Luxemburger Juristische Studien –
Luxembourg Legal Studies

Hofmann | Biber | Yusifli [eds.]

AI in European Law

A Deep View

Nomos

Luxemburger Juristische Studien – Luxembourg Legal Studies

edited by the

Department of Law, Faculty of Law, Economics and Finance,
University of Luxembourg

Volume 28

Herwig C. H. Hofmann | Elif Biber | Zahra Yusifli [eds.]

AI in European Law

A Deep View

Nomos

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

1st Edition 2026

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

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

ISBN (Print) 978-3-7560-3906-7
ISBN (ePDF) 978-3-7489-6930-3

DOI: https://doi.org/10.5771/9783748969303

Online Version
Inlibra

Acknowledgments

This volume emerged from the research of the Digitalisation Law and Innovation (DILLAN) doctoral training unit, supported by the Luxembourg National Research Fund (PRIDE19/14268506). It took further shape at the workshop "Frontiers of AI Law and Governance," held at the University of Luxembourg on 17–18 October 2024 as DILLAN's main academic event. The editors and contributors are deeply indebted to the many colleagues whose generous engagement shaped the chapters that follow.

We extend our sincere thanks to the distinguished discussants, whose insights from the bench, the academy, and professional practice immeasurably enriched our work. From the Court of Justice of the European Union and the General Court, we thank Advocate General Tamara Ćapeta, Judge Suzanne Kingston, Judge Tamara Perišin, and Judge Maria Lourdes Arastey Sahún, whose perspectives from the highest European courts offered invaluable guidance on questions of fundamental rights, criminal liability, taxation, consumer protection, and labour law in the digital age. We are likewise grateful to Hanns Peter Nehl, Legal Secretary at the European Court of Justice, for his contribution to the discussion on data sharing in competition law, and to Paul Nemitz, Principal Adviser at the European Commission, for his commentary on digital fundamental rights in the European Union.

Our academic discussants brought careful readings and probing questions that substantially improved the chapters that follow: Professor Niovi Vavoula (University of Luxembourg), Assistant Professor Simona Demková (Leiden University), and Assistant Professor Silvia de Conca (Vrije Universiteit Amsterdam). From professional practice, we are indebted to João Gião, General Counsel at the European Stability Mechanism; Steve Jacoby, Managing Partner for Europe at Clifford Chance; and Mirko Zichichi, Applied Research Engineer at the IOTA Foundation, whose engagement with the chapters on financial services, fund governance, and distributed ledger technologies brought essential practitioner perspectives to bear on our analyses.

Finally, we thank the doctoral researchers and early-career colleagues who made this book possible. Any errors that remain are, of course, our own.

Table of Contents

Herwig C. H. Hofmann, Elif Biber, and Zahra Yusifli
AI in European Law: A Deep View 9

Arthur Bianco
Beyond the Scope: AI, Tax Administrations, and the Limits of the EU AI Act 15

Zahra Yusifli
Labour rights and the EU Artificial Intelligence Act: How to get away with high-risk AI 47

Isabella Lorenzoni
From B2B to B2G and G2G data sharing in competition law. When data become a competitive advantage and an enforcement tool 73

Kalliopi Terzidou
Artificial Intelligence in the Service of Self-represented Litigants Accessing the Law with the Aid of Large-Language-Models 101

Anna Moraiti
The Recognition of Legal Personhood for Autonomous AI Systems 135

Grazia Bruzzese
In AI we trust – The use of artificial intelligence tools for payment fraud detection 159

Théo Antunes
The Formation and Execution of Contractual Securities through Smart Contracts under Luxembourg Law 183

Elif Biber & Herwig C.H. Hofmann
A Human-Centric Approach Through Rights: The *Missing Dialogues* between Rights, Risks, and Remedies under the EU's Artificial Intelligence 209

Contributors 231

AI in European Law: A Deep View

Herwig C. H. Hofmann, Elif Biber, and Zahra Yusifli

The Purpose of the Book

This book invites to explore the legal challenges posed by the growing use of technology across various sectors of EU law and policy. It addresses the legal challenges posed by AI technologies by looking at the intersections of policy-specific questions and regulation with general AI regulatory questions in fields such as competition law enforcement, fraud detection and criminal liability, labour law and working conditions, autonomous decision-making systems, and financial market regulation. These chapters are set against a discussion of questions of digital rights, open data policies and challenges of access to justice and judicial review of data-related matters.

By bringing together scholars from diverse legal backgrounds, spanning public, private, and criminal law, the book offers a multidimensional understanding of how latest technology is reshaping legal systems. Each chapter analyses both specific regulatory challenges and their implications for the EU's wider digital agenda. The case studies and legal analyses are grounded in real-world examples, linking legal theory to practical concerns.

The primary aim of this volume is to assess whether the current EU legal framework is equipped to address the profound transformations AI brings. It does so by analysing how AI affects decision-making by both private and public actors, and how legal systems respond to issues such as algorithmic opacity, autonomy, and the continuous learning capacity of AI systems. These features challenge conventional legal categories and demand new approaches to accountability, transparency, and risk management.

A key theme of the book is the need for interdisciplinary approaches in AI governance. Legal expertise alone is insufficient to evaluate high-risk AI systems, which also require input from computer science, ethics, and sector-specific knowledge. This is reflected in the volume's structure, with each chapter combining doctrinal analysis with technical and contextual insights. For example, the regulation of algorithmic management systems in labour law, or the review of smart contracts in private law, are discussed not

only in terms of legal compliance but also practical feasibility and societal impact.

The book also addresses the cross-border dimensions of AI regulation, considering how EU law interacts with national and international legal frameworks. Topics such as social scoring, fraud detection, and AI in supply chains illustrate the cross-border nature of these challenges.

In sum, this volume argues that AI regulation is not just a technical exercise, but a normative project concerned with balancing innovation, democratic control, and the protection of fundamental rights. The EU's evolving regulatory approach represents a significant attempt to govern AI in line with these values. By offering a combination of theoretical reflection and policy-relevant analysis, this book contributes to the ongoing re-evaluation of how law can and should respond to the digital transformation of society.

Structure of the Volume and Key Themes

A recurring theme of this volume is the issue of accountability in the public sector use of AI: who can be held responsible, and through which legal safeguards, when automated systems produce harmful or unjust outcomes? Chapter 2 by Arthur Bianco illustrates this challenge through the example of AI-driven tax administration in the European Union. Tax authorities increasingly rely on AI tools for fraud detection, audit selection and taxpayer services, yet these systems may produce opaque, biased, or erroneous outputs with severe consequences for individuals. Bianco shows that these risks are not matched by an adequate EU-level regulatory response. Although the AI Act frames certain public-sector AI as 'high-risk,' it expressly excludes tax administration from this category, leaving such systems subject only to minimal transparency obligations. The safeguards central to the AI Act therefore do not apply, despite the significant coercive powers tax authorities wield over citizens. Bianco argues that this gap reflects political sensitivities surrounding fiscal sovereignty and a narrow, market-focused legislative design. His contribution demonstrates that accountability in AI-supported public administration cannot be achieved through technical regulation alone: safeguards for fundamental rights must be embedded even in domains that lie at the heart of state authority.

Where Bianco exposes accountability gaps in public-sector AI, in Chapter 3, Zahra Yusifli turns to the workplace to show how similar gaps arise

under the EU AI Act in employment relations. Although Annex III classifies workplace AI systems as high-risk, the Act's derogation clause in article 6(3) opens a path for employers to argue that their tools merely support, rather than shape, decision-making. Yusifli shows how these derogations, combined with ambiguous thresholds for 'material influence' and narrow conceptions of harm, risk allowing high-impact employment systems to escape the obligations designed to ensure transparency, human oversight and accountability. This is particularly problematic in contexts of algorithmic management, where power imbalances already discourage human intervention and facilitate automation bias. Even when transparency duties exist, workers often lack the procedural tools, bargaining power and information needed to challenge AI-supported decisions. Her analysis demonstrates that meaningful worker protection cannot rest solely on risk classification and supply-chain obligations. Instead, robust safeguards rooted in labour law, collective rights, and enforceable oversight are required to prevent AI from entrenching managerial prerogatives and eroding workplace autonomy.

In Chapter 4, Isabella Lorenzoni explores how data's dual role, as both a competitive advantage for firms and a crucial input for public enforcement, reshapes EU competition law. She examines business-to-business data sharing, where dominant firms' control over proprietary datasets may exclude rivals and distort markets. Traditional mechanisms such as the Essential Facilities Doctrine provide limited ex post remedies. In contrast, emerging EU regulations like the Digital Markets Act and the Data Act introduce ex ante data-sharing obligations to address market failures in data access. Crucially, the chapter extends this analysis to the public sector, showing how business-to-government and government-to-government data sharing can equip competition authorities with the datasets necessary to deploy AI-based enforcement tools. In an era of 'computational antitrust,' authorities need high-quality, machine-readable data to detect collusive practices, assess mergers, and monitor digital markets. Lorenzoni identifies legal gaps in the Data Act's provisions and argues for tailored sectoral frameworks that enable proactive data sharing while safeguarding privacy and commercial confidentiality. The chapter calls for integrated legal approaches that not only regulate dominant firms' data power but also empower regulators to uphold fair competition in AI-driven markets.

The volume continues with a contribution examining how AI is reshaping access to legal services. In Chapter 5, Kalliopi Terzidou critically examines AI-powered legal chatbots, asking whether large language models such as ChatGPT and Gemini can support self-represented litigants in

navigating legal processes, with a focus on landlord-tenant disputes under Greek law. Drawing on the EU's commitment to trustworthy AI, Terzidou develops a normative framework grounded in fairness, human autonomy, explicability, and prevention of harm. Her evaluation reveals that while LLM-based tools score well on accessibility and readability, they often provide inaccurate, outdated, or incomplete legal information and lack proper citation of sources. Most troublingly, these systems can generate misleading outputs without safeguards such as consistent disclaimers or verifiable references. Rather than bridging justice gaps, such tools risk entrenching new inequalities when users rely on outputs that appear authoritative but cannot be trusted. Terzidou distinguishes sharply between AI augmentation, where professionals are supported by AI tools, and AI substitution, where users rely on these tools in place of professional advice. Her findings warn against uncritical optimism and call for rigorous design standards, public oversight, and institutional safeguards.

Chapter 6, authored by Anna Moraiti, addresses whether autonomous AI systems should be treated as bearers of legal rights and obligations, particularly under criminal law. This inquiry stems from real-world challenges in attributing liability when complex AI systems cause harm in ways that defy clear human attribution. Drawing on analogies with corporate criminal liability, proponents have suggested that recognising 'electronic personhood' could fill accountability gaps. However, Moraiti's analysis reveals significant doctrinal pitfalls in this analogy. While corporations embody human collectives and serve as proxies for their members' intentions, AI systems lack the capacity for reflection, moral agency, or relational grounding that personhood presupposes in criminal law. Extending criminal liability to AI risks creating an 'empty imputation,' allowing responsible human actors to evade scrutiny. Instead, Moraiti advocates for a more restrained approach rooted in civil law, where limited legal personhood could serve pragmatic goals such as fair risk allocation without overstepping conceptual boundaries. The chapter urges caution against anthropomorphising AI or diluting fundamental principles of criminal justice, while recognising the need for adaptive legal mechanisms in an age of emergent machine agency.

Chapter 7 by Grazia Bruzzese explores AI's deployment in digital fraud detection within the financial sector. Digital payment fraud has proliferated alongside the digitisation of banking, prompting institutions to turn to machine learning tools capable of detecting anomalous transactions in real time. Yet the adoption of such tools is not merely a technical upgrade but a legally intricate process. Banks must navigate overlapping obligations under

the AI Act, PSD2, the GDPR, and sector-specific supervisory standards. This layered regulatory context generates compliance complexity, particularly concerning the use, quality, and sharing of personal and transactional data. While AI can help mitigate fraud risk, its effectiveness hinges on access to large, high-quality datasets, a condition often hampered by interbank competition and privacy constraints. Bruzzese identifies this data bottleneck as a critical obstacle and proposes a coordinated EU approach that might include a centralised fraud data platform, analogous to existing anti-money laundering databases. The chapter demonstrates that technical tools alone cannot resolve fraud's persistence; effective AI deployment requires legal frameworks that facilitate data cooperation, uphold fundamental rights, and promote systemic resilience.

Chapter 8 by Théo Antunes focuses on smart contracts, automated, self-executing digital agreements, and their integration into legal frameworks governing contractual securities. Through a detailed analysis under Luxembourg law, one of Europe's most fintech-forward jurisdictions, the chapter probes how the immutable, automatic nature of smart contracts intersects with the flexible, interpretive requirements of private law. While Luxembourg has adopted several 'Blockchain Laws' to support innovation, Antunes shows that the legal regime remains primarily focused on blockchain as an object of investment rather than as a means of forming and executing obligations. This creates significant legal uncertainty when applying smart contracts to securities, which often demand human discretion and responsiveness to unforeseen events. The chapter identifies key challenges, from coding accuracy and oracle reliability to questions of enforceability, consent, and contractual form, and proposes practical safeguards including hybrid smart contracts that blend automation with human oversight. These recommendations aim to preserve the normative foundations of private law while enabling legal innovation. Antunes' contribution underscores that as automation technologies shape legal and economic life, their interaction with existing doctrines requires not only regulatory updates but conceptual clarity about what law demands when human judgment is outsourced to machines.

