Research*, *10*(1), Article 1. https://doi.org/10.7821/naer.2021.1.650

Decuypere, M., Grimaldi, E., & Landri, P. (2021). Introduction: Critical studies of digital education platforms. *Critical Studies in Education*, *62*(1), Article 1. https://doi.org/10.1080/17508487.2020.1866050

Dieter, M., Gerlitz, C., Helmond, A., Tkacz, N., van der Vlist, F. N., & Weltevrede, E. (2019). Multi-Situated App Studies: Methods and Propositions. *Social Media + Society*, *5*(2), 2056305119846486. https://doi.org/10.1177/2056305119846486

Eleftheriou, I., Embury, S. M., Moden, R., Dobinson, P., & Brass, A. (2018). Data journeys: Identifying social and technical barriers to data movement in large, complex organisations. *Journal of Biomedical Informatics*, *78*, 102–122. https://doi.org/10.1016/j.jbi.2017.12.001

Esposito, E. (2024). *Kommunikation mit unverständlichen Maschinen*. Residenz Verlag.

Eubanks, V. (2019). *Automating inequality: How high-tech tools profile, police, and punish the poor* (First Picador edition). Picador St. Martin's Press.

Franken, L. (2022). Digitale Daten und Methoden als Erweiterung qualitativer Forschungsprozesse. Herausforderungen und Potenziale aus den Digital Humanities und Computational Social Sciences. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *23*(2). https://doi.org/10.17169/fqs-22.2.3818

Gentzel, P. (2017). Praktisches Wissen und Materialität. Herausforderungen für kritisch- konstruktivistische Kommunikations- und Medienforschung. *Medien & Kommunikationswissenschaft*, *65*(2), 275–293. https://doi.org/10.5771/1615-634X-2017-2-275

Gerlitz, C. (2017). Data point critique. In M. T. Schäfer & K. van Es (Eds), *The datafied society: Studying culture through data* (pp. 241–244). Amsterdam University Press.

Hepp, A., Jarke, J., & Kramp, L. (Eds). (2022). *New Perspectives in Critical Data Studies: The ambivalences of data power*. Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0

Hillman, V. (2023). Bringing in the technological, ethical, educational and social-structural for a new education data governance. *Learning, Media and Technology*, *48*(1), 122–137. https://doi.org/10.1080/17439884.2022.2052313

Hintz, A., Dencik, L., & Wahl-Jørgensen, K. (2019). *Digital citizenship in a datafied society*. Polity Press.

Hirschauer, S. (2014). Un/doing Differences. Die Kontingenz sozialer Zugehörigkeiten. *Zeitschrift für Soziologie, 43*(3), 170–191.

Houben, D., & Prietel, B. (2018). *Datengesellschaft: Einsichten in die Datafizierung des Sozialen*. Transcript.

Jarke, J., Hartong, S., Raabe, T., Dabisch, V., Breiter, A., Lange, A., & Zakharova, I. (2023). Zur Erfassung und Modellierung der „Hinterbühne" von Datenflüssen: Das Beispiel Unterrichtsausfall. In A. Bock, A. Breiter, S. Hartong, J. Jarke, S. Jornitz, A. Lange, & F. Macgilchrist (Eds), *Die datafizierte Schule* (pp. 61–91). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-38651-1_3

Katzenbach, C. (2022). Der „Algorithmic turn" in der Plattform-Governance. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *74*(S1), 283–305. https://doi.org/10.1007/s11577-022-00837-4

Keller, T. R., & Klinger, U. (2019). Social Bots in Election Campaigns: Theoretical, Empirical, and Methodological Implications. *Political Communication*, *36*(1), 171–189. https://doi.org/10.1080/10584609.2018.1526238

Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication. *Krisis I*, 18–30. https://eprints.whiterose.ac.uk/id/eprint/12995

9/2/Krisis-2018-1-Helen-Kennedy-Living-with-Data-Aligning-Data-Studies-and-Data-Activism-Through-a-Focus-on-Everyday-Experiences-of-Datafication.pdf-18-30.pdf [30.01.2026].

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures & their consequences*. Sage Publications.

Knox, H., & Nafus, D. (2018). *Ethnography for a data-saturated world*. Manchester university press.

Law, J. (2021). From after method to care-ful research (a foreword). In C. Addey & N. Piattoeva (Eds.), Intimate accounts of education policy research (pp. xvi–xx). Routledge.

Law, J., & Lin, W. (2020). Care-ful Research: Sensibilities from STS. 17.

Lengersdorf & M. Wieser (Hrsg.), Schlüsselwerke der Science & Technology Studies (S. 123-132). Wiesbaden: VS.

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. https://doi.org/10.1177/1461444816675438

Lippert, I., & Mewes, J. S. (2021). Data, methods and writing: Methodographies of STS ethnographic collaboration in practice. *Science & Technology Studies*, *34*(3), 2–16. https://doi.org/10.23987/sts.110597

Livingston, S. (2018). Audiences in an age of datafication: Critical questions for media research. *Television and New Media*.

Mahrt, M. (2019). Beyond filter bubbles and echo chambers: The integrative potential of the Internet. ifpuk – Institute for Media and Communication Studies at FU Berlin. (Digital *Communication Research;* 1). https://doi.org/10.17174/DCR.V5.0

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. Murray.

Mol, A. (2002). *The Body Multiple: Ontology in Medical Practice*. Duke University Press. https://doi.org/10.2307/j.ctv1220nc1

Parsons, K. M. (Ed.). (2003). *The science wars: Debating scientific knowledge and technology*. Prometheus Books.

Pentzold, C., Brantner, C., & Fölsche, L. (2019). Imagining big data: Illustrations of „big data" in US news articles, 2010–2016. *New Media & Society*, *21*(1), Article 1. https://doi.org/10.1177/1461444818791326

Peters, U., & Chin-Yee, B. (2025). Generalization bias in large language model summarization of scientific research. *Royal Society Open Science*, *12*(4), 241776. https://doi.org/10.1098/rsos.241776

Pink, S., Berg, M., Lupton, D., & Ruckenstein, M. (2022). *Everyday automation: Experiencing and anticipating Emerging Technologies* (1st edn). Routledge. https://doi.org/10.4324/9781003170884

Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, *5*(1), Article 1. https://doi.org/10.1177/2053951717753228

Pütz, O., & Esposito, E. (2024). Performance without understanding: How ChatGPT relies on humans to repair conversational trouble. *Discourse & Communication*, *18*(6), 859–868. https://doi.org/10.1177/17504813241271492

Quinlan, E. (2008). Conspicuous invisibility: Shadowing as a data collection strategy. *Qualitative Inquiry*, *14*(8), 1480–1499. https://doi.org/10.1177/1077800408318318

Rogers, R. (2017). Foundations of digital methods. Query design. In M. T. Schäfer & K. van Es (Eds), *The datafied society: Studying culture through data* (pp. 75–94). Amsterdam University Press.

Sahin, A. (2025, March 5). What is Retrieval-Augmented Generation (RAG) in LLM and How it works? *Medium*. https://medium.com/@sahin.samia/what-is-retrieval-augmented-generation-rag-in-llm-and-how-it-works-a8c79e35a172 [27.01.2025].

Schäfer, M. S., & Wessler, H. (2020). Öffentliche Kommunikation in Zeiten künstlicher Intelligenz. *Publizistik*, *65*(3), 307–331. https://doi.org/10.1007/s11616-020-00592-6

Schäfer, M. T., & van Es, K. (Eds). (2017). *The datafied society: Studying culture through data*. Amsterdam University Press.

Schneider, T., Fuchs, K., & Mayer, S. (2021). Science and Technology Studies: Die Datafizierung von Alltagspraktiken: Datenaktivismus als neue Verantwortung? In A. Henkel (Ed.), *10 Minuten Soziologie: Verantwortung* (pp. 183–196). transcript Verlag. https://doi.org/10.14361/9783839451120-012

Sismondo, S. (2010). *An introduction to science and technology studies* (Second edition). Wiley-Blackwell.

Stehle, H., Bock, A., Wilhelm, C., Springer, N., Mahrt, M., Lobinger, K., Linke, C., Engelmann, I., Detel, H., & Brantner, C. (2024). In/Visibility in the digital age: A literature review from a communication studies perspective. *International Journal of Communication*, *18*(0), 5471–5493.

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, *38*(5), 503–516. https://doi.org/10.1177/0894439319843669

Troeger, J., & Bock, A. (2022). The sociotechnical walkthrough – a methodological approach for platform studies. *Studies in Communication Sciences*, *22*(1), 43–52. https://doi.org/10.24434/j.scoms.2022.01.3064

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

van Es, K., & de Lange, M. (2020). Data with its boots on the ground: Datawalking as research method. *European Journal of Communication*, *35*(3), 278–289. https://doi.org/10.1177/0267323120922087

van Es, K., Schäfer, M. T., & Wieringa, M. (2021). Tool criticism and the computational turn. A „Methodological moment" in Media and Communication Studies. *M&K Medien & Kommunikationswissenschaft*, *69*(1), 46–64. https://doi.org/10.5771/1615-634X-2021-1-46

von Scherenberg, F., Hellmeier, M., & Otto, B. (2024). Data sovereignty in information systems. *Electronic Markets*, *34*(1), 15. https://doi.org/10.1007/s12525-024-00693-4

Warschauer, M. (2003). *Technology and social inclusion: Rethinking the digital divide*. MIT Press.

Wiedemann, G. (2013). Opening up to Big Data: Computer-assisted analysis of textual data in social sciences. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *14*(2). Art. 23. https://doi.org/10.17169/fqs-14.2.1949

Williamson, B. (2016). Digital education governance: Data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy*, *31*(2), 123–141. https://doi.org/10.1080/02680939.2015.1035758

Wollin-Giering, S., Hoffmann, M., Höfting, J., & Ventzke, C. (2024). Automatic transcription of English and German Qualitative Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *25*(1) Art. 8. https://doi.org/10.17169/fqs-25.1.4129

Zakharova, I. (2022). *Understanding data studies—A methodological and conceptual inquiry into research on datafication* [Universität Bremen]. https://doi.org/10.26092/ELIB/1675

Zakharova, I., & Bock, A. (2023). Researching in/visibilities and data practices in education. *On_education. Journal for Research and Debate*, *6*(18). https://doi.org/10.17899/on_ed.2023.18.4

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile books.

Besprechungen

Hardy Gundlach (Hrsg.)
Medienregulierung

Nomos, 2025. – 783 S.

3., vollständig aktualisierte und erweiterte Auflage

ISBN 978-3-7560-0086-9 (Print)

Ohne Zweifel: Fragen der Ausgestaltung des Medien- und Kommunikationssystems, das sich unter digitalen Bedingungen und unter starkem Einfluss mächtiger amerikanischer und chinesischer Unternehmen entwickelt, sind von großer gesellschaftlicher Bedeutung. Und sie sind, mehr denn je, auch von hoher wissenschaftlicher Relevanz. Von daher ist es sehr zu begrüßen, dass der einst von Wolfgang Seufert maßgeblich konzipierte Band eine Aktualisierung und Neukonzeption erfahren hat. Der Band schließt im deutschsprachigen Raum eine Lücke, sieht man einmal vom Buch von Manuel Puppis ab (Manuel Puppis [2023]. Einführung in die Medienpolitik. UTB, UVK, 3. vollständig überarbeitete und erweiterte Auflage). Daher leuchtet es ein, dass der in Fribourg lehrende Kommunikationswissenschaftler mit einem systematischen Beitrag, über die Medienregulierung in der Schweiz, im vorliegenden Band vertreten ist.