The final contribution by Elif Biber and Herwig C. H. Hofmann addresses a structural tension at the heart of the EU's Artificial Intelligence Act: the *missing dialogues* between its stated commitment to fundamental rights and the technocratic, risk-based framework through which those rights are meant to be protected. The chapter shows that, despite repeated references to fundamental and human rights and the rule of law, the AI Act insuffi-

ciently translates these commitments into enforceable legal mechanisms. The authors develop this argument in three steps. Firstly, they situate the AI Act within the EU's broader digital acquis, revealing how it conflates fundamental rights protection with technical risk management and product-safety compliance. Second, they examine human oversight obligations and the Fundamental Rights Impact Assessment (FRIA), arguing that both risk becoming symbolic, checklist-based exercises ill-suited to capturing the real-world impacts of AI. Thirdly, they analyse the Act's review and conformity-assessment procedures, showing how reliance on self-assessment and private actors weakens accountability and access to justice. Finally, the chapter outlines concrete steps to align fundamental rights with their legal enforcement, highlighting improved access to justice as key to realising the EU's human-centric approach to AI governance. This contribution serves as both a capstone and a call to action for the volume.

Beyond the Scope: AI, Tax Administrations, and the Limits of the EU AI Act

Arthur Bianco

Abstract: The integration of artificial intelligence (AI) into public administration has accelerated rapidly in recent years, with tax administrations across the European Union increasingly adopting AI systems to enhance compliance, streamline operations, and improve services to taxpayers. Despite the growing reliance on AI in this context, the regulatory framework designed to govern such technologies—namely, the AI Act—offers limited protections for individuals affected by their use in tax systems. This article critically assesses the AI Act's scope, its risk-based classification approach, and the consequences of excluding tax-related AI systems from its most stringent safeguards. The analysis situates these regulatory omissions within the broader legal and institutional structures of the EU, questioning the adequacy of the Act in protecting taxpayers' rights and maintaining trust in automated decision-making.

1. The use of AI by tax administrations: an undeniable reality

Tax administrations in the EU, similarly to most European public administrations, have in the recent year adopted, at a particularly rapid pace, Artificial Intelligence systems to facilitate their work and improve on their capacities. In general, across all States, the vast majority have reported using these tools to some extent in their functioning.[1]

Reliance on AI systems can be, broadly, categorized in three major sections: the use for traditional compliance management, the improvement of taxpayer-facing services and reliance for policy-making purposes.

Compliance management focuses, in taxation, on the methods used by the tax administration to compel taxpayers in fulfilling their tax obligations, which beyond the payment of their tax debt also includes other duties such as declaratory obligations or communication duties.[2] A number of tax administrations have introduced systems related, on the one hand, to fraud detection, in order to determine which taxpayers are likely to be non-com-

1 OECD, *Global Survey on Digitalization in Tax Administration* (OECD Publishing 2021).

2 Jeffrey Owens, 'Taxpayers' Rights and Obligations' (1990) 18 Intertax 554.

pliant[3], as well as audit selection systems, which aim at establishing the best strategy to adopt in facilitating the recovery of due taxes.[4]

The role of the tax administration is not, in itself, limited to the straightforward recovery of due taxes. It also communicates with taxpayers and offers specific services targeted at facilitating their compliance. For example, the tax administration must remain available for any inquiries from the taxpayers through different channels including in-person appointments, telephone or on the internet. Tax administrations have harnessed the power of AI systems to provide support for taxpayers by implementing chatbots or artificial assistants[5], but also in improving communication strategies.[6]

Finally, AI systems have also been used to facilitate the work of policymakers, mostly by providing additional ways for tax administrations to assess policy projects. including a number of sandbox-based systems to work out issues with these experiments.[7]

Literature and operational reports have underlined the advantages of the use of AI systems in taxation, including the ability for the administrations to reallocate resources towards other tasks, recover additional taxes and respond to additional pressures from increased demands in regards to the fight against tax evasion and avoidance. Nevertheless, a number of difficulties and risks with the use of AI systems in tax administrations have been identified. On the one hand, the technologies themselves have raised concerns, mostly related to biases and discriminations, transparency

3 For example, Slovakia has introduced a VAT fraud detection system comparing declared information with company invoices. See IOTA, *Data-Driven Tax Administration* (IOTA 2016) 16.

4 For example, Ireland has been using regression analytics to compare taxpayer returns with that of his peers to determine if they should be investigated. OECD, *Advanced Analytics for Better Tax Administration: Putting Data to Work* (OECD Publishing 2016) 23.

5 Spain uses a AI-powered VAT assistant named Skatti with positive success rates for resolution. IOTA, 'Applying New Technologies and Digital Solutions in Tax Compliance' (Conference Report, IOTA Annual International Conference – Tax Compliance Technology Showroom, 2020) 47.

6 For example, the Czech tax authorities has introduced a tax portal which includes an automated messaging techniques related to the taxpayer's due date for the fulfilling of their tax obligations and tax liabilities. OECD, *Tax Administration 2022* (OECD Publishing 2022) 78.

7 Spain, for example, uses a policy visualization tool allowing for the presentation of relevant data and insight on newly introduced policies. CIAT, *Manual sobre Gestión de Riesgos de Incumplimiento para administraciones tributarias* (CIAT 2020) 106.

or the position of the human agent within the decision process.[8] On the other hand, the adoption of those tools within public administrations has also highlighted issues, such as the difficulty to ensure equality and fair treatment, enhance trust in the system and adequately allocate limited resources.[9] Additionally, the disparate approach to legislative actions by different States, including the lack of specific regulations for the use of AI systems (or technology in general) in taxation, have shown flaws in the protection of taxpayers in cases where AI tools may threaten their rights.

The prior shows both the positive aspects, but also the negative, of the reliance on AI tools in tax administrations and justifies a deeper analysis of the existing legal framework in relation to the protections available to taxpayers in regards to these uses. Therefore, it seems relevant to discuss the EU's action in regards to adoption of AI systems by both private and public actors. The AI Act is the Union's response to a number of concerns identified in relation to this use of AI tools, and thus one may wonder how the introduction of such a regulation may affect tax administrations.

2. The AI Act: Functioning and Justifications

An appreciation of the EU's response to the growing concerns linked to the use of AI both within public administrations and the European Union in general would not be complete without an assessment of its main foray into regulating the use of AI systems in the internal market: the AI Act.[10] Indeed, the following section will not only present this legislation but also critically discuss how it impacts – or not – the use of AI tools by tax administrations. The aim of this section is to consider what protections are accessible to taxpayers, but also outline both what remains outside the scope of the legislation and the potential missed opportunities for the EU legislator to extend protections to taxpayers.

8 Fernando Serrano Antón, 'Artificial Intelligence and Tax Administration: Strategy, Applications and Implications, with Special Reference to the Tax Inspection Procedure' (2021) 13(4) World Tax Journal 575.

9 María Amparo Grau Ruiz, 'Fiscal Transformations Due to AI and Robotization: Where Do Recent Changes in Tax Administrations, Procedures and Legal Systems Lead Us?' (2022) 19(4) Northwestern Journal of Technology and Intellectual Property 325.

10 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L/1689 (AI Act).

2.1. What is the AI Act and what are its purposes?

The AI Act is a regulation drafted by the European Commission which aims at regulating the use of Artificial Intelligence systems, or as defined by the regulation, within the European Union and its market. Indeed, the AI Act provides for a uniform legal framework for the development, the placing on the market and the putting into service of artificial intelligence systems in the Union.[11] The regulation was initially conceptualized and justified primarily by the need to regulate the use of these systems at European level rather than leaving it up to the Member States. The risk, potentially, would be that Member States may try to introduce their own legislation to regulate the use of AI, leading to discrepancies and asymmetries in the rules and protections established by those.

Before delving deeper into the actual functioning of the framework established by the Act, one should note that the European Legislator, in drafting the Act, included the protection of health, safety and fundamental rights as one of the main purposes of the legislation.[12] Similarly, the recitals of the Act also emphasize the fact that the regulation was drafted with a market orientation in mind.[13] Indeed, the European legislator states that the Act provides grounds for the Union fundamental freedoms enshrined in the Treaties, namely the protection of the freedom of movement of AI-based goods and services across the Union, to be ensured throughout these recent technological evolutions.

Therefore, this piece of legislation should be seen as a multifaceted attempt at regulating a complex area all the while ensuring that the rules and protections found therein are sufficiently comprehensive as well as future-proof to apply to a large array of technologies anticipating unforeseen potential uses characteristic of the development of a recently introduced new technological advancement.[14] Thus, in analysing the AI Act, one should nonetheless consider these potential difficulties faced by the European legislator when drafting the legislation, including in cases where the Act lacks finesse or details.

11 AI Act, recital 1.
12 AI Act, recital 1.
13 AI Act, recital 8.
14 Atte Ojanen, 'Technology Neutrality as a Way to Future-Proof Regulation: The Case of the Artificial Intelligence Act' (2025) 16 European Journal of Risk Regulation 1440.

2.2. How does the AI Act achieve those aims, and why does it choose to do so?

The most notable characteristic about the legal framework established by the Act is its risk-focused approach: The Act imposes more or less obligations on AI providers depending on the significance of the risk linked to the use of AI in certain circumstances. The Act establishes a scale of risk, that covers AI systems that pose an 'unacceptable risk' down to AI systems with 'minimal-to-no risk'. Each risk category includes rules that the European legislator considered to adequately pursue the aims of the directive (namely the protection of health, safety and fundamental rights) in relation to those risks[15].

This risk approach was adopted as a way to minimize unnecessary limitations to the introduction of AI systems on the European market and to focus the obligations implemented by the Act to use cases that needed them the most. According to the recitals, the European Legislator intended to focus stricter rules on AI systems that create the most significant risk.

The risk scale adopted by the Commission is separated in four categories of risks, which, in order of gravity, scale down from 'unacceptable risk' AI systems, following by 'high-risk' systems, and 'limited-risk' systems to finally end with 'minimal-to-no risk' systems. Each risk category is tied to different rules or requirements in the Act, which will be overviewed in the following. The first category focuses on AI systems presenting an 'unacceptable risk' and are therefore entirely banned from entering the EU market, according to article 5. This category includes cognitive behavioural manipulation of vulnerable people or to circumvent individuals' free wills, social scoring based on social behaviour or personal characteristics, biometric identification and categorization of persons and real-time identification systems, with some exceptions for law enforcement purposes.

According to article 6, high-risk AI systems are themselves separated into two categories. The first are AI systems that are safety components or the AI system being a product itself covered by a legislation listed in Annex I, or are safety components of products that are required to undergo a third-party assessment according to EU legislations listed in Annex I is comprised of a list of Union harmonization legislation (which covers medical equipment, toys and aviation…), one part based on the New Legislative Framework and the second listing other harmonization legislations. The second paragraph

15 AI Act, recital 7.

or article 6 also extends the category to systems that are used as a safety protocol to a product covered by the legislations listed in Annex II if they require third-party conformity assessment or is put into service of that product. The second category of high-risk AI systems are called 'standalone high-risk systems', which, according to article 6(2), are AI systems that fall in one of the classifications in Annex III. This annex includes eight areas in which are listed types of AI systems that should be considered high-risk, even if they do not fall within one of the legislations listed in Annex II. The eight areas covered by Annex III are: (1) Biometrics, (2) Critical infrastructure, (3) Education and vocational training, (4) Employment workers management and access to self-employment, (5) Access to and enjoyment of essential private services and essential public services and benefits, (6) Law enforcement, (7) Migration, asylum and border control management, (8) Administration of justice and democratic processes. Only the AI systems that fall within one of the specific categories within those areas shall be considered high-risk. One should note that article 7 gives the possibility for the Commission to expend the list of standalone high-risk AI systems.

It should be noted that most of the requirements found in the Act focus on high-risk systems only. Indeed, Chapter 2 and 3, articles 8 to 49, list requirements and obligations to which high-risk systems and their providers or users are subject to. These requirements cover a broad array of obligations that aim at ensuring that these systems do not negatively affect the health, safety or fundamental rights of individuals. These requirements, include, among others, the obligation to introduce a risk management system, the introduction of data governance rules, the writing of technical documentation, the respect of transparency obligations or the inclusion of human agents in the process. This chapter also deals with the notification procedure for the conformity assessment referred to in article 19, and provides for additional rules regarding that also apply to high-risk systems regarding applicable standards, conformity assessments, certifications and registration.

Article 50 imposes transparency obligations for certain AI systems that represent a limited risk, and are therefore neither banned under article 5, the unacceptable risk, or considered high-risk. This category covers AI systems that interact with natural persons, and aim at ensuring that they are developed in a way that ensures that these persons are aware that they

are dealing with an AI system.[16] Similarly, users of biometric categorization, except if used for law enforcement purposes, shall inform natural persons affected of the use of such systems.

Before its official adoption, the text of the AI Act went through several changes, the more relevant one being the addition of a chapter on General-Purpose AI ('GPAI') Models, Chapter 5. Following the recent technological developments, the EU legislator focused on amending the initial proposal of the Act to include safeguards in cases of use of GPAI systems. The Act currently provides for obligations in two categories, those that apply horizontally to all GPAI and those that only apply to GPAI systems that present a 'systemic risk'. Horizontal obligations are mostly summed up to transparency rules covering technical documentations and content on which the system was trained. As for systems that present a systemic risk, they will be subject to stricter rules, including the conducting of system evaluation to potentially respond to specific risks and the reporting to the Commission on certain incidents, among others.[17]

Lastly, AI systems that do not fit in any of the above categories are considered to present a minimal-to-no risk, and are not bound by any obligations. Nonetheless, the following of voluntary Codes of Conduct with similar guidelines that those included in the high-risk system requirements, by the AI providers or users, is encouraged.[18]

3. Effects of the AI Act on the use of AI systems in tax administrations

In the context of the use of AI systems in Tax Administrations, the relevance of the Act is easy to understand, and one can therefore wonder how the introduction of such a legislation in the European Union might impact those uses, especially in considering the issues introduced in the previous part. Considering the risk-based approach adopted by the European legislator in the Act, one should determine how tax fits within the tiers established in the legislation, which will then lead to the specific obligations that tax administration will potentially be subject to.