Der nun als Alleinherausgeber fungierende Hardy Gundlach hat, und das ist eine begrüßenswerte Innovation, dem D-A-CH-Raum viel Aufmerksamkeit geschenkt. Neben Puppis gibt es zwei weitere Beiträge aus Schweizer Federn, von Mark Eisenegger/Linards Udris zum Medien-Qualitätsmonitoring und von Matthias Künzler über die in der Schweiz erfolgreich erprobten Formen der ko-regulierten Medienqualitätssicherung. Josef Trappel und Autorenteam befassen sich auch mit diesen Problemstellungen in ihrem systematischen Überblicksbeitrag über die Medienpolitik in Österreich. Komparativ angelegt ist der Beitrag von Paul Clemens Murschetz zur Medienförderung im D-A-CH-Raum. Man lernt: Die Kleinstaaten übernehmen keineswegs nur Konzepte und Verfahren von ihren großen Nachbarstaaten, sie haben eigene Instrumente entwickelt, sind innovativ, so indem auf Medienqualität oder auf neue Akteure, bspw. Ombudsstellen, abgestellt wird. Da hat Deutschland noch einiges nachzuholen.

In 14 Kapitel wird das Thema vor allem aus medienökonomischer, selektiver aus kommunikationswissenschaftlicher Perspektive in den Beiträgen verfolgt. Die starke institutionenökonomische Perspektive ist zweifellos ein Alleinstellungsmerkmal für den Band, man spürt den prägenden Einfluss von Seufert. Gundlach und Seufert haben den sehr überzeugenden Leitbeitrag über „Regulierungsziele und -konzepte“ gemeinsam verfasst. Das überaus komplexe Thema behandeln sie auf fast 180(!) Druckseiten, ein Buch im Buch. Medienökonomisch ausgerichtete Beiträge steuern zudem M. Björn von Rimscha (Werbung) und das Autorenteam um Annika Stöhr (Regulierung des Medienwettbewerbs) bei.

Deutlicher geringer ist – sieht man vom Beitrag Gundlachs über die „Regulierung der Medien in Deutschland“ ab – die sozialwissenschaftliche Perspektive auf Medien und deren Regulierung vertreten. Dem in Transformation befindlichen Politikfeld der Medienpolitik (Telekommunikations-, Digitalpolitik) wird zu wenig Aufmerksamkeit geschenkt. Immerhin werden bestimmte neue Akteure, so Influencer, betrachtet. Der Beitrag von Juliane A. Lischka ist innovativ.

Es ist verständlich, und das formuliert der Herausgeber im Vorwort, dass nicht auch noch die medienrechtliche Perspektive verfolgt werden kann. Die dann aber getroffene selektive Auswahl (Geistiges Eigentum) vermag nicht so ganz zu überzeugen, wenngleich der Beitrag von Svenja Hagenhoff sehr lesenswert ist.

Der Verlag hat, so die Eigenwerbung, ein „Handbuch“ vorgelegt. Im Vorwort zur 2. Auflage wird von „Lehrbuch“ gesprochen, im aktuellen Vorwort zu 3. Auflage kommt zudem der Begriff „Sammelband“ vor. Aufgrund der Zielsetzung Lehrbuch sind die Beiträge mit „Diskussionsfragen“ versehen, und es werden wenige, ausgewählte Literaturempfehlungen gegeben. Die unterschiedliche Qualität dieser Angaben ist auffällig. Auf die „Diskussionsfragen“ wird nicht weiter Bezug genommen. Als Lehrbuch wäre der Band selbst für die Master-Stufe sehr anspruchsvoll, vor allem auch – mit fast 700 eng bedruckten Seiten – deutlich zu umfangreich. Dem „Handbuch“-Charakter wird mit einem Sachregister zu entsprechen versucht. Es würde aber Sinn machen, Kernbegriffe nicht nur ab und an in Texten zu behandeln, sondern einen systematischen Überblick zu bieten. Insoweit sollten die Beiträge dann einheitlicher angelegt sein. Der Band ist sehr voluminös, die Texte sind unterschiedlich lang, sie sind unterschiedlich gegliedert. Hier sind m. E. rasch redaktionelle Verbessrungen möglich, um die Texte besser erschließen zu können.

Inhaltlich ist auffällig, dass die Beschäftigung mit Digitalisierungsphänomenen zu bescheiden ist. Das etablierte Regulierungsregime der EU kommt in den Beiträgen natürlich vor, aber diese regulatorisch bedeutenden Veränderungen (Kompetenzordnung) werden nicht fokussiert

DOI: 10.5771/1615-634X-2026-1-102

gewürdigt. Auch Digitalpolitik wird nicht ausgeflaggt, im Band dominiert eine medienpolitische Sichtweise. Lediglich in einem Beitrag (Birgit Stark: Transparenzpflichten von Plattformen) wird grundlegend und überzeugend dargelegt, vor welchen neuen Herausforderung Regulierung steht. Im abschließenden Beitrag des Herausgebers zu den „Perspektiven der Medienregulierung" finden sich viele anregende Hinweise, vor allem in Form von Aufzählungen. Selbstverständlich kann ein Band, so aufgrund der raschen Veränderung im digitalen Sektor, nicht allerneueste Entwicklungen (KI, AI Act, European Media Freedom Act etc.) berücksichtigen. Aber das vorhandene Wissen sollte im Rahmen der Möglichkeiten systematisch bzw. problembezogen gebündelt werden, wenn das Ziel Lehrbuch verfolgt werden soll.

Das Themenfeld ist und bleibt wichtig. Eine Weiterentwicklung dieses Bandes ist sehr wünschenswert. Der Hausgeber sollte sich zwischen Hand- oder Lehrbuch entscheiden, Sammelbände und Journal-Beiträge zu den vielen Spezialfragen gibt es ja. Eine Fokussierung auf theoretisches Basiswissen- und Problemwissen macht Sinn. Dabei sollte die europäische Perspektive stärker Berücksichtigung finden, sowohl was die einzubeziehende Forschungscommunity als auch die internationale Forschungsliteratur anbetrifft. Das sind durchaus anspruchsvolle Anforderungen. Aber der Kern für ein solches Vorhaben ist vorhanden. Es ist das Verdienst von Wolfgang Seufert und von Hardy Gundlach, Grundlagen gelegt zu haben, auf denen man weiter aufbauen kann.

Otfried Jarren

Prof. em. Dr. Otfried Jarren, Universität Zürich, Institut für Kommunikationswissenschaft und Medienforschung (IKMZ), Andreasstrasse 15, 8050 Zürich, Schweiz, o.jarren@ikmz.uzh.ch

Marcel Lemmes

Visuelle Desinformation

Digitale Bilder zwischen Populismus und Manipulation

Halem: 2025. – 368 S.

ISBN (Print) 978-3-86962-741-0

Marcel Lemmer verfolgt das ambitionierte Ziel, populistische Bildkommunikation und visuelle Desinformation theoretisch zu fassen und begrifflich präzise zu systematisieren. Die Arbeit ist theoretisch-konzeptionell angelegt.

In den Kapiteln 3 und 4 werden mit Populismus und Desinformation zwei zentrale Begriffe der Arbeit ausführlich eingeführt. *Populismus* wird dabei ausdrücklich nicht nur als politischer Kampfbegriff verstanden, sondern als analytische Kategorie zur Beschreibung spezifischer Muster politischer Kommunikation (S. 38). Der Autor bewegt sich sicher innerhalb der medien- und kommunikationswissenschaftlichen Debatten und arbeitet die Relevanz populistischer Kommunikationslogiken für öffentliche Meinungsbildung überzeugend heraus. Auch wenn der Fokus hauptsächlich auf professionellen und institutionellen Akteur:innen liegt, trägt das Kapitel zur konzeptionellen Klarheit des Populismusbegriffs bei. Die Ausführungen zur digitalen Transformation politischer Kommunikation bleiben dagegen knapp und hätten durch die Einbindung aktuellerer Debatten noch gewinnbringend erweitert werden können (S. 50).

Das anschließende Kapitel zur *Desinformation* ist klar strukturiert und folgt einem konsistenten Aufbau: Auf die Darstellung bestehender Ansätze folgen deren Zusammenfassung und kritische Einordnung, bevor der Autor ein eigenes mehrdimensionales Begriffsverständnis entwickelt (S. 88). Besonders hervorzuheben ist die Entscheidung, Desinformation nicht über eine Täuschungsabsicht, sondern über den Begriff der Manipulation zu definieren (S. 96). Diese Verschiebung erweist sich als theoretisch produktiv und erlaubt eine differenzierte Betrachtung manipulativer Kommunikationsprozesse auf gesellschaftlicher Mikro-, Meso- und Makroebene, wenngleich die mittlere Ebene im Vergleich zu den anderen beiden teilweise unscharf bleibt (S. 101). Die anschließende Auseinandersetzung mit den Begriffen Propaganda, Verschwörungstheorien und Fake News verdeutlicht den Anspruch des Autors, den eigenen Desinformationsbegriff möglichst trennscharf zu positionieren, wenngleich Komplexität und Unschärfe der Gegenstände nicht immer aufzulösen sind.

Eine der größten Stärken der Arbeit entfaltet sich im umfangreichen fünften Kapitel zum *Bild*. Hier gelingt es Lemmer kunsthistorische, zeichentheoretische und kommunikationswissenschaftliche Perspektiven produktiv miteinander zu verbinden. Die Diskussion klassischer bildwissenschaftlicher Positionen (u. a. Bredekamp, Panofsky, Barthes, Eco, Mitchell, Belting) ist fundiert und wird konsequent auf den Forschungsgegenstand bezogen. Besonders überzeugend ist die zentrale Annahme, dass visuelle Kommunikation kein neutrales Abbild sozialer Realität darstellt, sondern stets kulturell, ideologisch und kontextuell geprägt ist (S. 198). Diese Einsicht wird sorgfältig hergeleitet und bildet ein tragfähiges Fundament für die Analyse populistischer Bildkommunikation.

Die Einführung einer integrativen Bildpragmatik stellt einen weiteren wichtigen Beitrag der Arbeit dar. Indem Bilder als Handlungsmedien verstanden werden, rückt ihre performative Dimension in den Mittelpunkt der Analyse. Die Betonung von Verwendungszusammenhängen, sozialen Praktiken und Machtbeziehungen erlaubt es, visuelle Kommunikation nicht isoliert, sondern als Teil gesellschaftlicher Aushandlungsprozesse zu begreifen. Gerade hier wird deutlich, dass die Arbeit über eine bloße Zusammenfassung bestehender Theorien hinausgeht und einen eigenständigen analytischen Rahmen entwickelt.

In den folgenden Kapiteln zum *digitalen Bild* wird diese Perspektive weiter vertieft. Die einbezogenen Fallbeispiele – Memes, Sharepics und Wahlplakate 2.0 – illustrieren sinnvoll den Zugang zum Gegenstand und verdeutlichen die Relevanz der theoretischen Überlegungen. Im siebten Kapitel werden schließlich *populistische Bilder* definiert (S. 297) und unterschiedliche Formen visueller Desinformation typologisch erfasst (S. 311). Die entwickelten Kategorien sind klar strukturiert und ermöglichen eine differenzierte Analyse manipulativer Bildstrategien. Die exemplarischen Bildanalysen illustrieren diese Typologien anschaulich und machen das analytische Potenzial des Ansatzes sichtbar.

Insgesamt überzeugt die Arbeit durch ihre begriffliche Präzision, ihre theoretische Breite und die sinnvolle Strukturierung eines schwer fassbaren Forschungsfeldes. Der Autor bewegt sich souverän durch unterschiedliche theoretische Traditionen und reflektiert die Grenzen begrifflicher Trennschärfe offen. Gerade in den bildtheoretischen Teilen entfaltet die Arbeit ihre größte Überzeugungskraft und stellt einen substanziellen Beitrag zur Forschung über visuelle politische Kommunikation dar.

Patrick Nehls

Dr. Patrick Nehls, Universität Trier, Institut für Medienwissenschaft, DFG-Projekt „Visuelle Kommunikation in Zeiten des Krieges“, Universitätsring 15, 54296 Trier, Deutschland, nehls@uni-trier.de, https://orcid.org/0009-0003-8041-5504

Karsten Rudolph

Sendestörung

Aufstieg und Krise des öffentlich-rechtlichen Rundfunks

Beck, 2025. – 240 S.