16 AI Act, recital 132.
17 AI Act, art 55.
18 AI Act, art 95.

3.1. Unacceptable risks systems

In regards to unacceptable risk systems, AI systems used by the tax authorities are very unlikely to fall within this category, which would otherwise completely ban them. No AI use that has been made public so far have fit in any of the specified categories that have been flagged as unacceptable. Nevertheless, should a use find itself in that classification, it would most likely be caught by domestic administrative law protecting citizens against authoritative administration (for example, should a system exploit vulnerabilities of specific populations or try to manipulate human behaviour to circumvent free will). However, depending on how the use of such systems evolves within tax authorities, AI systems may prove to be problematic in this regard.

Indeed, the use of AI tools for social scoring by public authorities based on social behaviour or personal characteristics is strictly banned by the Act. According to the Recital of the Act, social scoring corresponds to 'evaluat(ing) or classify(ing) the trustworthiness of natural persons based on their social behaviour in multiple contexts or known or predicted personal or personality characteristics.'[19]. Although the reference to 'social behaviour' and 'personality characteristics' seem to fall outside of what would be traditionally seen as relevant for tax authorities, one could imagine how the examination of a taxpayer's behaviour could be significant to the mission of the administration. For example, should a taxpayer's lifestyle broadly differ from their declared income, the tax administration could be inclined to pursued a more diligent investigation in their affairs to determine if the taxpayer is potentially not being compliant. One could thus easily imagine the development of an AI tool that considers such traits to determine the likeliness of non-compliance. The Commission has published guidelines on prohibited artificial intelligence practices under the AI Act.[20] In it, the Commission states specifically that: 'The 'social scoring' prohibition has a broad scope of application in both public and private contexts and is not limited to a specific sector or field.'. However, the main questions reside in the assessment of social behaviours or personality characteristics. According to the Commission guidelines, these are broad, and include, among others, income, address, level of debt and even the type of cars. Ad-

19 AI Act, recital 17.

20 European Commission, 'Commission Guidelines on Prohibited Artificial Intelligence Practices Established by Regulation (EU) 2024/1689' C(2025) 5052 final.

ditionally, these characteristics and behaviours have to lead to a detrimental or unfavourable treatment, which the Commission highlights does not have to be solely driven by the assessment of behaviours and characteristics. However, the AI output must 'must play a sufficiently important role in producing the social score'.[21] Additionally, the data used to determine the social score must be used either in a social context unrelated to the context for which the data was originally collected, or disproportionate to the social behaviour or its gravity. This is particularly interesting for the tax context, as tax administrations usually collect a large amount of data, but also receive data either from other public administrations within the State, or also from other tax administrations of other Member States. This means that, potentially, tax administrations may feed into an AI system, data on which taxpayers could be scored (for example, as to determine if they are likely or not to fraud or deviate from their obligations), without that data necessarily being obtained at first for that specific purpose. The Commission also highlights that unfavourable treatments may extend to cases 'of scoring practices where people are singled out for additional inspections in case of fraud suspicious'[22]. This means that AI systems specifically made to profile and flag individuals that may potentially fraud, or the best candidates for audits could be covered. Among the examples of prohibited systems provided by the guidelines of the Commission, one may find certain likely scenarios relevant for tax administration: 'National tax authorities use an AI predictive tool on all taxpayers' tax returns in a country to select tax returns for closer inspection. The AI tool uses relevant variables, such as yearly income, assets (real estate property, cars etc.), data on family members of beneficiaries, but also unrelated data, such as taxpayers' social habits or internet connections, to single out specific individuals for inspections.' It should be noted, however, that article 5 specifically prohibits the social scoring of natural persons. Taxation targets a diversified population, which includes both natural persons but also legal persons, that, according to article 5(1)(c), would not fall within the scope of the protection. This further complexify the question of the use by tax administration of scoring systems.

21 ibid [161].
22 ibid [165].

3.2. High Risks AI

Due to the importance given to those systems in regards to the number of safeguards and obligations found in the AI Act, it is evident that one should determine if AI systems used by the Tax Administrations fall within this category. As noted earlier, an AI tool is considered high-risk either because it is a safety component or a product subject to existing safety standards under EU law, or it falls in one of the class of systems listed in Annex III of the Act. Since AI tools employed by tax authorities would not be included in the first category of high-risk AI, one should determine if AI uses by tax authorities could fall in the latter classification.

Among the categories listed in Annex III, none of them specifically reflects the uses of AI tools currently used by tax administration within the EU. Indeed, while several categories refer to the use of AI systems by administrative bodies, none of them specifically include tax authorities. Paragraph 5 of Annex III states that certain AI employed in the sector of 'Access to and enjoyment of essential private services and essential public services and benefits', shall be considered high-risk. Furthermore, paragraph 5(a) classifies as high-risks systems tools that 'AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services'. However, this category of systems is relevant only if the tax authorities are also tasked with providing assessment for the distribution of such benefits and services. This would also fall outside the framework of research of this chapter. Beyond this category, no other uses described in paragraph 5 seem to fit the definition of a system potentially used by the tax administration.

Paragraph 6 of Annex III focuses on AI uses by law enforcement, and lists a number of AI uses that would lead to those systems being considered high-risk. If at first glance this category seems to fall outside of the type of systems used by tax authorities, this is further confirmed by the Recitals. Indeed, the legislator states explicitly that 'AI systems specifically intended to be used for administrative proceedings by tax and customs authorities as well as by financial intelligence units carrying out administrative tasks analysing information pursuant to Union anti-money laundering legislation should not be considered high-risk AI systems used by law enforcement

authorities for the purposes of prevention, detection, investigation and prosecution of criminal offences.'.[23]

This means that the AI tools employed or potentially employed by tax administrations in the EU are unlikely to ever be considered as high-risk by the AI Act, so long as the categories of Annex III remain the same. This also excludes tax administrations using such tools from being subject to a majority of the rules found in the Act.

This conclusion is particularly surprising in taking into account that the Act alternatively defines high-risk systems covered by the legislation as '*hav(ing) a significant harmful impact on the health, safety and fundamental rights of persons in the Union*'.[24] Considering the issues outlined in earlier part of this chapter, one would be remiss to exclude AI tools used by tax administrations from the category of AI systems potentially having a significant harmful impact on fundamental rights.[25] The discrepancy between this definition and the categorization of high-risk systems in Art. 6 of the Act has been outlined as a major concern. Indeed, it seems that the decision to limit high-risk tools to either product or safety components or the list in Annex III does not allow for the inclusion of all tools that could potentially pose significant issues and harm, and leaves out a large number of problematic uses. This is especially concerning since the vast majority of the safeguards and obligations found in the Act concern only high-risk AI systems. This in turns means that a number of systems, including those employed by the tax authorities, are not subject to the guarantees provided, in spite of the potential harm or risk that they present. One may note that Art. 7(1) provides for the possibility for the Commission to amend Annex III to add other systems to the list, further expanding the category of high-risk AI tools. Nevertheless, this amendment is limited by the fact that such system should already fall within one of the areas listen in points 1 to 8 of Annex III. It should be noted that before the adoption of the AI Act, article 7(1)(b), has been added, which state that an AI system could also be qualified as a High-Risk system if: '*the AI systems pose a risk of harm to health and safety, or an adverse impact on fundamental rights, and that risk is equivalent to, or greater than, the risk of harm or of adverse*

23 AI Act, recital 38.

24 AI Act, recital 27.

25 David Hadwick, 'Slipping Through the Cracks, the Carve-Outs for AI Tax Enforcement Systems in the EU AI Act' (2024) 9(3) European Papers 936.

impact posed by the high-risk AI systems already referred to in Annex III'.[26] This addition is particularly interesting, as limiting the amendment to areas already covered by Annex III would have severely limited the ability for the AI Act to be properly technology-proof, especially with the swiftness with which AI systems are being introduced in certain areas. Nonetheless, this possibility is limited by article 7(1)(a) which limits the application of those changes solely tot he areas already covered by the Annex. This, in turn, restricts these changes to predetermined uses of AI systems, which seems to exclude most technologies used in tax administrations, as evidenced earlier. Ebers also questions the lack of details as to the risk-assessment and classification carried out in the drafting of the Act in order to establish the list in Annex III, and states that the solution provided by article 7 is insufficient in ensuring the flexibility and adaptability needed in the context of a technology-driven legislation.[27]

It is possible that the choice of the Commission to limit the definition of high-risk systems to either systems that are a safety component or a product already covered by a European legislation, or to the systems comprised in one of the areas listed in Annex III comes from the fact that the European Council, during the drafting on the initial proposal, emphasized the need to provide a clear definition of high-risk AI.[28] Therefore, the Commission may have consciously moved away from a wide-applying definition that may create uncertainty in favour of a more limited definition. This may also be justified by the need to limit legal uncertainty for EU actors that would otherwise need to deal with an added layer of complexity if the definition of high-risk AI systems were to be changed.[29] However, the opposite can be argued as well, seeing as the lack of safeguards for taxpayers (and

26 AI Act, art 7(1)(b).

27 Matthias Ebers and others, 'The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)' (2021) 4(4) J 589, 593. Ebers here states that the amendment procedure for Annex III should allow for the addition of entirely new categories of systems, and not just specific uses inside pre-defined categories. Ebers also underlines the need to include ongoing research (and public consultation) in the amendment procedure to ensure that the risk assessment is based on significant and actual risks.

28 European Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' COM (2021) 206 final, 2 and 8.

29 This has been confirmed by the Commission in its Staff Working Document. European Commission, 'Commission Staff Working Document, Impact Assessment Accompanying the Proposal for AI Act' SWD (2021) 84 final, 50. However, the

individuals affected by AI uses not classified as high-risk) could lead to uncertainty as to their treatment and available remedies. While tax administration may be bound by domestic legislations in that regard, taxpayers operating on a cross-border basis would find themselves unable to predict how each Member State would handle infringement to their fundamental rights. Moreover, while the addition of new requirements could lead to legal uncertainty for private actors adopting AI tools, the same may not be said for public bodies. Indeed, even if a larger area of public administration were to be added to the list of standalone high-risk AI systems, and thus subject to more obligations, the national legislator would thus have to introduce such obligation at domestic level (as well as ensuring coordination with domestic protections), though the administrations themselves (including taxation) would not necessarily be as impacted by this additional uncertainty as private actors. Indeed, public actors are better equipped at dealing with potential legal uncertainty, especially when they themselves are at the source of administrative decisions.

In analysing the protection afforded by the Act, authors have underlined the fact that safeguards are dependent on the coverage of the list in Annex III, and should this list not be sufficiently comprehensive or up-to-date, the actual purpose of the Act would not be reached. This has been acknowledged as a difficult task, due to the evolving nature of the technology and the digital scene in the European Union.[30] Nevertheless, the importance of the adequacy between the list in Annex III and the systems potentially creating major harm to health, safety or fundamental rights seems paramount to the success of the Act.[31] Authors have criticized the list-based categorization of standalone high-risk AI systems by stating that

---

Commission here acknowledges the fact that the current approach may limit the Act's ability to cover all potentially harmful systems.

30 Nathalie Smuha and others, 'How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act' (LEADS Lab, University of Birmingham 2021) 13.

31 The Commission, in its Staff Working Document, has emphasized this by stating that the main objective of the legislation is to ensure that AI systems are 'safe and respect the existing law on fundamental rights'. Similarly, the Commission also pointed out that currently national and European authorities do not have sufficient power or resources to effectively ensure and monitor compliance with those requirements. This shows that protection of fundamental rights is paramount to the goals pursued by the Commission, and therefore the disparities in applications of requirements may limit the ability for the Act to ensure protection of fundamental rights. See European Commission (n 29).

the narrowly defined classification 'fails to capture the nuanced reality of the way AI systems are used – and have an impact on fundamental rights'.[32] This pushes potentially high-risk systems in the no risk category and limit the ability for the Act to effectively provide fundamental rights protection. Literature further indicates that while the prescriptive list-based approach might be intended to secure legal certainty, it might have the opposite effect in leaving major areas in the shadows.[33] Mahler also criticizes the risk assessment approach adopted by the legislation by underlining the fact that the Commission, in drafting the Act, claims to have used a quantitative risk-analysis to determine which systems should be qualified as high-risk when it most likely also took into account political and societal aspects that are hardly quantifiable. For that reason, it should also be noted that the risk-assessment carried out is necessarily thus context dependent and may vary according to the actor performing the assessment.[34] This may further provide explanation as to why AI systems employed by tax administrations may not have been considered as presenting sufficient risk to be included in that category. As tax remains a largely sovereign area of action, it is possible that the Commission underestimated the potential risk these used may cause. This argument could nevertheless be nuanced by the fact that VAT, for example, is a harmonized area and therefore gives the Commission sufficient knowledge of the relationship between taxpayers and fundamental rights protection for it to be able to determine exactly which risks may arise from the introduction of AI tools in this field.[35]

Stuurman and Lachaud explain the choice of the Commission in delimiting the categories of high-risk AI by the fact that the Act focuses not on the actual functioning of the system, but on the purpose intended by the use of the system.[36]

32 Smuha and others (n 30) 13.

33 Smuha and others (n 30). Smuha also argues in favour of a more extended risk-based approach rather than the sectorial one adopted by the Act, and underlines the solution used by the GDPR in allowing for the criteria used to cover most realities that would create risks.

34 Tobias Mahler, 'Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal' in *Nordic Yearbook of Law and Informatics 2020–2021* (2021) 267.

35 This is further comforted by the fact that the ECJ has rules in several instances on the issue of fundamental rights in regards to taxpayers, allowing the Commission to assess more easily how the adoption of AI tools may affect them.

36 Kees Stuurman and Eric Lachaud, 'Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence' (2022) 44 Computer Law and Security Review 105657, 3.

Additionally, the Commission has stated that it has elected to adopt a risk tier system for the Act following an in-depth evaluation of societal and economic impact of each policy option considered for the development of the legislation. This included, as underlined by the Commission itself, a focus on the potential impact of the policy on fundamental rights.[37] This is further proof of the significance placed on the risks of detriment to fundamental rights that AI systems could cause. If so, in approaching the question of high-risk AI systems, that is to say which systems should require further scrutiny from the EU legislator, this focus on fundamental rights should be considered. It is thus even more surprising that the Commission then opted to limit the category of standalone high-risk AI systems to the areas found in Annex III.

Indeed, seeing as the Commission opted to leave the enforcement of the AI Act requirements to national authorities, it could potentially have been more relevant for the purpose pursued by the Commission to also give the opportunity for such authorities to determine which AI systems should be considered as high-risk, also depending on the specific AI uses in each Member States. In that case, it could have been feasible for the Act to impose distinction between low and high-risk AI on the basis of their potential impact on health, safety or fundamental rights rather than adopting a sectorial approach.

Another issue in relation to the AI Act's treatment of tax administration systems is the lack of definition of the individuals affected by the use of those tools.[38] Indeed, it seems that the Act may associate this status as being users of the AI tool, although there are many situations where this is not the case. This is particularly relevant to the area of taxation, as most affected by the use of AI tools, that is to say the taxpayers, are distinct from the user, the administration. This could also inform the reasons behind the exclusion of tax systems from the high-risk tier, as the potential negative effects to fundamental rights, health and safe do not affect the user per se, usually the tax administrations, but a third party, the taxpayers, who is subject to the decisions made using the AI system. This may have impacted the risk-assessment carried out by the Commission in determining what system should be considered high-risk. It could be preferential for the Act to also consider the subjects for an AI system use, as it seems to have played

37 AI Act Proposal (n 28), 9.

38 Anna Kiseleva, 'Comments on the EU Proposal for the Artificial Intelligence Act' (SSRN, 2021) 1.

a role in determining which systems should be high risk. For example, law enforcement systems covered are mostly comprised of tools that have natural persons as their subject. This is particularly interesting in regards to taxation as the subjects of most AI uses are a broad and diverse category. AI systems in tax administrations could apply to both natural and legal persons, indiscriminately, which of course changes greatly the impact of such tools on fundamental rights. In the same vein, the diversity, even within natural individuals, could lead to complexity as to the Commission's ability to properly assess the fundamental rights effect of AI systems.

Interestingly enough, the inclusion of law enforcement AI tools in the high-risk category due to the possibility for 'over-policing' in case feedback looks exacerbates already existing discriminatory practices. One example is the SyRI case[39], where an AI system was used by the Dutch tax administration to detect welfare frauds, and led to indirect discrimination due to the criteria that were chosen for their risk-assessment.[40] The justification used for law enforcement tools could also very easily apply to tax administrations, especially in the context of tax fraud detection or audit selection. Although the Recitals of the Act explicitly excludes tax AI tools from the law enforcement category, it does not provide any justification for such exclusion. In taking into account the definition provided by the Act of high-risk AI systems (posing significant risk of harm, including to fundamental rights), this exclusion does not, once again, seem justified by such concern. The question of the reasoning behind the decision of the Commission thus remains.

Interestingly enough, the Commission justifies the exclusion of AI systems used by tax administrations from the high-risk category by saying that these are already sufficiently covered by existing legislation and that there is no evidence of significant potential harm. The Commission, however, does not reject the possibility for these areas to be reintroduced in the category by future amendments to the Annex.[41] These claims are particularly surprising, as these do not align with the current conclusions established in the context of this research.

Firstly, it would be hardly right to say that current legislation covers the area of taxation sufficiently to ensure that taxpayers are effectively protected

39 Adamantia Rachovitsa and Nora Johann, 'The Human Rights Implications of the Use of AI in the Digital Welfare State: Lessons Learned from the Dutch SyRI Case' (2022) 22 Human Rights Law Review 1.

40 Smuha (n 30) 33.

41 European Commission (n 29) 51.

in cases of AI uses. Indeed, beyond the requirements set by the AI Act (which has not yet been adopted, and may not even cover taxation), both primary and secondary EU legislation seems particularly lacking, due to the EU's lack of competence in direct taxation (which further justifies the need for this research). The Commission may refer to VAT, which is indeed harmonized, but does not necessarily include specific protections for taxpayers. In case the Commission here refers to domestic legislation, which, depending on the Member State, does indeed provide protection for taxpayers, but should not be the basis for the exclusion of taxation AI systems from the class of high-risk AI systems, as discussed earlier in the research: disparate and different level of protections in the EU may lead to severe limitations as to the freedom of movements of taxpayers.

Secondly, the Commission claims that no evidence of potential risks has been identified in the area of AI us in taxation. This assertion is difficult to contradict as the Commission has not made its risk-assessment practices public, and therefore it is difficult to establish which harms would constitute sufficient concern to be classified as 'high-risk'. Nevertheless, the relationship between high-risk systems and potential harm to fundamental rights could inform this assessment. As established earlier in this research, fundamental rights are currently and will be impacted by the use of such systems, and it is therefore doubtful that these uses show no evidence of potential harm.

Finally, the Commission asserts that should additional risks be unearthed in the area of taxation, Annex III could be amended for such systems to be included in the high-risk category. Once again, this claim seems dubious. As discussed earlier, the amendment procedure only empowers the Commission to add specific AI uses in the Annex should these uses fall within one of the eight predetermined categories, none of which seem to encompass taxation.

The treatment of AI systems used by tax administrations by the AI Act also raises several issues in relation to the use of such systems specifically in public administrations. Fink discusses the specificities of the risk presented by such systems when used by public authorities, and argues that these deserved accrued scrutiny from the European legislator when discussing the Proposal. Specifically, the use of AI tools in the decision-making process may require additional safeguards, which, for the time being, seem not to be the case for taxation. Indeed, in the absence of a high-risk system, the only transparency requirement is that under article 52 triggering if an AI

system is used directly with a natural person. This, as discussed earlier, does not extend to decision-making involving such tools. Fink identifies two distinct issues that have not been properly addressed by the Act: transparency and automation bias.[42] One could imagine the AI Act inserting such specific obligation specifically in all cases of uses of AI system by public bodies, separate from the existing requirements. This would also be in line with the philosophy adopted by the Commission in limiting obligations only to situations that pose sufficient risk. Furthermore, Fink also examines the issue of the potential lack of available remedies for individuals affected by AI-driven administrative decisions, which, has not entirely been addressed by the Act either. Indeed, the Proposal lacks a specific individual procedure in those cases[43], which would therefore be left up to national administrative rules and courts, further exacerbating the aforementioned discrepancies in protections available to taxpayers.

While the Commission expressly refers to the protection of fundamental rights against potential infringements by AI systems as one of the mains goals pursued by the Regulation, concerns over the effectiveness of the rules laid out by the Commissions in reaching this goal have been raised by several authors. The Act lacks effective protection due to the market-specific approach adopted by the Commission. Indeed, the focus granted to the preservation of innovation within the EU market may have discouraged the Commission from taking a more radical approach to the protection of fundamental rights. Moreover, the Regulation very explicitly balance these market interest with the protection of fundamental rights, which severely limits the Regulation's ability to ensure the latter. Authors have indeed criticized the fact that the interests of individuals affected by AI usage are given the same weight as that of other actors, included AI developers and providers. This results in a number of issues leading to a weakened fundamental rights protection.[44]

3.3. Low-Risk AI systems

Beyond the first two categories outlined before, AI systems used by tax authorities would therefore be considered low-risk according to the frame-

42 Melanie Fink, 'The EU Artificial Intelligence Act and Access to Justice' (EU Law Live, 2021) 2.
43 ibid 3.
44 Smuha (n 30) 16.

work set by the Act. In that case, the number of guarantees or safeguards applicable to them is highly limited.

As far as this research goes, there has been no use of General-Purpose AI use made public in any tax administration within, or outside of the European Union. This could be due to the very recent development of such technologies, or from the highly-specialized needs of tax administrations which lend themselves poorly to the use of GPAI. Nevertheless, this means that the rules introduced to better safeguard natural persons against harmful uses of GPAI do not concern this research.

Art. 50 of the Act does provide for additional transparency obligations for certain AI systems, depending on their purpose. While paragraphs 2, 2a and 3 concern respectively biometric categorization, emotion recognition and image, audio or video content manipulation, paragraph 1 of Art. 50 focuses on AI tools developed to interact with natural persons. This is relevant for tax administrations as AI-driven chatbots and digital assistants have been made available by several tax authorities. Art. 50(1) thus imposes the obligation for those providers to ensure that the natural persons are aware of the fact that they are interacting with an AI system. This may be paramount to tax authorities as the communications means between taxpayers and authorities have been rapidly evolving towards an AI-centric support to taxpayer rather than a human-to-human system.[45] One could however wonder if such transparency obligations could be extended to other cases beyond those specific cases: for example, in the context of the tax administration, could it be also preferable to impose a transparency obligation when a natural person is affected by a decision taken on the basis of an AI system? Depending on the risk targeted by those transparency requirements, it could also be beneficial to consider those other situations. Furthermore, authors have raised the issue of the content of the information to be provided to natural person affected by an AI system, as this has not been specified in the Act, and runs the risk of varying among providers.[46] This could mean than different providers would provide different information to individuals, depending on the transparency assessment of each system.

45 It should be noted that this category of AI tools lacks specific risk definition. The Commission in the draft seems to imply that as soon as an AI system interacts directly with a natural person, it presents an inherent risk that should be minimized (through additional transparency requirement), but there is no clear description of the actual potential harm these could pose. See Mahler (n 34) 249.

46 Stuurman and Lachaud (n 36) 5.

The Act only suggests that minimal-to-no risk AI system providers and users follow a Code of Conduct (and potentially holding themselves to the same standards as those imposed on high-risk systems) though this seems of limited impact. Indeed, while the costs of complying with the requirements found in the Act seem particularly high[47], which will most likely drive non-high-risk AI providers away from following these codes of conducts, these rules could add another layer of complexity that AI providers are unlikely to choose to follow by themselves. This is interesting as the Commission has underlined, in its justification for the Proposal, that voluntary self-regulation is unlikely to sufficiently ensure protection of fundamental rights and safety within the Union. While the Act provides regulation in regards to high-risk systems, the Commission must have known that simply suggesting the following of codes of conduct would not be sufficient in the context of low-risk systems. Therefore, even according to the logic of the Commission, the Act does not extend any regulation to low-risk systems.

Apart from the obligations outlined in the previous paragraph, AI tools employed by the tax administrations seem to pose, according to the tiered approach of the AI Act, no to minimal risk and therefore warrant little to no safeguard.[48]

Beyond the issues related to the determination of high-risk AI systems and exclusion of potentially harmful systems, the difference between the requirements imposed on high-risk systems and those (or lack thereof) for low-risk systems is particularly jarring.[49] Indeed, the Commission only suggest the following of a Code of Conduct for providers and users of low-risk AI tools, while high-risk ones are subject to a number of obligations. One may wonder if the Commission could have alternatively chosen to not simply reject any requirements for low-risk systems, but more so to impose

47 See Dimitar Lilkov, 'Regulating Artificial Intelligence in the EU: A Risky Game' (2021) 20(2) European View 169. According to research, the costs linked to the compliance with these requirements could reach the tens of thousands of euros, which could further increase the more staff is needed to fulfil these obligations.

48 Mahler does regret the fact that the Act puts the lowest level of regulation on the widest category of AI tools. Mahler (n 34) 250.

49 It should be noted however that literature does praise the Commission for foregoing a completely binary approach to risk and does not simply differentiate between high and low risks but also adds categories of unacceptable risks and lower-risk AI still requiring transparency obligations. For example, see Smuha and others (n 30) 2. This categorization still remains very limited, especially in regards to legal obligations, the majority of which only apply to high-risk AI systems.

a more limited number of requirements to systems that present less of a risk. While this would have required to adopt a different assessment of what constituted a high-risk system in the Act, it could also have avoided the situation where AI systems that are not considered to be entirely high-risk by the Commission would still be subject to requirements, albeit limited. Similarly, this would also have necessitated to elect which obligation should apply only to the highest risk AI systems, and which would also apply to lower risk systems. While this could have potentially added a layer of complexity to the functioning of the Act (which the Commission tried to curb), it could have solved a number of issues identified here regarding the treatment of AI systems used by tax administrations in the Act.

3.4. Why are AI systems in tax administrations are excluded from protections?

The prior sections have shown that despite focusing the use of AI in the EU, in both the private and the public sphere, the AI act has a limited impact on AI systems used in tax administrations. Thus, one may not rely on this legislation to ensure taxpayer protection. Nevertheless, these limitations do raise a question: For what reasons has the European Legislator considered that the use of AI by tax authorities should not be considered a High-Risk use of AI system? Could these reasons provide insight as to the state of taxpayer protections at EU level in general?

As mentioned earlier, the Commission has justified the exclusion of the AI systems used by tax administrations from the scope of the High-Risk AI category because it was 'considered sufficiently covered by existing legislation and there is no sufficient evidence for harms caused by AI'.[50] This justification is unfortunately not being evidenced by the Commission in its Staff Working paper, which, considering the prior section of this research, weakens its position. Interestingly, the prior could be contradicted by the European Parliament's initial identifications of high-risk sectors, which the Commission decided not to follow entirely. Indeed, the Parliament had considered the use of AI in tax administrations to raise a number of concerns, which would have justified its inclusion in Annex III of the Act.[51]

50 European Commission (n 29) 5.
51 European Commission (n 29) 51.

Nonetheless, one could also explain this exclusion by examining the specificities of taxation in allowing the European legislator to regulate it. Indeed, tax remains today a special case due to the importance of the collection of revenue for the State as a sovereign prerogative. This leads Member States to be somewhat reluctant to surrender the power to legislate in the area of taxation to the European Union. The EU has already legislated in this area, and even harmonized the Value Added Tax framework among Member States; yet tax remains a controversial topic at EU level.[52] Therefore, extending the obligations of AI providers to tax administrations may have been more politically complicated for the Commission and the European legislator to justify.