ISBN 978-3-406-83690-9

In zwei Jahren wird der Öffentlich-Rechtliche Rundfunk (ÖRR) in Deutschland seinen 80. Geburtstag feiern. Und die Rückblicke werden zeigen, wie eigentümlich diese Konstruktion ist: ein Implantat ohne Verankerung in der deutschen Medientradition, den Deutschen oktroyiert durch die britische und amerikanische Besatzungsmacht. Die „Anstalten“, etwa der WDR oder das ZDF, sind weder Unternehmen noch Behörden. Regional gegliedert, aber bundesweit aktiv. Hierarchisch organisiert, aber kontrolliert durch Gremien mit Vertretern aus Landtagsfraktionen und Verbänden. Mit großen publizistischen Freiräumen, aber medienpolitisch gerahmt und geformt.

Bereits jetzt kann man zurückblicken mit Hilfe eines Buchs von Karsten Rudolph, Historiker an der Ruhruniversität Bochum, ehemals SPD-Abgeordneter im nordrhein-westfälischen Landtag und Mitglied in den Aufsichtsgremien des WDR. In seinem Rückspiegel kann man sehen, welche gefährlichen Klippen diese Konstruktion in den 80 Jahren umschifft hat: so etwa den Versuch Adenauers, ein Staatsfernsehen zu gründen; die Konkurrenz durch private TV-Sender; die Osterweiterung im Zuge der Wiedervereinigung; die Übergriffe der EU mit Fernsehrichtlinie und Quotenvorgaben und und und.

Verblüffend ist im Rückblick: Aus jeder Bedrohung ist der ÖRR gestärkt hervorgegangen. Dies zeugt von einer enormen Resilienz. Die hat drei Gründe: Erstens hat das Bundesverfassungsgericht den Bestand und die Weiterentwicklung den ÖRR verfassungsrechtlich garantiert. Eine Säule dessen ist die Pflichtgebühr. Dadurch sprudelt eine nie versiegende Finanzquelle – der zweite Grund für die Resilienz. Dies ermöglicht vor allem, den Beschäftigten des ÖRR das Beste aus beiden Welten zu bieten: Gehälter, die mit der Privatwirtschaft konkurrenzfähig sind, und Versorgungsleistungen wie in den höheren Etagen des öffentlichen Dienstes. Und der dritte Grund ist die Einsicht von Union und SPD, nicht konfrontativ, sondern im Einvernehmen den ÖRR zu steuern.

Dafür entwickelten sie überaus komplizierte Verfahren. So wurden die Rundfunkanstalten entlang von Bundesländergrenzen zugeschnitten, was den Einfluss der jeweiligen Landesregierungen und Landesparlamente sicherte. Oder es wurde gesetzlich vorgeschrieben, dass die Berichterstattung politisch ausgewogen zu sein habe. Damit bekommen Schwarz und Rot immer etwa gleiche Chancen, öffentlich gehört zu werden, auch wenn die Regierung wechselt. Und besonders augenfällig: Die Auswahl von Führungskräften geschieht nach dem Reißverschlussprinzip, immer im Proporz nach Parteifarben. Ab und an geriet der Einflussmotor ins Stottern, und es brach öffentlich Streit aus. Da wurde mal

ein aufmüpfiger Journalist gegängelt, oder Bayern drohte mit Alleingängen. Aber das wurde immer schnell wieder eingefangen, und schon surrte die ÖRR-Maschine wieder. Folgenreicher waren die mehrfachen Anläufe der Union, ihren Einfluss auf die öffentliche Meinungsbildung dadurch zu vergrößern, dass kommerzielle Sender zum Zuge kamen. Das geschah, aber auch in dieser Konkurrenz konnte sich der ÖRR letztlich gut behaupten.

Höchst erstaunlich also, wie diese Konstruktion in all den Jahrzehnten so wachsen und gedeihen konnte. Doch nun scheint die Erfolgsgeschichte an ihr Ende gekommen zu sein. Dies liegt vor allem an einer völlig veränderten Mediennutzung: Das Web 1.0 mit der stationären Nutzung von Websites seit Mitte der 1990er Jahre und das Web 2.0 mit der mobilen Nutzung von Social Media ab Mitte der 2010er Jahre stellen die Kommunikation überall und insgesamt auf den Kopf und damit auch den ÖRR. Es schwindet die Nachfrage nach linear ausgestrahltem Hörfunk und Fernsehen. Selbst immer mehr Ältere brechen mit ihren Gewohnheiten und präferieren flexiblere, individuellere, attraktivere Quellen für Information und Unterhaltung, als sie der ÖRR mit seinen Mediatheken bietet. Ein gefährlich großer Teil der Bevölkerung kommt gut ohne die digitalen Angebote des ÖRR aus. Und zugleich schwindet die Akzeptanz des ÖRR in den politischen Eliten, wie der Streit um Erhöhung der nun „Beitrag" genannten Gebühr zeigt – befeuert durch Enthüllungen zur Selbstbedienungsmentalität einiger Führungsfiguren des ÖRR.

Viele suchen nach Auswegen aus dieser Akzeptanzkrise. Bei dieser Suche ist Karsten Rudolph keine allzu große Hilfe. Denn den Großteil seines Buches verwendet er dafür, die Verästelungen der Medienpolitik vor 40, 60 oder 80 Jahren zu rekonstruieren. Das im Detail nachzulesen, ist weder amüsant noch lehrreich. Zumal er keine neuen Quellen erschließt, sondern sich auf medienhistorische Standardliteratur stützt, ohne dabei aber eine neue Sichtweise auf den Wandel zu eröffnen. Das Nacherzählen der vielen Scharmützel kostet viel Zeit und Raum, so dass nur wenige abschließende Seiten dafür bleiben, Ratschläge zu geben, wie es weiter gehen soll und wie nicht.

Und guter Rat ist dringend erforderlich. Denn es geht ja um mehr als um Gebühren, Gehälter und Pensionen. Es geht um eine Berichterstattung, die sich an klassischen journalistischen Standards orientiert und der gemeinhin immer noch stärker vertraut wird als anderen Quellen. Es geht um ein Rückgrat der Öffentlichkeit in Deutschland. Und hoffnungslos ist die Lage des ÖRR sicher nicht. Es spricht einiges dafür, dass der ÖRR auch die Akzeptanzkrise überstehen wird. Schließlich hat er über 80 Jahre lang eine enorme Resilienz aufbauen können. Vor allem ist er im Unterschied zu Tageszeitungen ökonomisch abgesichert. Trösten kann vielleicht auch eine Erkenntnis der Mediengeschichte (Riepl 1913; Vowe & Henn 2025): Noch nie ist ein Medium vollkommen verschwunden, sofern es die Kraft hatte, sich an veränderte Bedingungen anzupassen. Selbst die Steintafel, gut 6000 Jahre alt, überdauert bis heute als Grabstein.

Fernsehen und Hörfunk und damit in Deutschland der ÖRR werden sich im Sturm der digitalen Disruption behaupten, wenn sich der ÖRR schneller und gründlicher wandelt als bislang. Noch wird seine Informationsleistung geschätzt. Fakten professionell zu überprüfen und Informationen zu kuratieren, ist arbeitsintensiv und bisher die Domäne klassischer Medienanbieter. Doch da droht schon die nächste Klippe: Dank KI ersetzt eine algorithmisch gesteuerte Informationsvermittlung Stück für Stück die redaktionell gesteuerte. Und noch wird der ÖRR auch deswegen anerkannt, weil er in einer individualisierten Welt Gemeinschaft stiftet: Noch versammelt sich die Nation rund um seine massenmedialen Angebote, bei großen Sportereignissen etwa oder bei Bundestagswahlen. Wobei dann parallel dazu in vielen kleinen Online-Gruppen über das Gesehene geredet wird. Aber auch dafür deuten sich bereits KI-basierte Lösungen an. Der ÖRR wird also rasch viel lernen müssen. Und erstmals in den 80 Jahren wird er lernen müssen zu schrumpfen.

Gerhard Vowe

Prof. Dr. Gerhard Vowe, Heinrich-Heine-Universität Düsseldorf, Seniorprofessur Kommunikations- und Medienwissenschaft, Universitätsstraße 1, 40225 Düsseldorf, Deutschland, vowe@uni-duesseldorf.de

Literatur

Riepl, W. (1913): Das Nachrichtenwesen des Altertums mit besonderer Rücksicht auf die Römer. Leipzig: Teubner

Vowe, G., & Henn, P. (2025). Öffentliche Kommunikation im Übergang zum Internetzeitalter. Beitrag für das Dossier "Medienpolitik" der Bundeszentrale für politische Bildung. https://www.bpb.de/themen/medien-journalismus/medienpolitik/572882/oeffentliche-kommunikation-im-uebergang-zum-internetzeitalter/

Zeitschriftenlese

AfP
Jg 56 (2025) Nr 6

Manns, Luca: Presserechtliche Ansprüche gegenüber den Nachrichtendiensten des Bundes. – S. 469–474

Mitzscherlich, Johann: Das Recht an der eigenen Stimme – Teil 2. Schutz vor ungewollter Kommerzialisierung und KI-Imitation. – S. 475–484

Hermann, Tobias: Kein Grundrechtsschutz für bekannte Personen bei satirischen Werbeanzeigen? Der BGH als „versixter" Geisterfahrer – eine kritische Bestandsaufnahme 20 Jahre nach dem Lafontaine-Urteil. – S. 484–489

Teichmann, Fabian: Phishing und Social Engineering im Mediensektor. Prävention und rechtliche Herausforderungen. – S. 490–499

Communicatio Socialis
Jg 58 (2025) Nr 4

Bohrmann, Thomas: Die Moral der Bilder. Ethische Fragen im Spielfilm. – S. 474–487

Wiedemann, Thomas: Privilegierte Blickwinkel. Zu den Deutungen von Welt in deutschen Kinospielfilmen. – S. 488–498

Radziwill, Sophie; Stüwe, Julia: Die Ethik der Lücke. Alter, Geschlecht und Sichtbarkeit in der Filmproduktion. – S. 499–510

Groß, Bernhard: Aufarbeitung oder Erinnerung? Zum Verständnis von Vergangenheit und Gegenwart in Holocaustfilmen der BRD, DDR und Italien 1945-1990. – S. 511–522

Endres, Susanna: Bildung. Grundbegriffe der Kommunikations- und Medienethik (Teil 42). – S. 540–547

Volkmann-Schluck, Sonja: Interessenkonflikte transparent machen. Der Presserat hat seine Richtlinien zur Trennung von Tätigkeiten aktualisiert. – S. 548–553

Hepach, Sophie: Wollen kleine Köpfe große Themen? Kriterien und Empfehlungen für Qualitätsjournalismus für Kinder. – S. 554–567

Maier, Tanja: Papst-Bilder in „Der Spiegel" 1949–2025. Politiken der Sichtbarmachung im Wandel. – S. 568–580

Stütz, Kira: Posten gegen den Trend. Kommunikation in der digitalen Diaspora. – S. 581–594

Communication, Culture & Critique
Jg 18 (2025) Nr 4

Karim, Mariam A.: The Colonial Politics of British Publishing. Revisiting Matiel Mogannam's (1937) the Arab Woman and the Palestine Problem. – S. 302–309

El Masry, Noura et al.: Gender Hierarchies in Reporting Genocide. An Analysis of the Dehumanization of Palestinian Men in Western Media. – S. 310–321

Okeke, Dennis Ekwemnachukwu: Race, Violence in the US, and Digital (News) Discourse. A Corpus-Based Critical Discourse Analysis of Jordan Neely's Racialization. – S. 322–337

Proust, Valentina: Collective Care and Solidarity as Resistance. An Analysis of the Life in Leggings Movement and Caribbean Feminism. – S. 338–345

Mohammed, Wunpini Fatimata; Fawe, Charity Markus: Disrupting the Netflix Hegemony in Nollywood. The Case of A Tribe Called Judah. – S. 346–353

De Clerck, Amber; Rutten, Kris: Queen Nikkolah's Reign. Revealing Racial Denial in Flanders through a Rhetorical Listening Analysis of the Queen Nikkolah Debate. – S. 353–361

Computer Law Review
Jg 26 (2025) Nr 3

Albrecht, Daniel: Framework for Autonomous Vehicles in China – An Overview of the Evolving Changes. – S. 101–110

Lloyd, Ian: UK: The Data (Use and Access) Act 2025 – Data Protection and More. – S. 111–115