Additionally, the AI Act was adopted on the basis of article 16 of the TFEU, regarding the adoption of legislation aiming at the protection of personal data, but more importantly on the basis of article 114 TFEU, which allows for the European legislator, in accordance with the ordinary legislative procedure, to adopt legislation to approximation of laws to ensure the functioning of the internal market. The reference to the ordinary legislative procedure is interesting, as it involves both the Parliament and the Council, thus no unanimity in the Council is required (unlike the special procedure). This could have alleviated some of the tensions regarding the introduction of tax rules in the Act, as all Member States would not have to agree on the extension of the protections for tax AI. However, article 114(2) states that '*Paragraph 1 shall not apply to fiscal provisions*'. While this seems to exclude the possibility for the act to apply to AI systems in tax administration, this solution does not seem satisfactory. If the use of AI in tax is not considered High-Risk, they are still in scope of the Act in general: tax administration may employ AI techniques falling under the category of unacceptable systems, or General-Purpose AI models, and these administrations are still in scope of the article 50 transparency requirements. Therefore, either the European legislator made a mistake in relying on article 114 as a legal basis for the Act if tax administrations intended to still be in scope of the Act, or the protections found in the Act do not constitute a fiscal provision, as understood in article 114.

52 Taxation is a sensitive topic for EU Member States, as tax remains an expression of a State's sovereignty, although the EU has legislated in both indirect and direct taxation. Member States remain reluctant to give up on their sovereignty in those areas. See, for example, Tjaša Zgaga, 'The Fiscal Sovereignty of the European Union after the COVID-19 Pandemic and the War in Ukraine' (2023) 45(4) Journal of European Integration 703.

This question raises concerns, as the Court of Justice has, previously, stated that article 114(2) should be interpreted as prohibiting the use of the article as a legal basis *'not only for all areas of taxation, without drawing any distinction between the types of duties or taxes concerned, but also all aspects of taxation, whether material rules or procedural rules'*.[53] The Court also extends this prohibition to any arrangements of the collection of tax, as they cannot be dissociated from the system of taxation of which they form part.[54] Therefore, the use of AI systems by tax administrations, as described in the first section of this research, would most likely be considered as forming part of the taxation system, and regulating them could not be done on the basis of article 114. Additionally, reliance on article 114 TFEU as a legal basis could also exclude the use of article 7(1)(b) of the Act in order to modify the list in Annex III and add AI systems used by tax administrations in the category of High-Risk AI systems, further limiting the future prospects of application of the Act to protect taxpayers.

One should conclude that the all AI systems used by tax administrations should be excluded from the scope of the Act. The fact that this has not been clearly stated by the legislator in the text of the Act or in the recitals could indicate that this was an oversight, and may not have been a conscious overall exclusion. It is nonetheless a precarious situation, as it even excludes taxpayers from the minimum protections introduced in the Act, including the relevant prohibition of the use of unacceptable risk AI tools. As discussed in an earlier section, certain systems that rely on the social scoring of taxpayer to assess risk of fraud could fall under the definition of article 5 and Prohibited AI Practices.[55]

4. Could and should the AI Act apply for tax administrations?

4.1. Future possibilities for applicability

A first possibility would be to introduce a reform to the recently adopted AI Act in order to ensure that AI systems used by tax administrations are included within the category of high-risk AI, and would then benefit from the obligations and protections that have been provided in cases of use of other AI systems. As discussed in chapter 5, Annex III of the AI Act, which lists

53 Case C-674/20 *Airbnb Ireland UC v Région de Bruxelles-Capitale* EU:C:2022:303 [27].
54 ibid.
55 AI Act, art 5.

the sectors in which AI usage would be considered high-risk, and currently does not include tax administrations, may have changes brought by the Commission in order to expand what can be considered high-risk. These changes are, however, impossible, as article 7(1)(a) limits those changes to the areas already covered by Annex III. As discussed, the use of AI by tax administrations does not fit in any of the predetermined categories set by the European legislator in the Annex, and an expansion to cover these systems is not an appreciable solution. Moreover, had the possibility been opened to the Commission, such changes to Annex III would not be likely, as, in the context of tax administrations, the Commission has stated that it does not find the adoption of AI systems to be creating particular risks of harm to fundamental rights. This could change, should the use of AI in taxation prove to be harmful, although cases such as the Dutch childcare allowance scandal may have evidence this fact already.

One could imagine an entire reform of the AI Act to better accommodate protections for AI in tax administrations. This is particularly unlikely, for two distinct reasons. Firstly, the political opportunity of such an amendment might be difficult to imagine, in taking into account both the recency of the adoption of the Act, but also in considering the four years that passed between the initial proposal and the adoption, which is proof of the difficulty the Member States found in reaching a consensus, especially considering that the use of AI in tax was consciously exempted from the definition of high-risks systems. However, this possibility is also limited by the legal basis on which the Act relies on: indeed, the Act was adopted on the basis of both article 16 TFEU, which focuses on the protection of personal data, and article 114, which allows for the European legislator to adopt legislation, but not in the area of taxation or fiscal policy. While article 114(1) allows for the action of the legislator according to the ordinary legislative procedure, it also states in its second paragraph that *'Paragraph 1 shall not apply to fiscal provisions, to those relating to the free movement of persons nor to those relating to the rights and interests of employed persons.'*. Therefore, reliance on article 114 for any tax-related legislation is impossible. Article 114 has however been relied upon to draft several legislations that are relevant to this research, including the AI Act, mentioned in a prior chapter. This is of particular relevance in the context of this research, as underlined in chapter 5, due to the fact that the main AI regulation drafted by the EU legislator is based on an article that specifically excludes fiscal provisions from that legislation. Beyond the fact that this does show the lack of willingness from the European legislator to introduce AI rules at EU level, it is also possible

that relying on article 114 was the only path forward for the adoption of such rules. This could raise a number of questions relating to the feasibility, going forward, of the adoption by the European legislator of AI-centred rules that also include tax administrations and taxpayers. Thus, it seems unlikely that the legislator may rely on another legal basis to introduce an amendment to the Act.

Another possibility open to the legislator is the adoption of a technology-centric text, similar to the AI Act, but would cover the use of AI systems in tax administrations. This is all the more unlikely as the AI Act was drafted as a general legislation that would be the European Union's overall and unique response to the issues raised by the use of AI systems. This means that the Union may see little interest in writing additional legislation which, on top of proving difficult to pass as a consensus across Member States, could lead to overregulation and complications in ensuring that all legal frameworks function together.

4.2. Assessment of the usefulness of the protections for tax purposes

Following the previous conclusion, one may wonder if it would be preferable for the AI Act to extend applicable rules and safeguards also to AI systems used in tax administrations. Indeed, this reflects on the question of the efficiency of the Act in properly regulating the introduction of AI systems on the European market all the while guaranteeing adequate protections against potential harms posed. Before arguing in favour of an amendment of the Act to include tax administration use of AI systems, it may therefore be preferable to first determine how effective would the protections provided in the Act be in those cases. In doing so, one should compare the obligations and safeguards offered by the Act with the issues identified in relation to the adoption of AI systems by tax authorities. As outlined in the previous part, most of the relevant standards set by the Act apply solely to high-risk systems, and these are thus the rules that should be analysed here.

The second chapter of the Act focuses on requirements for High-Risk AI systems, articles 8 to 15, providing broad requirements which all systems should comply with. The aim of the requirements seems not to set up strict rules that would inherently dictate how high-risk AI systems should be developed, rather provide for general obligations focusing on allowing developers and providers to ensure themselves that their systems present

minimal risks. One of the better examples of this is article 9, which requires a high-risk AI system provider to establish a risk management system, thus all foreseeable risks should be identified, their probability of arising evaluated and potential countermeasures should be adopted in the view of eliminating or minimizing those risks. Although the phrasing of the article remains particularly vague, as the article remains silent on the nature of the risk-assessment test or how it should be carried out, probably in order to leave enough flexibility for AI providers to tailor their risk-assessment mechanisms to the specificities of their systems, it does ensure that the AI providers is tasked with the responsibility to carry out this assessment. In the same vein, article 13 imposes transparency obligations on high-risks system providers in order to secure proper information for users. While this article is slightly more precise as to the requirement imposed on the provider, including the nature of the information to be made available, it still gives providers the liberty to determine how this transparency obligation should be applied to their systems all the while requiring that specific steps be taken to ascertain that the use of the system is in line with the goals set by the Act.

For the purpose of providing adequate protection to taxpayers in cases of use of AI systems in tax administrations, the inclusion of those systems in the category of high-risk AI tools would have been preferable, as these requirements correspond to major issues outlined in literature as well as the limitations established by the Act itself (mainly related to risks linked to fundamental rights). Specifically, the adaptability of the articles aforementioned fits relatively well in the context of the discrepancies of uses and mechanisms used in tax administrations, which seems to align with the goals of the EU legislator. Furthermore, the obligations found in this chapter of the Act, while broad, may inform a more prudent approach from AI users and providers in tax administrations, decreasing potential risks linked to their uses. It is difficult to determine how much a soft law approach would impact the use of AI systems by tax administrations, however it could be considered a first step towards a uniformed approach.

Interestingly, the Commission in developing the Act emphasized the need to avoid unilateral national actions by Member States to regulate AI systems themselves, as it runs the risk of leading to market fragmentation and probably would effectively limit the movement of AI tools on the European Market. Moreover, the existence of distinct national rules could

also lead to increased compliance costs by AI system providers or users.[56] This reflects well on the issues identified in this research regarding the discrepancies that currently exist in the treatment of AI systems used by tax administrations and the need for additional safeguards for taxpayers. Indeed, by focusing on finding solutions at European level rather than leaving Member States to their own devices, one may hope that a somewhat uniform taxpayer protection would limit either compliance costs or fragmentations. Certain authors have criticized the choices made by the Commission in introducing this proposal by outlining the risks that such a legislation could still stifle innovation by leading AI providers to move outside of the European Union to jurisdiction that do not apply these rules. While this could be mitigated by nudging other countries to adopt such rules, certain authors have indicated that an alternative to a wide-ranging legislation could be to focus on specific areas of the introduction of AI systems, especially areas that may create specific issues.

Nonetheless, the obligations found in this chapter of the Act have been heavily criticized for their lack of effectiveness in relation to the goals pursued. Indeed, it seems that the broadness of the requirements also leads to a limited impact of those rules on the issues the Act is trying to fix. One of the better examples of that fact is article 14, which establishes the need for High-risk AI to function with the proper human oversight. Fully automated decision-making based on AI systems is one of the major issues which have been discussed in literature in the last few years, although the Act does very little to ensure that the human oversight it tries to implement is going to be necessarily relevant to the high-risk system it should supervise. the lack of specific definition of oversight, what constitute sufficient control or even who should be entrusted with the task, severely limits the effectiveness of the rule. Similarly, article 10 seems to try to establish a framework of data governance in the functioning of high-risk tools, although leaves major gaps as to how this data governance relates to the functioning of the AI system and what exactly is understood by the EU legislator as 'good data governance'.

These prior conclusions paint a mixed picture of the potential performance of the AI Act as an adequate tool in facilitating the protection of taxpayers in the context of AI uses by tax administrations. The EU legislator did address several issues that impact taxpayers and have to be corrected, much seems to have been left out of the Act that significantly limits its

56 Lilkov (n 47) 169.

effectiveness. Does that mean that the exclusion of tax administration AI tools from the high-risk tier is a non-issue? Although the impact of the Act would be limited, the message sent by the European legislator that those systems present minimal-to-no risks to those affected by them is concerning. It is more than possible that these systems will, and already have had, potentially harmful consequences, although the Act does not consider them as risky.

Furthermore, if the Commission has acknowledged the fact that self-regulation is most likely insufficient to ensure that the eventual risks posed by AI systems be suppressed or limited, the current obligations imposed on AI providers and users still leave a lot up to these actors. Indeed, every rule found Chapter 2 mostly provides general instructions but entrusts providers and users with filling themselves the gaps in the requirements set by the Act. For example, if article 9 requires that a risks assessment test be carried out for high-risk systems, the nature, content and consequence of such test is left entirely up to the providers. The same can be said of the necessity to introduce human oversight in the AI process.

The exclusion of tax AI systems from the framework of obligations for high-risk systems could also present another problem. Indeed, one should consider the relationship between the introduction of a binding legal framework that commits to the protection of fundamental rights in the area of AI regulation (as well as considers the rule of law and democracy implications) with the EU goal of trustworthiness established by the HLEG.[57] Trustworthy AI can only be attained by establishing responsibilities in cases of harms caused by the systems, but also by maintaining a legal framework that effectively enforces the protection of fundamental rights and ensuring accountability and transparency.[58] In the absence of such prerequisite, trust in these systems, especially for those affected by them, is unlikely to subsist. While such ideals seem to take form in the requirements set for high-risk AI systems, the same cannot be said for systems that fall outside the category. Therefore, the lack of protection provided by the Act to taxpayer could also endanger the trust placed in the tax administration's use of AI tools, especially when those systems are used in processes leading to potentially adverse decisions, and to an extent, the tax administration itself. As discussed in the earlier chapter of this research, trust is paramount to ensure taxpayer compliance, and it is likely that that adoption of AI systems

57 Smuha (n 30) 5.
58 Smuha (n 30) 6.

in tax administration without prior introduction of safeguards may weaken taxpayer confidence in the tax system. Therefore, the current definition of high-risk AI could lead to adverse effect on taxpayer trust.

Similarly, if the Commission considers the use of AI tools in tax administration to not be high-risk, this does not necessarily mean that Member States will think of it the same way. Since the Act effectively does not regulate these uses, Member States could be inclined to unilaterally provide safeguards and protections by imposing their own requirements on AI systems used by their tax systems. This is interesting, as some may argue that Member States are better placed to determine how to provide adequate protections for taxpayers in the specific context of their own tax systems. Nonetheless, the potential disparity the introduction of unilateral legislations by each Member States could lead to negative effects on the abilities of taxpayers to seamlessly move within the internal market, as identified in the earlier parts of this research.

Specifically, the lack of risk-assessment requirements for AI systems not considered high-risk is regretful, which would have given providers the opportunity to review their systems and identified risks and voluntarily followed a code of conduct or certain requirements set for high-risk AI tools.[59] This risk-assessment requirement creates doubt as to when exactly are risks considered when establishing the high-risk category. Indeed, all high-risk systems are required to carry out this risk assessment which should also have been initially done by the legislator when establishing this category. This identification of risks at several level may put into question this initial step in classifying systems according to their risks.

Another regrettable consequence to the exclusion of tax AI systems from the high-risk category is the fact that according to the Proposal, standalone high-risk systems would be listed in an EU-wide database, providing increased transparency as to the use of AI tools in Europe. This could have been particularly beneficial for taxpayers in knowing the types of systems are used by the authorities. This might however also explain why tax has been sidelined from the classification; it is possible that this inventory of systems could restrict the tax administration's mission if specific knowledge of the tools being used by the tax were to be disseminated to taxpayers. Indeed, the information provided in the database could endanger the necessary secrecy needed by tax administration to ensure that no taxpayer may

59 Smuha (n 30) 13.

'game' the system by collecting confidential information on the workings of the tax administration.

Beyond those requirements, taxpayers could have greatly benefited from the obligations set in article 10 regarding data management, specifically in allowing taxpayers to communicate with the administration to ensure that the data on which the authorities base themselves (and their AI processes) are up-to-date and accurate. Such obligation would incite tax administrations to keep a flow of communication with data subjects.

One of the main criticisms from literature focuses on the reliance by the AI Act on internal compliance by AI providers and users, with particularly limited cases of external and independent assessment of certain requirements. Entrusting AI providers with complying with those obligations with little supervision could effectively limit the effectiveness of the Act.[60] This is particularly exacerbated by the fact that the requirements in the Act are formulated in a broad way that give providers a large margin of discretion.[61] Accordingly, even if tax administrations were to be subject to the requirements established in the Act, it is, for the time being, difficult to determine how effective those obligations would be to ensure the safeguards necessary to avoid harms to fundamental rights.[62] In the same vein, the fulfilment of the rules in the Act actually hinges on the Art. 9 risk assessment procedure, and whether it has been carried out properly by the provider. Indeed, providers, in order to fully respond to those very broad requirements, have to actively know the weaknesses and risks of their systems to provide adequate response, even within the confines of those rules.[63] This is further exacerbated by the fact that the risk assessment procedure aims at determining what potential harms an AI system may pose to health, safety or fundamental rights, the latter which is potentially very difficult for every provider to assess properly. Indeed, it is up to the provider to decide which risk is to be deemed 'acceptable', which gives the provider

60 See, for example, Mauritz Kop, 'EU Artificial Intelligence Act: The European Approach to AI' (2021) 2 Transatlantic Antitrust and IPR Developments 1, 8.

61 Ebers and others (n 27) 593.

62 For example, Ebers discusses the human oversight requirement in art 14, and highlights how impractical the rule seems, especially in regards to the ability for a human to have a good enough understanding of the system to provide sufficient oversight, leading either to an effective ban of certain AI tools if supporting staff is not trained, or a circumvention of those rules, rendering the obligation in the Act useless. Ebers and others (n 27) 596.

63 Smuha (n 30) 34.

much liberty in deciding how to treat their systems.[64] This concern is also echoed by authors in analysing the other obligations set in the Act. For example, he data governance rules in article 10 do not define exactly what should be considered an appropriate and accurate data set, leaving it to the consideration of the provider. Similar doubts were raised in relation to the article 14 human oversight rules, where the lack of specificities could lead to discrepancies among providers on their application of the rules.

5. Conclusion

The AI Act marks a significant step in the European Union's effort to regulate the development and deployment of artificial intelligence systems, yet its current configuration reveals a critical shortcoming: the failure to account for the specific risks posed by the use of AI in tax administrations. By categorizing such systems as low-risk or excluding them altogether from the high-risk designation, the legislation overlooks potential harms to taxpayers' fundamental rights and fails to offer meaningful safeguards against opaque or discriminatory practices. This exclusion appears rooted in political sensitivities surrounding fiscal sovereignty and the limitations of the EU's legislative competences. It should be noted that both the legal basis on which the Act is constructed as well as the opportunities for expansion to the Act's scope of protection severely limit the opportunities for AI systems used by tax administrations to be included and meaningfully treated by the obligations under the legislation. Thus, the consequences are practical and far-reaching. Without harmonized protections, the burden of safeguarding rights falls unevenly across Member States, reinforcing disparities and undermining legal certainty. In light of these findings, the article calls for a reconsideration of the regulatory approach to AI in public administration, particularly in areas where the stakes for individual rights and democratic accountability are high.

64 Smuha (n 30) 34.

Labour rights and the EU Artificial Intelligence Act: How to get away with high-risk AI

Zahra Yusifli

Abstract: The European Union's Artificial Intelligence Act (AI Act) establishes a regulatory framework that primarily approaches worker protection through the lens of employer obligations in the AI supply chain. This chapter critically examines this approach by analysing employers' legal responsibilities when deploying high-risk AI systems for workforce management, particularly their role as providers and deployers of these systems.

A central focus is the examination of article 6(3) derogations and their implication for classifying high-risk AI systems in workplace contexts. Drawing on interdisciplinary research on workplaces utilising AI automation tools, the chapter identifies potential regulatory gaps in the AI Act's effectiveness. An examination of article 6's legislative history reveals that the final text of article 6(3) includes safeguards that mitigate the large-scale classification of high-risk AI, though several challenges remain.

The analysis shows the fundamental limitations in the AI Act's approach to workplace AI regulation. Although the AI Act establishes employer obligations, its heavy reliance on preconditions and product safety-oriented framework fails to adequately address the power asymmetries inherent in employment relationships. The chapter concludes by advocating for supplementary sector-specific policies to better regulate AI-driven workplace management and protect workers' interests.

1. Introduction: AI systems for workplace management

The wide scholarly discussion on emerging new technologies has focused on regulatory interventions in employment relations and the resilience of labour law institutions.[1] The use of artificial intelligence (AI) systems is inextricably linked to expanding employer prerogatives. Regional and domestic authorities have enacted and proposed regulations to address cer-

1 See, *inter alia*, Jeremias Adams-Prassl, 'What If Your Boss Was an Algorithm? The Rise of Artificial Intelligence at Work' (2019) 41 Comparative Labor Law and Policy Journal 30; Cynthia Estlund, *Automation Anxiety: Why and How to Save Work* (OUP 2021); Valerio De Stefano and Antonio Aloisi, *Your Boss Is an Algorithm: Artificial Intelligence, Platform Work and Labour* (Bloomsbury Publishing 2022) <www.bloomsbury.com/ca/your-boss-is-an-algorithm-9781509953172/> accessed 1 January 2025.

tain aspects of AI deployment.[2] The European Union Artificial Intelligence Act (AI Act) is one of the most comprehensive regulations on AI matters to date.[3] With several provisions aimed at protecting workers, the AI Act introduces a novel dimension of responsibility for employers utilising AI systems for work management.

The employer's responsibilities under the AI Act ultimately depend on two factors: (1) the role the employer plays in the supply chain of the AI product and (2) the type of AI the employer deploys. Since most work-related AI is classified as high-risk, this chapter examines and evaluates the legal obligations that the AI Act places on employers concerning the use of high-risk AI systems (HRAIS) to manage their workforce. The central research question is whether the derogation of the HRAIS classification in article 6 can compromise the effectiveness of the AI Act. This chapter answers this question by first identifying the parties in work relations within the AI supply chain created by the AI Act. Second, it analyses the responsibilities assigned to employers regarding HRAIS and distinguishes between their roles as deployers and providers. Third, it scrutinises the derogation in article 6. Finally, this chapter concludes with a synthesis of findings and implications of this regulatory framework for protecting workers.

2. The EU AI Act: Examining labour law matters under the omnibus law

The AI Act, which entered into force in August 2024, is scheduled for phased implementation over the period leading up to 2030. Despite the limited attention given to employment-related matters since the introduction of the first draft in 2021 and throughout the legislative process,[4] the

2 See, eg, OECD.AI, 'National AI Policies and Strategies' (*OECD.AI: Policy Observatory*, July 2022) <https://oecd.ai/en/dashboards> accessed 1 January 2025.

3 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L/1689 (AI Act).

4 For example, in 2021, the European Parliament established the Specialised Committee on Artificial Intelligence in a Digital Age (AIDA), tasked with evaluating the prospective impacts of AI across various sectors, including the labour market. See Special Committee on Artificial Intelligence in a Digital Age, 'AIDA Working Paper on "AI and the Labour Market" Following the AIDA/EMPL Public Hearing on 25 May 2021' (2021) <www.europarl.europa.eu/cmsdata/237745/Working%20Paper%20on%20AI%20and%20the%20Labour%20Market.pdf> accessed 1 January 2025.

AI Act governs certain aspects of AI tools used in workplace management. Since the omnibus law stems from product safety rather than labour law legal tradition, the AI Act regulates AI systems through the supply chain. This means the AI Act oversees the life cycle of AI products, including their development, market introduction and deployment. Although the legislation references parties to the work relation, such as workers, employers and workers' representatives, it does not explicitly rely on these terms. Instead, the AI Act delineates responsibilities for various actors within the AI supply chain, including deployers, providers, authorised representatives, importers, distributors, operators, downstream providers and product manufacturers.[5]

While these stakeholders have been integral to the EU's regulatory tradition since the inception of the common market, the AI Act introduces multiple actors that may not be immediately recognisable in labour law scholarship. Labour scholars have focused on institutes that regulate work relations and, in the age of advanced automation, assess the impact of AI systems on labour protections. Considering the focus of the AI Act towards product responsibility, scholars have voiced their concern that the Act may have limited implications for employment relations and, in particular, for workers' labour rights.[6] This product-centric approach prioritizes technical standards and conformity assessments over, for example, the contextual power dynamics and collective bargaining frameworks that traditionally safeguard workers' interests in the workplace. Others contend that existing national and European data protection regulations might prove more valuable for protecting workers' rights.[7] These scholars point to the GDPR's established mechanisms and its provisions on automated decision-making, which provide workers with rights to explanation and human intervention that may be more applicable than the AI Act's broader risk-based framework. To explore general rights and obligations established by the AI Act

5 See AI Act art 3 for the introduction and definitions of the terminology.

6 For the limitations of the AI Act in the workplace, see Aída Ponce Del Castillo (ed), *Artificial Intelligence, Labour and Society* (ETUI 2024) <https://www.etui.org/sites/default/files/2024-03/Artificial%20intelligence,%20labour%20and%20society_2024.pdf > accessed 1 January 2025; Aislinn Kelly-Lyth, 'Dispatch No. 39 – The AI Act and Algorithmic Management' [2021] Comparative Labor Law and Policy Journal 9 <https://cllpj.law.illinois.edu/content/dispatches/2021/Dispatch-No.-39.pdf> accessed 1 January 2025.

7 See Zahra Yusifli and Luca Ratti, 'Labour Law and Automated Systems in the EU' in Massimo Durante and Ugo Pagallo (eds), *Handbook on Law and Digital Technologies* (De Gruyter 2025).

in the employment context, this chapter proceeds by examining prohibited HRAIS and their implications for labour rights.

2.1 Prohibited practices under the AI Act

The AI Act establishes obligations based on the risks posed by AI systems.[8] One of the categories concerns *general prohibitions* of AI practices deemed to pose unacceptable risks which must be avoided at work. Article 5 considers several categories of AI systems unfit for use due to their potential infringement upon fundamental rights, including systems allowing social credit scoring, emotion recognition, exploitative practices, biometric categorisation and predictive policing.

The prohibition regime restricts AI systems within the EU internal market. The AI Act identifies four primary areas of prohibition: subliminal techniques, exploitations of vulnerabilities, social scoring and real-time remote biometric identification.[9] Some parts of this provision are directly relevant to employment. Article 5(1)(f) explicitly forbids 'the placing on the market, the putting into service for this specific purpose, or the use of AI systems to infer emotions of a natural person in the areas of workplace'. It represents a significant stride in protecting individuals' privacy in professional environments. The second part, however, significantly diminishes this provision's overall importance due to a major limitation. It states that AI systems 'intended to be put in place or into the market for medical or safety reasons' in the workplace are acceptable. This part leaves a big loophole in the prohibition as employers may use a broad set of AI systems

8 The EU AI Act was presented as a regulatory setting based on a risk-based four-tier pyramid: prohibited, high-risk, limited and minimal. See information disseminated at the European Commission website on AI Act <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 1 January 2025. However, some argue that this is not representative of the categories created by the content of the regulation. For example, Almada and Petit argue that the AI Act has unacceptable risks, high risks and minimal risks. See Marco Almada and Nicholas Petit, 'The EU AI Act: Between the Rock of Product Safety and the Hard Place of Fundamental Rights' (2025) 62 Common Market Law Review 1. Most workplace tools are prohibited and classified as high-risk AI, with few exceptions that do not fall under the high-risk category by derogation of article 6(3) discussed in this chapter.

9 Rostam J Neuwirth, 'Prohibited Artificial Intelligence Practices in the Proposed EU Artificial Intelligence Act (AIA)' (2023) 48 Computer Law and Security Review 105798 <https://linkinghub.elsevier.com/retrieve/pii/S0267364923000092> accessed 1 January 2025.

that would collect data related to the emotional state of workers for medical purposes.

The assessment of data that is used can be intimately related to the gathering of personal data, including biometric data, and its collection in the workplace. These data are broadly regulated by the General Data Protection Regulation (GDPR), which allows employers to collect and process workers' data.[10] Biometric data are held to a higher standard when employers can use biometric consent upon workers' explicit consent or when such biometric data is required for workers' verification or safety purposes.[11] The key principle is that biometric data should be used upon absolute necessity. The AI Act relies on the biometric data definition of the GDPR.[12] If the AI Act permits the collection of facial images, voice recognition data and movement data to protect workers, it must include safeguards that the current Act does not adequately provide.