Jg 26 (2025) Nr 4

Bertrand, Alix et al.: The Impact of a Future Digital Fairness Act on Influencer Marketing. – S. 137–146

Computer und Recht
(2025) Nr 11

Becker, Maximilian; Birnstill, Pascal; Ebert, Andreas: Erklärbarkeit von KI-Systemen unter der KI-Verordnung. Vorgaben der KI-VO für Hochrisiko-KI-Systeme im Lichte aktueller Erklärverfahren aus dem Feld der Explainable AI (XAI). – S. 701–708

Nebel, Jens: Das Ende der Auftragsverarbeitung? Die Auswirkungen des EuGH-Urteils in Sachen

DOI: 10.5771/1615-634X-2026-1-106

Einheitlicher Abwicklungsausschuss auf die Notwendigkeit von Auftragsverarbeitungsvereinbarungen. – S. 711–718

Teichmann, Fabian: NIS-2-Schulungspflicht der Geschäftsleitung – Regulatorische Grauzonen und nationale Umsetzung. Eine Übersicht zum Handlungsbedarf in Unternehmen. – S. 718–725

Schiefer, David; Kremer, Sascha: Deepfakes in der Unternehmenskommunikation – Teil 4. Eine Einordnung aus haftungsrechtlicher Perspektive hinsichtlich des Urheberrechts und der KI-VO. – S. 762–768

(2025) Nr 12

Zech, Herbert: Vom IT-Recht zum Recht der Digitalisierung. – S. 769–775

Coenen, Jonas: Datenweitergabe durch Vertrag. Der Kontrahierungszwang im Data Act im Lichte bestehender Kontrahierungspflichten. – S. 775–782

Piltz, Carlo; Kukin, Ilia: „Naming and Shaming" durch Datenschutzbehörden – Dürfen in Pressemitteilungen zu Bußgeldern auch Namen der Unternehmen genannt werden? Öffentlichkeitsarbeit der Behörde als datenschutzrechtliche Sanktion. – S. 791–797

Schweinoch, Martin; Peintinger, Stefan: Die Rechtsprechungslinie des EuGH zur Reichweite des Personenbezugs von Daten. Wieviel Rechtssicherheit für Verantwortliche und Nicht-Verantwortliche folgt aus dem relativeren Personenbezug? – S. 797–805

Kreutzer, Till: Softwareurheberrecht zwischen Rechtsschutz und Nutzerautonomie. Eine Einordnung der BGH-Rechtsprechung in Sachen Action Replay und Werbeblocker. – S. 819–823

Hoeren, Thomas; Pinelli, Stefan: Quantencomputing und Recht – Herausforderungen für Datenschutz, Beweisrecht und Produkthaftung. Warum es normative Kriterien für den Umgang mit Unsicherheit zur Stärkung digitaler Rechtsstaatlichkeit braucht. – S. 834–840

(2026) Nr 1

Streitz, Siegfried; Schneider, Jochen: Urheberrechtschutz für Computerprogramme. Was unterscheidet heutige State-of-the-Art Softwarelösungen von Computerprogrammen der 80er Jahre? – S. 1–9

Sassenberg, Thomas; Hunstein, Dorothee: Der Einsatz von KI-Systemen zur Erfüllung der eigenen Leistungsverpflichtungen – Vertragliche Pflichten und Auswirkung auf die Vertragsgestaltung. – S. 10–16

Voigt, Paul; Schmalenberger, Alexander: Das BSI-Gesetz nach der NIS-2-Umsetzung – Überblick für die Praxis. Von der KRITIS-Regulierung zum umfassenden Cybersicherheitsrecht. – S. 17–25

Hartung, Jürgen: Die divergierende höchstrichterliche Rechtsprechung zum immateriellen Schadenersatz wegen Kontrollverlust nach Art. 82 DSGVO. Warum der Ansatz des BGH unvereinbar ist mit den Vorgaben des EuGH. – S. 25–32

Quiel, Philipp; Kukin, Ilia; Piltz, Carlo: Bestandskundenwerbung nach der „Inteligo Media" Entscheidung des EuGH. Die UWG-Vorgaben im Anwendungsbereich von Art. 13 Abs. 2 ePrivacy-RL. – S. 45–51

Käde, Lisa: Datenlos durch die Nacht – Von „Memorization" zur Vervielfältigung in KI-Modellen. Anmerkung zu LG München I - GEMA ./. OpenAI. – S. 67–72

Convergence
Jg 31 (2025) Nr 6

McGuigan, Lee: Direct Marketization. Platform Surveillance and the Contextual Integrity of Democratic Media. – S. 1781–1798

Lobato, Ramon: Automated Content Recognition (ACR), Smart TVs, and Ad-Tech Infrastructure. – S. 1799–1814

Hessler, Jennifer: The Ratings Panel from Participation to Personification. Nielsen and the Big-Data Reconfiguration of the Audience. – S. 1815–1831

Søltoft, Johan Irving; Munk, Anders Kristian: Recalcitrant Audiencing. Between Analytics and Creative Practice in the Film Industry. – S. 1832–1846

Saunders, Rebecca: Big Data on Pornhub Insights. Datafication and the Making of a New Sexual Culture. – S. 1847–1864

Liu, Anna Yan et al.: Stakeholder-Desired Audiences. Fans' Audience Data Imaginaries and How They Shape Industry Data Practices. – S. 1865–1885

Hesmondhalgh, David: Subjectivity, Culture and the Datafication of Music. – S. 1886–1900

Jg 32 (2026) Nr 1

Digital Journalism
Jg 13 (2025) Nr 9

Reveals About the Epistemology of Data Journalism. – S. 1599–1624

Mattis, Nicolas et al.: Nudging News Readers. A Mixed-Methods Approach to Understanding When and How Interface Nudges Affect News Selection. – S. 1647–1667

Amit-Danhi, Eedan R. et al.: An Ecosystem of Collective Futures. How Journalists and Experts Co-Construct Projections in Hybrid Media Environments. – S. 1668–1691

Jg 13 (2025) Nr 10

Thomson, T. J.; Thomas, Ryan J.; Matich, Phoebe: Generative Visual AI in News Organizations. Challenges, Opportunities, Perceptions, and Policies. – S. 1693–1714

Di Salvo, Philip: Dealing with the Black Box. European Journalists and the Threats of Spyware. – S. 1715–1734

Harkin, Diarmaid; Mann, Monique: Electronic Surveillance and Australian Journalism. Surveillance Normalization and Emergent Norms of Information Security. – S. 1735–1754

van der Lubben, Sebastiaan et al.: A Discourse Community of Livebloggers? Routines, Conventions, and the Pursuit of Credibility in Dutch Liveblogs. – S. 1755–1772

Parks, Perry: „Find the Joy". A War Correspondent's Tweets and the Rise of an Affective Age in News. – S. 1773–1795

Moore, Martin; Ramsay, Gordon Neil: Local News in National Elections. An „Audit" Approach to Assessing Local News Performance During a National Election Campaign. – S. 1796–1815

Lincoln, Louisa: „A Hard Road with Personal Costs". How Digital-First Nonprofit News Founders Navigate Precarity. – S. 1816–1834

Moon, Ruth et al.: „Scooped by the Town Drunk". Unpacking the Effects of COVID-19 on Rural Journalism Work. – S. 1835–1854

Reyna, Víctor Hugo: Mobilities, Immobilities and News Work. The New Normal of Award-Winning Journalists. – S. 1855–1872

European Journal of Communication
Jg 40 (2025) Nr 6

da Silva, Jaércio: The Emergence of Intersectional Multi Micro Media on the French Web. – S. 559–575

Mariotti, Claudia; Stanziano, Anna: To Polarize or Not to Polarize? Unveiling 2024 European Election Communication Strategies. – S. 576–597

Van den Bulck, Hilde et al.: Epistemic Welfare and Public Service Media's Algorithmic Recommender Systems. A Theoretical Framework, Operationalization and Relevance for Governance. – S. 598–618

Seigneurin, Marion; Balagué, Christine; Lyubareva, Inna: Navigating Misinformation and Disinformation. How Definition Ambiguity Limits the DSA's Implementation. – S. 619–646

Naderer, Brigitte; Niederkrotenthaler, Thomas; Till, Benedikt: Exploring Media Responses to Terrorist Attacks. A Content Analysis of Portrayed Mitigation Strategies to Terrorism in German and Austrian Newspapers. – S. 647–665

Human Communication Research
Jg 52 (2026) Nr 1

Peng, Tai-Quan: Human Communication Research at a Crossroads. Reaffirming Standards, Reimagining Possibilities. – S. 1–10

Skurka, Chris et al.: News Finds Me as the Illusion of Competence. Evidence for Overconfidence in Discernment of Political Misinformation. – S. 11–23

Dreston, Jana H.; Nanz, Andreas: Incidental and Intentional Exposure in a Mock Election Setting on Social Media. Effects on Subjective and Objective Political Knowledge and Aligned Voting. – S. 24–37

Li, Zixi; Shi, Jingyuan: Message Effects on Psychological Reactance. Meta-Analyses. – S. 38–52

Kreling, Rebekka J. et al.: What Do People Watch under Adversity? Testing Interactions of Semantic Affinity and Coping Style Using Netflix Data Donations. – S. 53–65

Javnost
Jg 32 (2025) Nr 4

Löblich, Maria; Pollack, Elisa: Unequal Media Worlds. East Berlin Media Criticism after 1990. – S. 417–434

Fiedler, Anke; Morrison, James: How Different Models of Media Regulation Address Social Inequality. A Qualitative Textual Analysis of Key Media Regulation Texts from Germany and the United Kingdom. – S. 435–455

Ferreira Fernandes, Inês: Censorship in the News. Understanding Social Inequalities in Portuguese Printed News in the Second World War. – S. 456–469

Venema, Niklas; Michael, Hendrik: High Expectations, Hostile Perception, and Biased Coverage? The Labour Movement and the Press Coverage of the Paris Commune, 1871. – S. 470–486

Ribeiro, Nelson: Broadcasting and the Creation of a Black Public Opinion in the Portuguese Empire. Using Radio to Promote and Counter Cultural Dominance. – S. 487–501

Battisti, Fabiana; Dalvit, Lorenzo: Children, Angels and Demons. The Coloniality of Print Media Representations of (Dis)ability in South Africa (1978–2023). – S. 502–520

Morrison, James: Decay, 'Dependency' and Deindustrialisation. The Evolving Discourse of 'The Left Behind'. – S. 521–538

Journal of African Media Studies
Jg 17 (2025) Nr 2

Osmondi, Grace; Diers-Lawson, Audra: Visual Frames of Resilience. Crisis History's Influence in the Visual Representation of Post-Conflict Somalia. – S. 105–129

Abubakar, Abdullahi Tasiu: The Risks and Resilience in Reporting Violent Extremism. – S. 131–147

Yue, Qi et al.: Analysis of Zimbabwean Students Studying in China from the Perspective of Afrokology. Breaking Through, Reconstruction and Integration. – S. 149–168

Agina, Añulika: The Reception of China's Startimes in Nigeria. – S. 169–182

Anoke, Ijeoma Nwakama; Stubbs, Jonathan: Women's Voices and the Representation of Africa in the BBC Documentary Sex for Grades. – S. 183–198

Nyamnjoh, Francis B.: The Digital Skull. Navigating Contemporary Romance and Values in a Reimagined the Palm-Wine Drinkard. – S. 199–211

Journal of Children and Media
Jg 20 (2026) Nr 1

Jenkins, Marina C. et al.: Association between Parent Technology Rules and Loneliness among Adolescents in the United States. – S. 1–12

Wald, Rebecca et al.: The Google Family Home. Smart Speaker Use in Dutch Families and the Impact of Parents' Individual and Contextual Characteristics. – S. 13–32

Piper, Douglas J. et al.: The Intersection of COVID Lockdown Policies and Parent Resources with Children's Media Exposure in Canada, the United States, Israel, Sweden, and New Zealand. – S. 33–59

Andriono, Maria Angela et al.: Exploring the Effectiveness of Mobile Learning on Early Literacy Skills of Kindergarteners in Indonesia. – S. 60–80

Stewart, Kate; Martins, Nicole: Brief Report – A Quantitative Content Analysis of Adolescence Depictions in Popular Coming-Of-Age Movies. – S. 81–91

De Wolf, Ralf et al.: Exploring Children's Perspectives on Parental Surveillance in Belgium. – S. 92–110

Griffith, Shayl F. et al.: Leveraging a Parenting Intervention to Support Healthy Media Use in U.S. Preschoolers with Externalizing Behavior. A Pilot Randomized Trial. – S. 111–131

Edgerly, Stephanie: Developing the Habit. The Socialization of U.S. Teens into Distinct Repertoires of News Consumption. – S. 132–150

Swit, Cara S.; Friesen, Myron D.; Xue, Xinyi: Predictors of Young Children's Problematic Screen Media Use in Aotearoa New Zealand. Testing the Interactional Theory of Childhood Problematic Media Use Model. – S. 151–172

Johnson, Opeyemi Victoria; Hust, Stacey J. T.: Empowering Narratives. Understanding Consent, Personal Boundaries, and Body Autonomy in US Children's Picture Books. – S. 173–185

Panneerselvam, Theerthaana; Waheed, Kareem Abdul; Srikanth, Snehalathaa: Going Gray for Gains. Exploring the Role of Monochrome Displays in Enhancing Children's Well-Being in India. – S. 186–198

Journal of Communication
Jg 75 (2025) Nr 5

Minich, Matt et al.: Pictorial Warning Labels Reduce Sharing Intentions, Blunt Self-Relevance

Processes Elicited by Social Media Posts Promoting Cannabis Edibles. – S. 321–334

Huang, Vincent Guangsheng; Lyu, Yuexin: Diffusion as Translation. Transediting and the Diffusion of Social Movement Discourses across Linguistic Borders in the Age of Globalization. – S. 335–347

Ma, Rong et al.: Words That Trigger. A Meta-Analysis of Threatening Language, Reactance, and Persuasion in Health. – S. 348–357

Chung, Sungeun; Lee, Byeong-Hyeon; Kim, Youllee: Diagnosticity of Perceived Message Effectiveness in Campaign Message Pretesting. Multilevel Analysis of the Between-Message Correlation and Message-Pair Standing Comparisons. – S. 358–370

Ali, Christopher; Van den Bulck, Hilde; Kropko, Jonathan: An Island of Trust. Public Broadcasting in the United States. – S. 371–384

Nothias, Toussaint: An Intellectual History of Digital Colonialism. – S. 385–397

Journal of Communication Inquiry
Jg 50 (2026) Nr 1

Matsilele, Trust; Hlungwani, Trevor: Factional Journalism in South Africa's Privately Owned Print Media. An Exploration. – S. 9–29

Andrews, Karen; Jaffer, Safeera; Shariff, Shaheen: Powerful yet Disempowered. A Thematic Literature Review Exploring the Challenges of Media Reporting on Sexual Violence. – S. 30–51

Velloso, Carolina: Traditional Journalism Norms Revisited. Journalistic Reconceptualizations of Objectivity. – S. 52–72

Ismail, Amani; Torosyan, Gayane: Press Freedom, State Interests, and a Murder Case. Editorial Coverage of Jamal Khashoggi in the Washington Post. – S. 73–96

Mathews, Nick; Wolgast, Stephen: „Shared Press Distress" and the Police Raid on a Newspaper. – S. 97–113

Marinos, Martin: Ethnography of Media Production. An Endangered Species? – S. 114–134

Journal of Computer-Mediated Communication
Jg 30 (2025) Nr 6 (ohne Seitenzahlen)

Zoizner, Alon; Levy, Avraham: How Social Media Users Adopt the Toxic Behaviors of Ingroup and Outgroup Accounts.

Hall, Jeffrey A.; Pennington, Natalie: How Change in Sociality over Time Moderates the Association between Mobile and Social Media Use and Well-Being.

Chen, Y. Anthony; Mares, Marie-Louise: Social Media Self-Presentation of LGBTQ+ Youth in the United States. The Role of Identity Exploration, Context Collapse, and Supportive Feedback.

Kang, Hyunjin et al.: Latent Privacy Management Profiles on Algorithmic Social Media. Cross-Cultural Insights into Privacy Protection Motivations and Management Behaviors.

Journal of Health Communication
Jg 30 (2025) Nr 7–9

Ratzan, Scott C. et al.: Advancing Health Communication, a Call for the Future. 30 Years of the Journal of Health Communication – International Perspectives. – S. 232–237

Hust, Stacey J. T. et al.: Washington State Teens' Perceptions of Cannabis-Infused Product Packaging. A Qualitative Study. – S. 238–246

Kaden, Jana: Informed Decision-Making about COVID-19 Vaccination – Development and Feasibility of a Decision Support Intervention. A Mixed-Methods Study. – S. 247–261

Kim, Rayun et al.: Identifying Reference Groups in Social Norms Campaigns Intended to Promote College Students' Mental Health Help-Seeking in South Korea. – S. 262–282

Jg 30 (2025) Nr 10–12

Wang, Buduo; Shibo, Bruce Wang; Kafle, Jiwan: When ChatGPT Speaks About Health. Examining Perceptions of Warmth and Competence Toward AI as a Health Information Source. – S. 285–295

Yi, Victoria N. et al.: Comparing Large Language Models as Health Literacy Tools. Evaluating and Simplifying Texts on Gender-Affirming Surgery. – S. 296–314

Shi, Jingyuan; Xu, Kun; Chen, Xiaobei: Tweaking the Messages and Approaching the Glass Box.

Wright, Paul J.: Sex Education, Public Opinion, and Pornography. Replication and Experiment. – S. 602–617

Kim, Kayoung; Lee, Chul-Joo; Jeong, Se-Hoon: Narrative Persuasion by Celebrity Source. The Moderating Role of Involvement. – S. 618–629

Journal of Media Psychology
Jg 38 (2026) Nr 1

Adam-Troian, Jais: The Fitness-Validation Model of Propagandist Persuasion and the 5Ds of Propaganda. – S. 1–13

Bailey, Rachel L. et al.: Eating Point-of-View in ASMR Videos Alters Motivational Outcomes. – S. 14–27

Jansen, Martin-Pieter; van Ooijen, Iris: For Your Eyes Only? An Eye-Tracking Experiment Investigating Microtargeting Transparency, Visual Attention, and Critical Processing. – S. 28–40

Teunissen, Lauranna; Decorte, Paulien; Cuykx, Isabelle: The Food Media Content Gratifications Scale Conceptualization, Development, and Validation. Measuring What Audiences Seek from and Find in Food Media Content. – S. 41–55

Zhao, Danyang; Raney, Arthur A.: Mood Enhancement, Self-Transcendence, and Prosociality. Selective Exposure to Self-Transcendent Media Entertainment and Prosocial Behavior. – S. 56–67

Journalism & Mass Communication
Jg 102 (2025) Nr 4

Kang, Seok; Valadez, Kayla: Deepfakes' Cognitive, Emotional, and Behavioral Impact. A Systematic Review and Meta-Analysis of Individual Responses. – S. 958–992

Rasul, Muhammad Ehab et al.: „It's All Fake News!". How Perceptions of Misinformation and Disinformation Influence News Consumption across Traditional Media, Social Media, and AI. – S. 993–1019

Weikmann, Teresa; Egelhofer, Jana Laura; Lecheler, Sophie: Beyond Credibility. The Effects of Different Forms of Visual Disinformation. – S. 1020–1043

Tang, Rita Rongwei et al.: Automating Accuracy. Scalable Approaches to Correcting Disinformation with Artificial Intelligence on Social Media. – S. 1044–1070

Sun, Yibing et al.: Data Visualization or Visual Exemplars? Testing the Differential Effects of AI-Generated Visual Correction Enhancements. – S. 1071–1101

Zhang, Bingbing; Kim, Sang Jung; Scott, Alex: Immunizing the Public Against AI-Generated Disinformation. Testing the Effects of Inoculation Mode and Issue Attitude on Inoculation Likelihood of Political Deepfakes. – S. 1102–1134

Vogler, Daniel; Rauchfleisch, Adrian; de Seta, Gabriele: The Short-Term Impact of an On-Site Literacy Intervention on Discerning Deepfake Videos Based on Visual Features. – S. 1135–1156

Journalism Studies
Jg 26 (2025) Nr 15

Tong, Jingrong: Serving the Public Interest? A Computational Analysis of the Topics of UK National Newspaper Coverage Using Freedom of Information (FOI) Requests Between 2005 and 2023. – S. 1803–1827

Palmer, Ruth; Paul, Subin: Is it Just Fluff? Topics, News Values, and Normative Democratic Implications of BBC's The Happy Pod. – S. 1828–1850

Renaud, Pauline: A Game Worth the Candle? French Journalists and Their Illusio in the Face of Disinformation. – S. 1851–1870

Kopelman, Sara: The Insistent Image. The Photojournalistic GIF as a Storytelling Form in Online News. – S. 1871–1891

Koo, Gyo Hyun; Coleman, Renita: Solutions-Oriented Framing and Responsibility Attribution in Urban Youth Crime. Effects on Policy Support and the Mediating Role of Emotions and Social Inequality Perception. – S. 1892–1913

Perreault, Gregory; Ohme, Jakob: ChatBots as Artificial Intermediaries? Adaptation to Artificial Intelligence in Newsrooms. – S. 1914–1935

Riemann, Robin; Ehrlich, Gianna L.: Up or Out? The Role of Audience Analytics in Post-Publication Gatekeeping. – S. 1936–1958

Malik, Saadia Izzeldin: Foregrounding Collective Trauma. Experiences of Palestinian Journalists Covering the 2023 War in Gaza. – S. 1959–1980

Jg 26 (2025) Nr 16

Diamantopoulos, Mitch: Theorizing „Co-operative Advantage" in News-Markets. Rethinking Media-Ownership, Renewing the Sociology of

Jg 27 (2026) Nr 1

Journalistik
Jg 8 (2025) Nr 3–4

Kommunikation & Recht
Jg 28 (2025) Nr 12

Datenschutz und digitale Souveränität bei Microsoft 365. – S. 755–760

Piltz, Carlo; Kukin, Ilia: Private Nutzung betrieblicher E-Mail-Postfächer. Aktuelles zur Frage der Geltung des Fernmeldegeheimnisses im Beschäftigungsverhältnis. – S. 761–765

Rippert, Stephan; Fischer, Anne-Sophie; Issels, Maximilian: Wem gehört das Glasfasernetz? – S. 765–772

Thiele, Clemens: Länderreport Österreich. – S. 772–774

Jg 29 (2026) Nr 1

Ettig, Diana: Die Entwicklung des Presserechts 2025. – S. 1–5

Himmelsbach, Gero; Gößwein, Arthur: Urheber(persönlichkeits)rechtliche Grenzen des Redigats bei Tageszeitungen. – S. 5–8

Barwick, Susanne: Der US-Fall Bartz vs. Anthropic. Was Urteil und Vergleich für Rechteinhaber bedeuten. – S. 9–12

Sprenger, Tim; Kahl, Jonas: Der Schutz von Abgeordneten im äußerungsrechtlichen Streit – Braucht die Indemnität ein Update? – S. 12–17

Kreitz, Benedict: Der Einsatz personenbezogener Daten als Trainingsdaten für Künstliche Intelligenz. Ein Beitrag am Beispiel der Automobilindustrie im Lichte der aktuellen Rechtsprechung. – S. 17–23

Kochinke, Clemens: Goldrausch im medienrechtlichen Umfeld der USA. – S. 23–24

Mass Communication & Society
Jg 29 (2026) Nr 1

Gondwe, Gregory et al.: Misinformation and Digital Inequalities. Comparing How Different Demographic Groups Get Exposed to and Engage with False Information. – S. 1–15

Dalisay Francis S.; Yamamoto, Masahiro: News Literacy and Critical Thinking in the Pacific. Evidence from Guam and the Northern Mariana Islands. – S. 16–38

Yoo, Joseph J.; Johnson, Thomas J.; Lacasa-Mas, Ivan: The Dynamics of Misinformation Sharing. The Mediated Role of News-Finds-Me Perception and the Moderated Role of Partisan Social Identity. – S. 39–65