While article 5(1)(f) allows for safety reasons as it can prevent workplace accidents, there is also another side of the coin. The global biometric technology market is rapidly expanding, and adoption rates of those technologies are significant across EU countries. While safety applications show promise in benefiting both workers and employers—such as preventing accidents and health issues—data indicates that increased automation and technology adoption have not necessarily resulted in a corresponding decrease in workplace injuries.[13] Meanwhile, there is a growing concern regarding workers' privacy. The conundrum with the AI Act lies in allowing employers to use AI systems for safety reasons, which opens a window to

10 See for discussion Frank Hendrickx, 'Privacy 4.0 at Work: Regulating Employment, Technology and Automation, Artificial Intelligence, and Labor Law' (2019) 41 Comparative Labor Law and Policy Journal 147 <https://heinonline.org/HOL/P?h=hein.journals/cllpj41andi=159> accessed 1 January 2025; Frank Hendrickx, 'Article 8 – Protection of Personal Data' in Filip Dorssemont and others (eds), *The Charter of Fundamental Rights of the European Union and the Employment Relation* (Hart Publishing 2019) <www.bloomsburycollections.com/book/the-charter-of-fundamental-rights-of-the-european-union-and-the-employment-relation> accessed 1 January 2025.

11 Dutch Data Protection Authority, 'Boete Voor Bedrijf Voor Verwerken Vingerafdrukken Werknemers (Fine for Processing Employees' Fingerprints)' (30 April 2020) <www.autoriteitpersoonsgegevens.nl/actueel/boete-voor-bedrijf-voor-verwerken-vingerafdrukken-werknemers> accessed 1 January 2025.

12 See recital 14 of the AI Act.

13 Mia Hoffmann and Mario Mariniello, 'Biometric Technologies at Work: A Proposed Use-Based Taxonomy' (Bruegel Policy Contribution 2021) Research Report 23/2021 <www.econstor.eu/handle/10419/270503> accessed 1 January 2025.

access workers' sensitive data. There is also a significant risk of collecting and using data beyond its original purpose.

With regards to biometric data, article 5(1)(g) of the AI Act refutes the use of biometric categorisation systems 'to deduce or infer […] race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation'.[14] AI systems' inference of trade union affiliation is prohibited, as such information is classified as sensitive personal data. However, this provision also allows for an exception for the labelling or filtering of lawfully acquired biometric datasets. Therefore, the prohibition set by article 5(1)(g) is also somewhat limited within the labour context.

There are other prohibited AI practices that might be broadly translated to be relevant in the workplace.[15] However, unacceptable AI practices in the regulation come with either caveats or significant legal uncertainties. While article 5 appears to establish general prohibitions for AI systems in employment contexts, both preventing employer usage and protecting workers, these prohibitions are drafted to extend beyond workplace environments. Their applicability in employment law, which is often governed by private law, remains subject to interpretation.

2.2 Employers as deployers and providers of HRAIS

The second category is 'high-risk' AI, which includes permitted systems subject to stringent regulations. The AI Act requires employers to use HRAIS to implement safeguards and oversight mechanisms. The obligations of employers with regard to HRAIS are based on their roles in the AI supply chain. In the context of AI regulation, employers can assume multiple roles. In fact, within the AI Act, employers are most likely to be deployers. More so, providers of HRAIS owe transparency obligations

14 Some of this sensitive data is protected in the EU under GDPR.

15 For example, article 5(1)(a) could be interpreted to prohibit employers from using AI to influence workers' behaviour subconsciously through subtle psychological manipulation. Article 5(1)(b) could ban employers from using AI systems that take advantage of workers' vulnerabilities to influence their behaviour. This might apply, for instance, to AI systems that target financially vulnerable employees with certain workplace policies or decisions. Furthermore, article 5(1)(c), in a workplace context, could prevent employers from using AI to create comprehensive behavioural profiles of workers, which could lead to unfair treatment.

pertaining to reporting the AI system's infrastructure to the deployers (employers). Along with the general responsibility of deployers of article 9 to oversee the AI system 'throughout the entire lifecycle of a high-risk AI system',[16] article 26 lists steps the employers should take within the enterprise to ensure AI compliance. Under article 26, the employer must assign competent individuals to oversee AI systems in the workplace; ensure that input data for the AI system is relevant and representative of its intended purpose; monitor the AI system's operation; inform providers of risks or incidents and suspend use if necessary; keep AI system logs for at least six months or longer; conduct data protection impact assessments; and cooperate with relevant authorities in implementing actions related to the HRAIS.

Article 26 stands out as it explicitly addresses labour matters by imposing obligations to inform workers about HRAIS. Specifically, article 26(7) states that 'deployers who are employers shall inform workers' representatives and the affected workers that they will be subject to the use of the high-risk AI system.' A similar obligation can be inferred from article 26(11), which requires informing 'the natural persons that they are subject to the use of the high-risk AI system.' The requirement aligns with broader information dissemination obligations included throughout the AI Act, particularly article 4's provisions on AI literacy. Article 4 exemplifies this approach by mandating that providers and deployers of AI systems take measures 'to their best extent' to ensure that actors who handle or are in the AI system's ecosystem have all the necessary information. Under article 4 alone, the workers are owed the comprehensive and general responsibility to be AI literate as they are the actors 'on whom the AI systems are to be used'. The addition of these provisions marks a considerable improvement, especially compared with the 2021 draft of the AI Act that did not include article 26(7)'s HRAIS notification requirements.[17]

While the AI Act establishes a certain degree of transparency in the workplace, the above-discussed provisions have noteworthy limitations. Firstly, their effectiveness ultimately depends heavily on labour laws and collective labour presence, which vary significantly across the EU Member States, industries and individual enterprises. Oftentimes, workers lack the

16 AI Act, art 9(2).

17 European Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – General approach 2021' COM (2021) 206 final.

mechanisms to be meaningful participants in implementing AI systems in their workplace environment. Instead, they are presented with the deployment of AI systems as a *fait accompli* under the EU framework with few options to contest automated decisions. Secondly, the AI Act does not foresee any enforcement of measures if the literacy principles are not upheld. Article 99, in particular, does not mention penalties for lack of literacy obligations or any more grounds for fines. While it does mention fines for the breach of transparency obligations mentioned in article 4(g), the obligations in article 99 are only relevant for providers and deployers, who are employers in the AI supply chain.

Finally, the obligations outlined in this part are also extended to non-EU-based employers who intend to use AI on their EU employees.[18] In its attempt to create a level playing field, the cross-border application of the AI Act extends primarily to multinational enterprises.[19] As outlined in the AI Act, the EU should promote mutual recognition of conformity assessments from qualified bodies, regardless of their geographic location, as long as they adhere to regulatory standards in line with those of the World Trade Organization.[20]

If employers purchase HRAIS applications in business-to-business transactions, their responsibilities are somewhat limited compared to those of the providers. In certain circumstances, employers, particularly large tech enterprises, may occupy dual roles as both providers–who develop and place AI systems on the market–and deployers when developing and utilising AI systems internally.[21] Employers who are providers build their own AI systems for workforce management or as occupational tools.

Under the AI Act, providers bear varying obligations according to risk categories, including those falling under HRAIS. Their responsibilities are generally more significant than those of employers who are only deployers. At the heart of these obligations lies a comprehensive framework of accountability measures, which include risk management and assessment,

18 AI Act, art 2.

19 See recitals 21 and 22 of the AI Act. Recital 22 further elaborates that 'To prevent the circumvention of this Regulation and to ensure an effective protection of natural persons located in the Union, this Regulation should also apply to providers and deployers of AI systems that are established in a third country, to the extent the output produced by those systems is intended to be used in the Union.'

20 See recital 127 and art 39 on 'Conformity assessment bodies of third countries' of the AI Act.

21 AI Act, art 3(3).

data quality and protection, documentation and traceability, human oversight and transparency, and technical standards and security.

Requirements that fall upon providers follow through the AI system's lifecycle. At the development stage, providers must build AI systems using curated datasets for training, validation, and testing purposes. Quality assurance extends beyond development and proceeds into deployment stages. Providers must implement technical documentation practices that demonstrate compliance with EU standards, though the Commission offers some relief to small and medium enterprises through simplified documentation options under article 11. The systems must incorporate automatic logging capabilities, creating an audit trail that facilitates ongoing monitoring.

Transparency and human agency form another essential pillar of provider obligations. Article 13 mandates 'transparency by design', which requires providers to equip users with clear instructions about system capabilities and oversight mechanisms. This obligation is connected with articles 14 and 15, which require effective human-machine interfaces and consistent performance standards throughout the AI system's operational life.

These threads — development controls, documentation requirements, transparency measures, and human oversight provisions — are included in Section 3 of Chapter III. Here, the regulation crystallises providers' core responsibilities. They are completed with specific requirements for maintaining documentation, automated logs, and channels for reporting non-compliance.[22]

2.3 Workers and the HRAIS

The AI Act presents a web of procedural requirements to ensure AI product safety. In this HRAIS scheme, limited obligations are directly owed to workers who are the subjects of AI system. The territorial scope of the AI Act suggests that provisions mainly apply to AI systems managing workers within EU territories, though the AI Act's overall protections for workers remain limited. A notable gap in the legislation is the absence of explicit conferral of rights to workers and workers' representatives. Instead, workers' protections can mainly be understood by examining employers' obliga-

22 AI Act, arts 16–21.

tions within the supply chain. In effect, employers' obligations become the gateway to understanding if and how the AI Act protects workers.

The AI Act does not address the underlying dynamic between employers and workers. The privileges workers might enjoy in AI-powered workplaces are largely indirect, dependent on proper implementation by employers, and tied to the obligations related to the HRAIS. This article further discusses the derogations of article 6(3) of the classification of HRAIS and assesses whether it may further weaken these already narrow obligations towards workers.

3. Analyses of derogations under HRAIS

Annex III provides the list of HRAIS referred to in article 6(2) of the AI Act. Article 7 further grants the EU Commission the right to modify the list to keep pace with the AI developments if necessary. At first glance, a broad array of AI systems created to manage workers would fall under Annex III and classify as high-risk. Annex III(4) includes:

(a) AI systems intended to be used for the recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates;
(b) AI systems intended to be used to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics or to monitor and evaluate the performance and behaviour of persons in such relationships.

Article 6, however, also contains an additional nuance. Even for systems whose purpose ostensibly falls under the HRAIS category, article 6(3) delineates several derogations from the general high-risk classification rule. The purpose of AI systems that may not qualify as HRAIS includes the following:

1. Performing a specific, narrow procedural task;
2. To improve the results of a previously completed human activity;
3. To detect decision-making patterns or deviations from prior decision-making patterns;
4. To perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.

Even if the purpose of the AI system falls under the derogation, there are three more preconditions that the AI system must meet so as not to be classified as high-risk. The AI system must:

(1) not pose a significant risk of harm to the health, safety or fundamental rights of natural persons;
(2) not materially influencing the outcome of decision-making, and
(3) not profile natural persons.

Paragraph 3 of article 6 was absent from the original draft of the AI Act presented in 2021.[23] Initially, article 6 outlined HRAIS as safety components within bigger products and standalone products that require a third-party conformity assessment under EU harmonisation legislation. The amendments on derogations have also been absent from most of the legislative process.

On 17 June 2022, the Czech Presidency of the Council of the EU shared a working paper outlining the priorities of the AI Act.[24] One of the priorities included 'Classification rules for standalone high-risk AI systems', in which some EU Member States expressed concerns about overly broad high-risk AI classifications in Annex III and the need to maintain effective regulatory flexibility while taking into account the AI risks.[25] The working paper laid out several ways in which HRAIS would be classified.[26]

Interestingly, one of the options proposed to link the risk classification to whether the AI system leads to a fully automated decision. Since automated decisions pose greater risks, the proposal suggested, they should always be classified as high-risk. When an AI system only provides information to help humans make decisions rather than autonomously, the proposal is considered high-risk only if its output can cause serious harm: health risks, safety concerns, violations of fundamental rights, and impact on important decisions. The essential point is that these 'advisory' AI systems are subject

23 See European Commission (n 17).
24 'Working Party on Telecommunications and Information Society on Options Paper for the Policy Orientation Debate on the Artificial Intelligence Act, Prepared by the Incoming CZ Presidency in View of the Discussion in WP Telecom on 5 July 2022.' (Council of the European Union 2022) WK 8862/2022 INIT <https://artificialintelligenceact.eu/wp-content/uploads/2022/07/AIA-CZ-Options-Paper-17-June-2022.pdf>.
25 ibid 2.
26 ibid 3–4.

to less stringent classification unless their recommendations could result in serious consequences.[27]

This debate defined the drafting of article 6 and made the classification of HRAIS a subject of amendments during the Parliamentary readings. On 14 June 2023, the European Parliament adopted further amendments to propose a new paragraph 2 for article 6 suggesting that providers who believe that their systems do not pose risks shall submit a notification to the national authorities and, if relevant, to the AI Office, alleviating them from the obligations.[28]

Article 6 caught the attention of human rights groups, which criticised the direction of the amendment. Some argued this approach would effectively allow AI developers to self-determine whether the systems qualify as HRAIS, undermining the original European Commission's more objective classification system. This self-assessment would create legal uncertainty, potential market fragmentation across the EU Member States, enforcement challenges and unfair competitive advantages for those willing to bypass safety requirements.

The Draft version of paragraphs 3 and 4 on derogations reflected in the final version of the AI Act was included only on 26 January 2024.[29] The final text appears to be a compromise that moves beyond pure self-classification while preserving an arguably rather broad path for HRAIS exemption from classification requirements. The legislative development showed that the text was decided as a middle ground – not to put categorical classification and to leave room for interpretation.

As a result, the version of article 6(3) of the AI Act that entered into force presents several layers of legal consideration. This section further examines these three preconditions and argues why article 6(3) creates a loophole in the AI Act for high-risk systems.