Yeo, Sara K. et al.: Understanding Knowledge Among White, Black, and Hispanic Audiences. Media Attention and Inequities in Factual and Perceived Knowledge. – S. 66–100

Corbu, Nicoleta et al.: Illusory Superiority About Misinformation Detection and Its Relationship to Knowledge and Fact-Checking Intentions. Evidence from 18 Countries. – S. 101–120

Kim, Young; Lim, Hyunji Dana: „Truth Sandwich" To Dispel Crisis Misinformation. Exploring the Separate and Combined Impacts of Inoculation, Narratives, and Stealing Thunder on Correcting Crisis Misinformation on Social Media. – S. 121–141

Chan, Michael; Vaccari, Cristian; Yamamoto, Masahiro: Cognitive Drivers of Misinformation Belief and Sharing on Social Media. A Cross-National Comparison. – S. 142–160

Media Perspektiven
(2025) Nr 35

Maurer, Torsten; Weiß, Hans-Jürgen: Das Informationsangebot von Das Erste und RTL. Ergebnisse der ARD-Programmanalyse 2024. – S. 1–21

(2025) Nr 36

Doh, Michael et al.: Medienumgang von Menschen ab 60 Jahren. Ergebnisse der SIM-Studie 2024. – S. 1–21

(2025) Dokumentation I

Medienrechtliche Staatsverträge in der Fassung des Fünften Medienänderungsstaatsvertrages. – S. 1–165

(2025) Dokumentation II

Medienrechtliche Staatsverträge in der Fassung des Sechsten Medienänderungsstaatsvertrages – S. 1–170

(2025) Dokumentation III

Medienrechtliche Staatsverträge in der Fassung des Siebten Medienänderungsstaatsvertrages (Reformstaatsvertrag). – S. 1–189

(2026) Nr 1

Burggraf, Jürgen: Paradigmenwechsel in der EU-Mediengesetzgebung. EU immer zentraler als Akteur der Gestaltung der europäischen und mitgliedstaatlichen Medienordnung. – S. 1–16

Kirschtein, Claudia: Online-Lernen und gleichzeitig fühlen? Das Verständnis von Emotionen in Theorie und Praxis. – S. 133–145

Biehl, Brigitte: „What the Aktual F*ck". Affektive Lernprozesse in der Identitätsarbeit im Bereich Führung mit TV-Serien. – S. 146–158

medien & zeit
Jg 40 (2025) Nr 2

Hammer, Hilke: „Proletarische Frauen, seid bereit!" Antimilitaristische Agitation proletarischer Frauen in der Zeitschrift Die Gleichheit (1892-1923). – S. 4–13

Irler, Jakob: „Spectacle of Broken Cities and Broken Lives". Media Innovations in U.S. Journalistic Reporting of the Spanish Civil War. – S. 14–23

Kuduzovic, Hanna: Digital Memory Practices and Platform Narratives. How TikTok Remembers the Srebrenica Genocide. – S. 24–33

Biringer, Katharina: „Nein, das hat es bei uns noch gar nicht gegeben." (Birgit, Z 116) Aufwachsen mit Medien in einer prä-digitalen Zeit. – S. 34–45

Mediendiskurs
Jg 29 (2025) 2

Schwartz, Maria: „Wer die Rache sucht, gräbt zwei Gräber." Rachemotive in Computerspielen. – S. 18–21

Künzel, Christine: Ausgleichende Gerechtigkeit? Rachenarrative im Rape-and-Revenge-Film. – S. 28–31

Völzmann, Berit: Revenge Porn? Der Schutz vor bildbasierter sexualisierter Gewalt. – S. 32–37

Hestermann, Thomas: Mediale Verzerrung so hoch wie noch nie. Wie deutsche Leitmedien über Gewalt berichten. – S. 68–71

Witter, Stefanie; Schwab, Frank; Unz, Dagmar: Toleranz statt Schlagabtausch. Was Talkshows von Dilemmadiskussionen lernen können. – S. 72–77

Meese, Christina: Jugendmedienschutz zwischen Brüssel und Berlin. – S. 80–85

MedienJournal
Jg 49 (2025) Nr 1

Kaltenbrunner, Andy et al.: Austrian Media Funding from the Perspective of Innovation in Journalism. An Analysis of National Programmes and a Comparison with Innovation Funding Strategies in Other European Countries. – S. 1–30

Frech, Jannis: Dimensionen und Herausforderungen eines transformativen Redaktionsumbaus. Der „Aktivierende Journalismus" des Stern. – S. 31–53

Paulitsch, Luis: „Gegenseitig vorantreiben und zugleich korrigieren". Zum symbiotischen Verhältnis zwischen der FPÖ und „alternativen Medien". – S. 54–65

Jg 49 (2025) Nr 2

Adolf, Marian: Von der Objektivität der Erkenntnis zur reflexiven Normativität. Perspektiven für eine öffentliche Kommunikationswissenschaft. – S. 17–38

Prinzing, Marlis; Robertson-von Trotha, Caroline Y.: Öffentliche Medien- und Kommunikationswissenschaft als Public Service. Anregungen zu einer notwendigen Debatte. – S. 39–58

Pfeiffer, Juliane; Damerau, Frederik; Pollak, Christopher: Interdisziplinarität als Kern von Transdisziplinarität. Erkenntnisse und Erfahrungen aus der Forschung zu Demokratie- und Journalismuskompetenz. – S. 59–76

Scheu, Andreas M.; Fischer, Liliann; Wigger, Julia: Annäherung von Forschung und Praxis im Bereich Wissenschaftskommunikation. Erfahrungen aus dem transdisziplinären Verbundprojekt Transfer Unit Wissenschaftskommunikation. – S. 77–90

Oehmer-Pedrazzi, Franziska; Pedrazzi, Stefano: Forschung mit, über und für die Bürger:innen? Ein Erfahrungsbericht zum Mehrwert und den Herausforderungen von Citizen Science für die Kommunikationswissenschaft. – S. 91–108

Vogler, Daniel et al.: 15 Jahre gelebte öffentliche Kommunikationswissenschaft. Erfahrungen und Lehren im Rahmen des Jahrbuchs Qualität der Medien in der Schweiz. – S. 109–123

Krüger, Uwe: Zu Risiken und Nebenwirkungen der Transformativen Kommunikationswissenschaft. Lehren aus dem Projekt „Reklame für Klimakiller". – S. 124–140

Jg 49 (2025) Nr 3

Asadi Pour, Fatemeh: Perceptions of Gender Portrayals in Movies and Series. Their Perceived Effects on Gender Identities and Sexual Orientation. – S. 9–37

Mobile Media & Communication
Jg 14 (2026) Nr 1

Multimedia und Recht
Jg 28 (2025) Nr 11

Jg 28 (2025) Nr 12

Jg 29 (2026) Nr 1

Werner, Daniel: Plug and Play für Datennutzung? Der Data Act und sein neues Vertragsregime. – S. 9–15

Wäger, Katharina: Das verfassungsrechtliche Doppeltürmodell bei strafprozessualen Datenabfragen. Defizite der Auskunftsnormen. – S. 15–20

Hüger, Jakob; Radtke, Tristan: Anbieter als Haftungs- und Sanktionssubjekt nach der KI-VO. Gestaltungsspielräume und Umgehungskonstruktionen. – S. 21–27

Carathanassis, Fay; Schillmöller, Jan: Sperrung von @elonmusk? Gesetzlich vorgesehene Account-Sperren vor dem Hintergrund von GgdG-E und DSA. – S. 27–32

New Media & Society
Jg 27 (2025) Nr 11

Hietajärvi, Lauri et al.: Latent Profiles of Adolescents' Digital Skills across Six European Countries. – S. 5915–5937

Schindler, Johanna et al.: „Check This Out!". Collective Functions of Instant Messaging about Media Content. – S. 5938–5961

Bulat, Beril et al.: The Psychology of Volunteer Moderators. Tradeoffs between Participation, Belonging, and Norms in Online Community Governance. – S. 5962–5985

Helsper, Ellen Johanna; Veltri, Giuseppe Alessandro; Livingstone, Sonia: Parental Mediation of Children's Online Risks. The Role of Parental Risk Perception, Digital Skills and Risk Experiences. – S. 5986–6005

Boulianne, Shelley; Hoffmann, Christian P.; Bossetta, Michael: Social Media Platforms for Politics. A Comparison of Facebook, Instagram, Twitter, YouTube, Reddit, Snapchat, and WhatsApp. – S. 6006–6037

Nguyen-Thu, Giang; Munn, Luke: Taming the Algorithm. The Platform Realism of GrabBike Delivery Workers. – S. 6038–6056

Hong, Joo-Wha; Cruz, Ignacio Fernandez; Kim, Donggyu: Justice behind the Virtual Mask. The Influence of Race of the Virtual Influencer and the Creator on Promoting the Black Lives Matter Movement. – S. 6057–6078

Burai, Krisztina; Solti, Ábel; Bene, Márton Bene: Feel Local, Post Local. An Ethnographic Investigation of a Social Media-Based Local Public. – S. 6079–6096

Strikovic, Edina et al.: Opportunity Structures for User Acceptance of News Recommender Systems (NRS). A Multi-Country Survey Study of Relationships between Individual-Level Factors and Evaluations of NRS. – S. 6097–6118

Dereymaeker, Julie et al.: Smarter Homes, Smarter Surveillance? Exploring Intimate Surveillance Practices in Modern Day Households. – S. 6119–6138

Li, Luzhou; Zhou, Kui: When Content Moderation Is Not about Content. How Chinese Social Media Platforms Moderate Content and Why It Matters. – S. 6139–6155

Tan, Yan; Liu, Shih-Diing: „Our Advice Is to Break Up". Douban's Intimate Public and the Rise of Girlfriend Culture. – S. 6156–6177

Butkowski, Chelsea P.: Record, Revise, Reinvent, and Resist. The Politics of Social Media Self-Representation during the COVID-19 Pandemic. – S. 6178–6197

Zhou, Zhiqiu Benson: Performing Lowbrowness. How Chinese Queer People Negotiate Visibility on Short-Video Platforms. – S. 6198–6214

Suenzo, Facundo N.: The Quest for Meaningful Connections. Navigating Scripted Sexuality on Instagram amid Grindr and Tinder Fatigue. – S. 6215–6232

Blum, Irene; Uldam, Julie: Faking, Optimising and Conceding to Power. Social Movement Understandings of Social Media Power. – S. 6233–6251

Mathé, Jan Jasper; Bauwens, Jo; Verstrynge, Karl: Being and Becoming in the Culture of Immediacy. An Existential-Ethical Approach. – S. 6252–6272

Golan, Oren; Stadler, Nurit: Digital Mary. Religious Mediatization and the Re-Enchantment of a Mega Symbol. – S. 6273–6294

Setty, Emily; Dobson, Emma: Young People's 'Post-Digital' Relationships during COVID-19 'Lockdowns' in England. – S. 6295–6314

Pinch, Annika et al.: „Is It Time for Me to Be Authentic?". Understanding, Performing, and Evaluating Authenticity on Bereal. – S. 6315–6333

Cazzamatta, Regina: Global Misinformation Trends. Commonalities and Differences in Topics, Sources of Falsehoods, and Deception Strategies across Eight Countries. – S. 6334–6358

in Social Media Conversations. An AI-Powered Computational Approach. – S. 99–118

Li, Mengyu et al.: Platform Affordances, Discursive Opportunities, and Social Media Activism. A Cross-Platform Analysis of #MeToo on Twitter, Facebook, and Reddit, 2017–2020. – S. 119–147

Laba, Nataliia: Beyond Magic. Prompting for Style as Affordance Actualization in Visual Generative Media. – S. 148–168

Sharabi, Liesel L.; Von Feldt, Paige A.; Ha, Thao: Burnt Out and Still Single. Susceptibility to Dating App Burnout over Time. – S. 169–190

Chae, Je Hoon; Tewksbury, David: Perceiving AI Intervention Does Not Compromise the Persuasive Effect of Fact-Checking. – S. 191–211

Van Bruyssel, Sara; De Wolf, Ralf; Vanden Abeele, Mariek: From Bliss to Burden. An Ethnographic Inquiry into How Social, Material and Individual Obstacles to Digital Well-Being Shape Everyday Life. – S. 212–229

Zhang, Maggie Mengqing; Ng, Yee Man Margaret: Beyond Dislike Counts. How YouTube Users React to the Visibility of Social Cues. – S. 230–249

Sharon, Tzlil; John, Nicholas A.: „It's between Me and Myself". Inverse Parasocial Relationships in Addressing (Imagined) Podcast Listeners. – S. 250–269

Deng, Ziyu: „They Don't Mean to Hurt". Female Gamers' Reluctance in Recognizing and Confronting Sexism in Gaming as an Online-Offline Juxtaposition. – S. 270–288

Ahmed, Saifuddin et al.: Pathways from Incidental News Exposure to Political Knowledge. Examining Paradoxical Effects of Political Discussion on Social Media with Strong and Weak Ties. – S. 289–311

Maes, Chelly; Van Ouytsel, Joris; Vandenbosch, Laura: Active Bystanders in the Forwarding of Sexting Messages. Applying a Theory of Planned Behavior in Youth. – S. 312–332

Brugman, Britta C.; van Huijstee, Dian; Droog, Ellen: Debunking the Corporate Paint Shop. Examining the Effects of Misleading Corporate Social Responsibility Claims on Social Media. – S. 333–356

Wischerath, Darja et al.: Rage Against the Machine. Exploring Violence and Emotion in Conspiracy Narratives on Parler. – S. 357–379

Trillò, Tommaso: „POV: You Are Reading an Academic Article." The Memetic Performance of Affiliation in Tiktok's Platform Vernacular. – S. 380–400

Roussel, Sarah; Restivo, Léa; Apostolidis, Thémistoklis: 'All Naked at the Gyno'. Psychosocial Approach to the Gynaecological Examination from Digital Media in a French Context. – S. 401–422

Nordicom Review
Jg 46 (2025) Nr 2

Sakariassen, Hilde: Multidimensional Digital Vulnerability among Older Adults. – S. 137 –160

Ruona, Mikaela Wikström; Sjölander, Annika Egan; Krutrök, Moa Eriksson: Dominant Discourses of Green Growth and the Swedish North. A Topic Model of the „Green" Transition in Swedish News Media, 2015–2023. – S 161–193

Åkerlund, Mathilda: The Co-Construction of Political Scandal in Hybrid Media. A Swedish Case Study. – S. 194–216

Steen-Johnsen, Kari; Fladmoe, Audun; Ihlebæk, Karoline Andrea: Right-Wing Alternative Media Users in Scandinavia. A Comparative Study of Ideological Attitudes and (Dis)Trust. – S. 217–249

Freudendal, Jakob: Rationalising Creative Uncertainty? The Negotiated Use of Audience Feedback in Danish Film Production. – S. 250 –272

Political Communication
Jg 42 (2025) Nr 6

Blair, Robert A. et al.: Depolarizing within the Comfort of Your Party. Experimental Evidence from Online Workshops. – S. 925–950

Kraitzman, Alon P.; DeMora, Stephanie L.; Albarracín, Dolores: Motivating Future Voters. Comparing the Effects of 'I Voted' and 'I Will Vote' Stickers on Intention to Vote. – S. 951–967

Klüser, K. Jonathan: From Entertainment to Engagement? Entertainment Figures' Political Messaging and Audience Responses in the Digital Age. – S. 968–991

Hayes, Danny et al.: The Local News Crisis and Political Scandal. – S. 992–1014

de León, Ernesto et al.: Misperceptions of Public Opinion During Crises. Evidence from the COVID-19 Pandemic. – S. 1015–1036

Jg 43 (2026) Nr 1

Publizistik
Jg 70 (2025) 4

SCM – Studies in Communication | Media
Jg 14 (2025) Nr 4

Studies in Communication Sciences
Jg 25 (2025) Nr 2

search on Politicians' Use of Social Media. – S. 77–90

Kaltschmitt, Tasmin et al.: „Hello Hackis, Stay at Home!" A Standardised Content Analysis on the Representation of Politics and COVID-19 in German Entertainment Podcasts. – S. 91–105

Reinbold, Judith: From 'War Hero' to 'Peace Hero'? The Bundeswehr's Portrayal of German Soldiers on the Peacekeeping Mission MINUSMA on Facebook. – S. 107–125

Zai, Florin; Wirz, Dominique: News Usage on Instagram. Frequency, Motives, Evaluation, and the Role of Political Interest. – S. 127–143

Burger, Johanna: No Local Newspapers but Whatsapp? Political Information Repertoires, Local Media Use and Their Effects in Switzerland. – S. 145–159

Jg 25 (2025) Nr 3

Monnier, Angeliki: Beyond the Discipline, beyond the Institution. The Outward Reach of French Information and Communication Sciences (SIC) through the Prism of Editorial Collaborations (2012–2018). – S. 167–183

Pereira, Fabio Henrique; Bilodeau Maxime: Qui Fabrique l'Information Scientifique au Québec? Analyse du Portrait des Journalistes Scientifiques 2023. – S. 185–206

Televizion
Jg 38 (2025) Nr 2

Krug, Maximilian: Rechte Influencerinnen im digitalen Raum. Visuelle Inszenierungen, emotionale Anschlussfähigkeit und Plattformlogiken. – S. 4–7

Zywietz, Bernd: „Online-Islamismus". Herausforderungen und Risiken religiöser und politaktivistischer Beeinflussung im Netz. – S. 8–11

Holler, Andrea; Götz, Maya: „TikTok ist die größte Informationsquelle, die ich habe". Wie Jugendliche mit Informationen von TikTok umgehen. – S. 12–17

Wenzel, Lukas; Holler, Andrea: Kommunikationsstrategien der AfD auf TikTok. – S. 18–22

Götz, Maya: „I Hear Your Voice". Detailanalyse eines TikTok-Clips von Alice Weidel. – S. 23–24

Breitweg, Jörg: Emotionale Radikalisierung. Was extremes Verhalten motiviert und Menschen steuert. – S. 29–33

Götz, Maya et al.: Können TikTok-Videos prodemokratische Werte vermitteln? Ein Experiment mit rechten Jugendlichen. – S. 34–39

Hampel, Kilian; Schnetzer, Simon; Hurrelmann, Klaus: „Zwischen persönlicher Hoffnung und gesellschaftlichem Pessimismus". Ergebnisse der Trendstudie „Jugend in Deutschland 2025". – S. 40–43

Zeitschrift für Urheber- und Medienrecht
Jg 69 (2025) Nr 11

Beyerle, Tulga: Was macht Design künstlerisch? Vortrag im Rahmen des Symposions „Was ist Kunst? – Design als Herausforderung für den Werkbegriff" des Instituts für Urheber- und Medienrecht am 4.7.2025 in München. – S. 741–746

de la Durantaye, Katharina: Gerichte als schlechte Kritiker – Urheberrecht an Gebrauchsgegenständen. Vortrag im Rahmen des Symposions „Was ist Kunst? – Design als Herausforderung für den Werkbegriff" des Instituts für Urheber- und Medienrecht am 4.7.2025 in München. – S. 747–755

Bullinger, Winfried: Ein Streit um den Brutalismus. Vortrag im Rahmen des Symposions „Was ist Kunst? – Design als Herausforderung für den Werkbegriff" des Instituts für Urheber- und Medienrecht am 4.7.2025 in München. – S. 755–758

Kur, Annette: Unité de l'art ohne cumul total (?) – das Verhältnis von Urheber- und Designrecht nach der Rechtsprechung des EuGH. Vortrag im Rahmen des Symposions „Was ist Kunst? – Design als Herausforderung für den Werkbegriff" des Instituts für Urheber- und Medienrecht am 4.7.2025 in München. – S. 759–767

Dobinsky, Hendrik: Beobachtungen zur Verfolgung von Produktnachahmungen in der Rechtspraxis unter besonderer Berücksichtigung des Urteils OLG Hamburg vom 29.4.2025 – 5 U 60/24 – Birkenstocksandale „Madrid" (ZUM 2025, 804). Vortrag im Rahmen des Symposions „Was ist Kunst? – Design als Herausforderung für den Werkbegriff" des Instituts für Urheber- und Medienrecht am 4.7.2025 in München. – S. 767–771

Schäffner, Steffen: Urheberrecht, Kunst und Design in der Praxis. Vortrag im Rahmen des Symposions „Was ist Kunst? – Design als Herausforderung für den Werkbegriff" des Instituts für Urheber- und Medienrecht am 4.7.2025 in München. – S. 772–774

Jg 69 (2025) Nr 12

Jg 70 (2026) Nr 1

Hinweise für Autor*innen

Die wissenschaftliche Vierteljahreszeitschrift „Medien & Kommunikationswissenschaft wird seit 1953 vom Leibniz-Institut für Medienforschung | Hans-Bredow-Institut herausgegeben und redaktionell betreut. Seit 2021 erscheint sie als Open-Access-Journal unter der Lizenz CC-BY.

Die Zeitschrift ist ein interdisziplinäres Forum für deutsch- und englischsprachige theoretische und empirische Beiträge aus der gesamten Medien- und Kommunikationswissenschaft.

Für die Publikation in „Medien & Kommunikationswissenschaft“ kommen wissenschaftliche Beiträge in Betracht, die:

- ein Moment originärer theoretischer Leistung beinhalten bzw. einen theoretisch weiterführenden Argumentationsgang bieten;
- Befunde zu einem ausgewiesenen Problem von theoretischer oder medienpraktischer Relevanz darstellen;
- innerhalb eines wissenschaftlichen Diskurses Position beziehen und die Diskussion voranbringen können oder
- Literatur bzw. ausgewählte Literatur zu bestimmten Problemstellungen systematisch und vergleichend zusammenfassen und eine Übersicht über den Stand der Theorie und/oder Empirie geben.

Die Redaktion bietet außerdem die Möglichkeit zur Stellungnahme und Erwiderung zu publizierten Beiträgen der oben genannten Kategorien. Stellungnahmen und Erwiderungen, die den in „Medien & Kommunikationswissenschaft“ üblichen inhaltlichen und formalen Standards entsprechen und geeignet sind, die wissenschaftliche Diskussion zu fördern, werden im nächstmöglichen Heft publiziert. Die Redaktion räumt dabei dem Autor bzw. der Autorin des Beitrages, auf den sich die Stellungnahme bezieht, die Möglichkeit einer Erwiderung ein.

Manuskripte, die zur Publikation in „Medien & Kommunikationswissenschaft“ eingereicht werden, dürfen *nicht anderweitig veröffentlicht* sein und bis Abschluss des Begutachtungsverfahrens *nicht anderen Stellen zur Veröffentlichung angeboten* werden.

Im Sinne der Förderung des wissenschaftlichen Diskurses und der kumulativen Forschung sowie der Qualitätssicherung legt die Redaktion bei der Begutachtung von Beiträgen besonderen Wert darauf, dass größtmögliche Transparenz hinsichtlich der verwendeten Daten hergestellt wird.

Autor*innen empirischer Beiträge verpflichten sich mit der Einreichung des Manuskripts, dass sie die Art und Weise der Datenerhebung bzw. den Zugang zu Datenbeständen, die von Dritten (z. B. Datenbanken) zur Verfügung gestellt worden sind, ausreichend dokumentieren, um so die Voraussetzungen für Sekundäranalysen und Replikationen zu schaffen. Zugleich erklären sie sich bereit, die verwendeten Daten bei wissenschaftlich begründeten Anfragen im Rahmen der jeweils gegebenen Möglichkeiten für weitere Analysen zur Verfügung zu stellen.

Im Hinblick auf eine transparente Dokumentation des methodischen Vorgehens entwickeln sich zunehmend standardisierte Kriterienkataloge. Dies ist in der Kommunikationswissenschaft insbesondere für die standardisierte Inhaltsanalyse der Fall. Aufsätze, in denen standardisierte Inhaltsanalysen dargestellt werden, sollten folgende Angaben enthalten: Grundgesamtheitsdefinition, Stichprobendesign, Aufgreifkriterium (inkl. Angabe der Identifikationsreliabilität), Design des Reliabilitätstests (Art der Reliabilitätsberechnung, Art der Stichprobe und Stichprobenumfang, Zeitpunkt, Zahl der Codierer, Begründung der Koeffizientenwahl), Zahl der bei der Codierung eingesetzten Codierer, Zahl der Ausprägungen und vorzugsweise zufallskorrigierte Reliabilitätskoeffizienten für alle dargestellten Variablen, Bewertung und Konsequenzen des Reliabilitätstests. Aufsätze, in denen andere Methoden

dargestellt werden, sollten sich in diesem Sinne an den der jeweiligen Methode entsprechenden Kriterien orientieren.

Formalien:

- Manuskripte sind der Redaktion per E-Mail im rtf-Format an die Adresse c.matzen@leibniz-hbi.de zuzuschicken.
- Da die eingereichten Manuskripte anonymisiert begutachtet werden, sind *zwei Titelblätter* erforderlich: eines mit Angabe des Titels und der Namen und Anschriften der Autor*innen, eines ohne Anführung der Namen und Adressen. Das Manuskript selbst darf keine Hinweise auf die Autor*innen enthalten.
- Beizufügen ist eine *kurze Zusammenfassung* des Beitrags (max. 15 Zeilen), die dem Leser als selbständiger Text einen hinreichenden Eindruck vom Inhalt des jeweiligen Beitrags vermittelt.
- Der Umfang der Beiträge soll *20 Manuskriptseiten* (55.000 Buchstaben inkl. Leerzeichen) nicht überschreiten.
- *Gliederung* des Textes: Jedes Kapitel und Unterkapitel sollte mit einer Überschrift (in Dezimalzählung) versehen sein.
- *Hervorhebungen* im Text sind kursiv oder fett zu kennzeichnen.
- *Hinweise* und *Literaturbelege* sollen erfolgen durch Angabe von Autor, Erscheinungsjahr und Seitenziffer im fortlaufenden Text – z. B.: . . . (Müller 1990: 37) . . . –, wobei der vollständige bibliographische Nachweis über ein *Literaturverzeichnis* im Anschluss an den Beitrag ermöglicht wird. Aufgrund der in verschiedenen Fachtraditionen unterschiedlichen Zitierregeln gibt M&K für das *Literaturverzeichnis* keine bestimmte Variante vor. Die einzige Vorgabe besteht darin, dass alle erforderlichen Literaturangaben vorhanden sind und dass diese für das gesamte Manuskript konsistent gestaltet werden.
- Artikel in M&K sollen in einer diskriminierungsfreien und gendergerechten Sprache verfasst werden. Den Autor*innen ist es freigestellt, hierfür eine geeignete Form zu wählen.

Über eine Annahme des Manuskripts und den Zeitpunkt der Veröffentlichung entscheidet die Redaktion auf der Grundlage redaktionsinterner und externer Gutachten. Dem/der Autor*in wird die Redaktionsentscheidung schriftlich mitgeteilt. Im Falle einer Entscheidung für Überarbeitung, Neueinreichung oder Ablehnung legt die Redaktion die Gründe für ihre Entscheidung offen.

Dazu werden die anonymisierten Gutachten, evtl. auch nur in Auszügen, zugesandt. Das Begutachtungsverfahren ist in der Regel acht Wochen nach Eingang des Manuskripts abgeschlossen; falls die Begutachtung längere Zeit erfordert, werden die Autor*innen benachrichtigt.

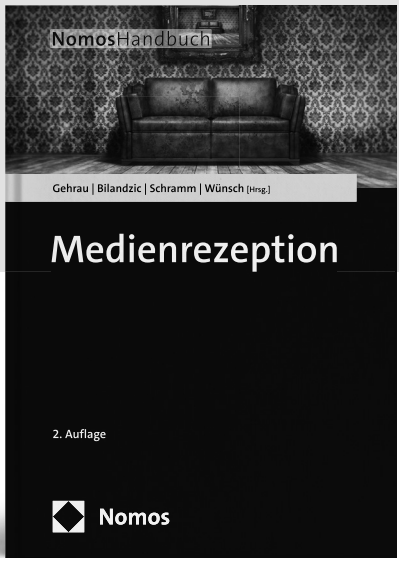
Den Autor*innen, deren Beiträge veröffentlicht werden, entstehen keinerlei Kosten. Nach Erscheinen erhalten die Autor*innen je ein Belegexemplar der gedruckten Ausgabe des Heftes, in dem ihr Artikel erschienen ist.

Verlag und Redaktion haften nicht für Manuskripte, die unverlangt eingereicht werden.

Anschrift der Redaktion: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut, Christiane Matzen, Warburgstraße 30 B, 20354 Hamburg, c.matzen@leibniz-hbi.de, Tel. 040/45 02 17-41.

Medien & Kommunikationswissenschaft (M&K)
ISSN 1615-634X
Herausgeber:
Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI), vertreten durch den Vorstand Kristina Hein, Prof. Dr. Judith Möller und Prof. Dr. Wolfgang Schulz
Schriftleitung:
Christiane Matzen, M.A. (V.i.S.d.P.)
Einsendungen bitte an:
Christiane Matzen, M.A. Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI), Warburgstraße 30 B, 20354 Hamburg, E-Mail: c.matzen@leibniz-hbi.de, www.muk.nomos.de
Manuskripte und andere Einsendungen:
Alle Einsendungen sind an die o. g. Adresse zu richten. Es besteht keine Haftung für Manuskripte, die unverlangt eingereicht werden. Sie können nur zurückgegeben werden, wenn Rückporto beigefügt ist. Die Annahme zur Veröffentlichung muss in Textform erfolgen. Mit der Annahme zur Veröffentlichung überträgt die Autorin/der Autor der Nomos Verlagsgesellschaft mbH & Co.KG an ihrem/seinem Beitrag für die Dauer des gesetzlichen Urheberrechts das einfache, räumlich und zeitlich unbeschränkte Recht zur Vervielfältigung und Verbreitung in körperlicher Form, das Recht zur öffentlichen Wiedergabe und Zugänglichmachung, das Recht zur Aufnahme in Datenbanken, das Recht zur Speicherung auf elektronischen Datenträgern und das Recht zu deren Verbreitung und Vervielfältigung sowie das Recht zur sonstigen Verwertung in elektronischer Form. Hierzu zählen auch heute noch nicht bekannte Nutzungsformen. Das in § 38 Abs. 4 UrhG niedergelegte zwingende Zweitverwertungsrecht der Autorin/des Autors nach Ablauf von 12 Monaten nach der Veröffentlichung bleibt hiervon unberührt. Eine eventuelle, dem einzelnen Beitrag oder der jeweiligen Ausgabe beigefügte Creative Commons-Lizenz hat im Zweifel Vorrang. Zum Urheberrecht vgl. auch die allgemeinen Hinweise unter www.nomos.de/urheberrecht.
Unverlangt eingesandte Manuskripte – für die keine Haftung übernommen wird – gelten als Veröffentlichungsvorschlag zu den Bedingungen des Verlages. Es werden nur unveröffentlichte Originalarbeiten angenommen. Die Verfasser erklären sich mit einer nicht sinnentstellenden redaktionellen Bearbeitung einverstanden.
Urheber- und Verlagsrechte:
Alle in dieser Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Der Rechtsschutz gilt auch im Hinblick auf Datenbanken und ähnlichen Einrichtungen. Kein Teil dieser Zeitschrift darf außerhalb der engen Grenzen des Urheberrechtsgesetzes oder über die Grenzen einer eventuellen, für diesen Teil anwendbaren Creative Commons-Lizenz hinaus ohne schriftliche Genehmigung des Verlags oder der Autor:innen in irgendeiner Form vervielfältigt, verbreitet oder öffentlich wiedergegeben oder zugänglich gemacht, in Datenbanken aufgenommen, auf elektronischen Datenträgern gespeichert oder in sonstiger Weise elektronisch vervielfältigt, verbreitet oder verwertet werden.
Namentlich gekennzeichnete Artikel müssen nicht die Meinung der Herausgeber/Redaktion wiedergeben.
Der Verlag beachtet die Regeln des Börsenvereins des Deutschen Buchhandels e.V. zur Verwendung von Buchrezensionen.
Anzeigen:
Verlag C.H.Beck GmbH & Co. KG, Media Sales, Dr. Jiri Pavelka, Wilhelmstraße 9, 80801 München
Tel.: (089) 381 89-687 687, mediasales@beck.de
Verlag und Gesamtverantwortung für Druck und Herstellung:
Nomos Verlagsgesellschaft mbH & Co. KG, Waldseestr. 3-5, 76530 Baden-Baden, Telefon: 07221/2104-0, Telefax 07221/2104-899, www.nomos.de
Geschäftsführer: Thomas Gottlöber
HRA 200026, Mannheim
Sparkasse Baden-Baden Gaggenau, IBAN DE05662500300005002266, (BIC SOLADES1BAD).
Erscheinungsweise: vierteljährlich
Bezugspreise für die Druckausgabe: Individualkunden: Jahresabo Druckausgabe € 124,–, Institutionen: Jahresabo € 239,–, Einzelheft: € 42,–; Vertriebskostenanteil (Inland € 20,80/Ausland € 37,50) und Direktbeorderungsgebühr € 3,50 jeweils zuzüglich. Alle Preise verstehen sich einschließlich der gesetzlichen Umsatzsteuer. Die Rechnungsstellung erfolgt nach Erscheinen des ersten Heftes des Jahrgangs.
Bestellungen über jede Buchhandlung und beim Verlag. Der Digitalzugang wird über die digitale Plattform des Verlags – Inlibra – bereitgestellt.
Kundenservice:
Telefon: +49-7221-2104-222 E-Mail: service@nomos.de
Kündigung:
Abbestellungen mit einer Frist von sechs Wochen zum Kalenderjahresende.
Adressenänderungen:
Teilen Sie uns rechtzeitig Ihre Adressenänderungen mit. Dabei geben Sie bitte neben dem Titel der Zeitschrift die neue und die alte Adresse an.
Hinweis gemäß Art. 21 Abs. 1 DSGVO: Bei Anschriftenänderung kann die Deutsche Post AG dem Verlag die neue Anschrift auch dann mitteilen, wenn kein Nachsendeauftrag gestellt ist. Hiergegen kann jederzeit mit Wirkung für die Zukunft Widerspruch bei der Post AG eingelegt werden.

Das neue Standardwerk zur Rezeptionsforschung



Volker Gehrau | Helena Bilandzic |
Holger Schramm | Carsten Wünsch [Hrsg.]

Medienrezeption

2., aktualisierte und erweiterte Auflage 2025,
770 S., geb., 99,– €
ISBN 978-3-8487-7384-8
E-Book 978-3-7489-1359-7
(NomosHandbuch | NomosHandbook)

Das Handbuch bietet einen umfassenden Überblick über die Gegenstände und Theorien der Rezeptionsforschung. In den 37 Beiträgen wird ein systematischer Zugang zum State of the Art der jeweiligen Thematik aus einer kommunikationswissenschaftlichen Perspektive gelegt.

Die umfassend aktualisierte und erweiterte 2. Auflage befasst sich in vier Teilen: (1) mit grundlegenden Konzepten der Medienrezeption, (2) mit Fragen der Zuwendung und Selektion, (3) mit spezifischen Phänomenen und Erlebnisweisen und (4) mit den wichtigsten Kontexten der Medienrezeption. Moderne Medienwirkungstheorien kommen ohne die Betrachtung von Rezeptionsprozessen nicht mehr aus. Im Handbuch werden daher die Prozesse vor und während der Mediennutzung im Detail betrachtet. Zielgruppe des Handbuchs ist das breite Fachpublikum der Kommunikationswissenschaft und angrenzender Fächer wie Soziologie, Psychologie, Politikwissenschaft und Erziehungswissenschaft.

Digital verfügbar auf Inlibra.com

Bestellen Sie im Buchhandel oder versandkostenfrei unter nomos-shop.de

Alle Preise inkl. Mehrwertsteuer

Nomos