---

27 See Option 2 ibid 3.

28 See 'Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9–0146/2021 – 2021/0106(COD))1 (Ordinary Legislative Procedure: First Reading)' (European Parliament 2023) P9_TA(2023)0236.

29 'Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts – Analysis of the Final Compromise Text with a View to Agreement' (2024) 2021/0106(COD) <https://artificialintelligenceact.eu/wp-content/uploads/2024/02/AIA-Trilogue-Coreper.pdf>.

3.1 Assessment of health, safety and fundamental rights

Labour rights are an integral part of the Charter of Fundamental Rights of the European Union. The opus of labour rights is broad, ranging from the freedom of assembly and of association, workers' right to information and consultation within the undertaking, and the right to collective bargaining and action to the freedom to choose an occupation and right to engage in work, non-discrimination, protection in the event of unjustified dismissal, and fair and just working conditions.[30]

The AI system should not qualify for derogation if it presents potential adverse implications for individuals' physical health and safety or the fundamental rights of natural persons. Article 27 of the AI Act outlines the necessary procedures for the fundamental rights impact assessment (FRIA). Evaluating the harms AI systems may cause to fundamental labour rights raises questions. Firstly, article 27(1) does not mandate the assessment of employers' fundamental rights. FRIA is mandated only for public entities, private entities providing public services, and credit and insurance evaluations. While it cannot be inferred from article 6(3) that it imposes mandatory FRIA either, it can be concluded that deployers who believe their AI system falls under the derogation are responsible for ensuring that their systems do not negatively impact fundamental rights.[31] For that purpose, the requirements outlined in article 27 can serve as benchmarks.

These requirements lead us to the second point: what the FRIA entails and whether it can provide a meaningful assessment. Specifically, article 27 mandates that FRIA should include:

(a) a description of the deployer's processes in which the high-risk AI system will be used in line with its intended purpose;
(b) a description of the period of time within which, and the frequency with which, each high-risk AI system is intended to be used;

30 For literature on labour rights and the EU fundamental rights, see Brian Bercusson (ed), *European Labour Law and the EU Charter of Fundamental Rights: Summary Version* (ETUI 2002); Filip Dorssemont, Klaus Lörcher, Stefan Clauwaert and Mélanie Schmitt (eds), *The Charter of Fundamental Rights of the European Union and the Employment Relation* (Hart Publishing 2019); Christina Hiessl (ed), *EU Labour Law: A Commentary* (Kluwer Law International B V 2025).

31 Article 95 encourages the AI Office and EU Member States' authorities to develop voluntary codes of conduct that encourage non-high-risk AI systems to adopt Chapter III requirements for high-risk systems in accordance with technical capabilities and industry standards.

(c) the categories of natural persons and groups likely to be affected by its use in the specific context;
(d) the specific risks of harm likely to have an impact on the categories of natural persons or groups of persons identified pursuant to point (c) of this paragraph, taking into account the information given by the provider pursuant to article 13;
(e) a description of the implementation of human oversight measures, according to the instructions for use;
(f) the measures to be taken in the case of the materialisation of those risks, including the arrangements for internal governance and complaint mechanisms.

The AI Act's application of product safety risk assessment methodologies to fundamental rights protection presents critical limitations. The requirements for the AI system's safety look like a box-ticking exercise of the documentation of possible harms.[32] The process is more procedural than fundamental rights-oriented. There is a significant discussion on the value of this exercise in assessing risks related to fundamental rights. As Isabel Kusche points out, fundamental rights are used as a 'blanket term' associated with all the harm related to HRAIS. The AI Act imposes the binary exercise, which determines the existence or absence of the harm. While there is little on proportionality, Kusche argues that the AI Act fails to provide clear methods for quantifying risks or comparing risk levels between different AI systems.[33]

Furthermore, Marco Almada and Nicolas Petit identify a fundamental tension in the AI Act's regulatory approach: the challenging integration of product safety regulation with fundamental rights protection.[34] The AI Act's reliance on product safety methodologies creates several critical issues. Firstly, fundamental rights impact often resists quantification, making traditional risk assessment methods inadequate. Secondly, probabilistic estimations, while suitable for mechanical failures, struggle to capture complex

32 Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, Legal Obligations and Key Elements for a Model Template' (2024) 54 Computer Law and Security Review 106020 <https://linkinghub.elsevier.com/retrieve/pii/S0267364924000864> accessed 1 January 2025.

33 Isabel Kusche, 'Possible Harms of Artificial Intelligence and the EU AI Act: Fundamental Rights and Risk' [2024] Journal of Risk Research 1 <www.tandfonline.com/doi/full/10.1080/13669877.2024.2350720> accessed 1 January 2025.

34 Almada and Petit (n 8) 20.

human-AI interactions. Thirdly, the discrete risk approach fails to address the cumulative nature of fundamental rights violations. The tension is further exacerbated by conflicting regulatory philosophies. Product safety regulation operates on a 'satisficing' principle, where meeting minimum requirements suffices. In contrast, fundamental rights protection demands optimisation that maximises protection to the fullest extent possible. This methodological inconsistency creates significant implementation challenges and reveals the AI Act's core dilemma. While aiming to protect both product safety and fundamental rights, its reliance on product safety frameworks may ultimately compromise its effectiveness in safeguarding fundamental rights. The unidimensional assessment approach proves particularly problematic when evaluating competing, incommensurable fundamental rights values.[35]

In employment matters, the issues with the FRIA also become evident. Studies make it clear that AI systems can make executive decisions traditionally reserved for managers, a phenomenon known as 'algorithmic management'.[36] These decisions can include determining workers' remuneration, adjusting working hours, and establishing conditions of employment which affect both standard and non-standard forms of employment.[37] As discussed above, they would all fall under HRAIS due to their function as per Annex III.

35 For further discussion on the AI Act's product safety risk assessment and fundamental rights protection, see Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22 Computer Law Review International 97; Margot E Kaminski, 'Regulating the Risks of AI' (2023) 103 Boston University Law Review 1347; Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template' (2024) 54 Computer Law and Security Review 106020; Mélanie Gornet and Winston Maxwell, 'The European approach to regulating AI through technical standards' (2024) 13 Internet Policy Review; Samuele Bertaina and others, 'Fundamental rights and artificial intelligence impact assessment: A new quantitative methodology in the upcoming era of AI Act' (2025) 56 Computer Law and Security Review 106101; Margot E Kaminski and Gianclaudio Malgieri, 'The Right to Explanation in the AI Act' (SSRN, April 2025) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5194301> accessed 31 August 2025.

36 Adams-Prassl (n 1); De Stefano and Aloisi (n 1).

37 Alex J Wood, 'Algorithmic Management: Consequences for Work Organisation and Working Conditions' (Joint Research Centre (Seville site) 2021) 2021–07 <https://ideas.repec.org/p/ipt/laedte/202107.html> accessed 1 January 2025.

The extent of possible harm caused by AI systems on labour rights is not always apparent. In employment, the FRIA, even when mandated, might not capture the risk of AI systems. Taking the example of working time, AI systems are created with the goal of optimising the work schedule. That optimisation might lead to neglecting working hours, breaks and downtime, all of which are essential for workers' well-being. Even if the maximum working hours are protected by domestic labour laws, it has been noted that AI intensified the amount of work presented to workers.[38] As a result, the rapid execution of work under algorithmic management questions the established labour standards on the hours of work and perception of time. This framework on fundamental rights indicates that even if an AI system presents a potential risk to fundamental rights, systems may fall outside the protective obligation of article 6.

Another example of an AI Act-mandated assessment that might be missed is occupational safety and health and physical and psychosocial health risks. Those risks might not be apparent at first, but they might be secondary to AI's impact on other working conditions. The nexus between the working conditions cannot be captured by the proposed assessment of article 27. While AI has the potential to make workplaces safer, especially in dangerous industries, there are significant concerns that its benefits might not be distributed equally. For example, AI systems work less well for people with darker skin tones or from minority groups, and job displacement from AI could disproportionately affect communities that already face disadvantages.[39] Similar observations can be made about the health and safety of mistakes made by AI systems and allocating work shifts, bonuses, and discrimination due to trade union activities or family responsibilities. Given all of that, article 27 may easily miss the purpose of the evaluations bestowed upon it.

38 Agnieszka Piasna, 'Algorithms of Time: How Algorithmic Management Changes the Temporalities of Work and Prospects for Working Time Reduction' (2024) 48(1) Cambridge Journal of Economics 115.

39 Elizabeth Fisher and others, 'Occupational Safety and Health Equity Impacts of Artificial Intelligence: A Scoping Review' (2023) 20 International Journal of Environmental Research and Public Health 6221 <www.ncbi.nlm.nih.gov/pmc/articles/PMC10340692/> accessed 1 January 2025.

3.2 Material influence on the outcome of an employment decision

The next precondition for derogation concerns the material influence of the AI system on the decision. In the legislative process, this precondition was included to provide flexibility and allow the industry to assess whether AI systems are high-risk. The reasoning is substantiated in recital 53 of the AI Act, which provides that the purpose of AI systems that do not 'materially influence the outcome of decision-making could include' performing a narrow procedural task, enhancing the results of a previously completed human activity, identifying decision-making patterns or deviations from previous patterns, and carrying out preparatory tasks related to assessments relevant to the use cases of HRAIS. Those are the categories of the AI systems also mentioned in article 6(3). The AI Act ties 'materiality' with the purpose of the AI systems and uses the significance of the AI system in the said decision as a legal threshold to determine whether the AI system might fall under the derogation or not.

This approach is a slight departure from the current data-protection EU legal acquis on automation. The significance of the decision is an essential factor under article 22 of the GDPR, which prohibits solely automated decision-making. However, according to the GDPR, the significance of the outcome depends not on how much the automated system determines the outcome of the decision (article 22 only concerns *fully* automated decisions) but on the type of legal decision that is made.[40] The Data Protection Working Party report gives examples of contract termination and dismissal of applications in e-recruitment processes as examples falling within the scope of this provision.[41] This means that to qualify for the prohibition

40 It must be noted that in the SCHUFA case (case C-634/21), the Court of Justice of the European Union broadens the scope of article 22 applicability to third parties. The Court provided that when parties heavily rely on decisions based on probability values such as scoring, these decisions fall within the purview of article 22 of the GDPR. This can potentially extend the responsibility to the third parties that provide automated decisions but do not necessarily make the final decision. While this case does not challenge the automated nature of the decision that SCHUFA made as a credit score agency, it illustrates that machine suggestions can play a considerable role in important final decisions. See Case C-634/21 SCHUFA Holding (Scoring) EU:C:2023:957.

41 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' 21 <https://ec.europa.eu/newsroom/article29/items/612053> accessed 1 January 2025.

under article 22 of the GDPR, the decision should be legally consequential or otherwise significant for workers.

Despite different approaches, the criteria for what defines 'significant' impact can remain ambiguous in both article 22 of the GDPR and article 6 of the AI Act in the employment context. The matter of what materially influences the decision is the one that legal scholars have been grappling with for a long time. In the labour context, this consideration is particularly relevant due to the power imbalance inherent in the employment relationship, as employers have the right to make decisions about the direction of work.[42] From the perspective of EU bureaucracy, tying the significance of an AI system to an ex-ante evaluation of its intended purpose can be beneficial. However, this approach may lead to legal complications, as benchmarks are open to interpretation. This vulnerability can be attributed to two main factors: (1) the fragmentation of AI systems and (2) the law's inability to reflect that fragmentation. These two points are considered below.

3.2.1 Fragmentation of decision-making by AI Systems

AI tools are increasingly complementing human labour across a wide range of sectors. Rather than simply replacing human labour, AI systems, in many instances, augment workers' capabilities by automating repetitive or data-intensive tasks.[43] Some AI systems can make executive decisions traditionally reserved for managers, a phenomenon known in academic literature as 'algorithmic management.'[44] These decisions can include determining workers' remuneration, adjusting working hours, and establishing conditions of employment which affect both standard and non-standard forms of employment.[45]

AI-enhanced management extends across industries such as commerce, education, healthcare, finance and retail services, with some sectors more

42 On the prerogatives of employers, see Gali Racabi, 'Abolish The Employer Prerogative, Unleash Work Law' (2021) 43 Berkeley Journal of Employment and Labor Law 79.

43 For an assessment of the impact of generative AI on jobs, see Paweł Gmyrek, Janine Berg and David Bescond, 'Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality' (International Labour Office 2023) 96 <https://doi.org/10.54394/FHEM8239> accessed 1 January 2025.

44 Adams-Prassl (n 1); De Stefano and Aloisi (n 1).

45 Wood (n 38).

susceptible to automation than others.[46] The coupling of hardware and software gadgets enables the totality of digital workplace management. Digital surveillance infrastructure includes wearable tracking devices (rings and bracelets),[47] Global Positioning System trackers,[48] video footage[49] and standard software suites like Microsoft 365[50] that now incorporate worker activity monitoring features. AI tools offer to collect vast amounts of workers' data, process it at unprecedented speeds, and produce more diverse and complex outputs.[51]

As fully automated and AI-driven workplaces do not yet exist, managerial input remains a constant factor.[52] AI systems in the workplace could be implemented as modular components rather than as a single comprehensive system, further complicating the assessment of their impact on decision-making. When employers or managers use AI-assisted tools, part of their decision-making is partially taken over by machines.

---

46 Annette Bernhardt, Lisa Kresge and Reem Suleiman, 'Data and Algorithms at Work: The Case for Worker Technology Rights' (*UC Berkeley Labor Center*, 3 November 2021) <https://laborcenter.berkeley.edu/data-algorithms-at-work/> accessed 1 January 2025; Wood (n 38); Alexandra Mateescu and Aiha Nguyen, 'Algorithmic Management in the Workplace' (Data and Society 2019).

47 Kateryna Maltseva, 'Wearables in the Workplace: The Brave New World of Employee Engagement' (2020) 63(4) Business Horizons 493; Philippa Collins and Stefania Marassi, 'Is That Lawful? Data Privacy and Fitness Trackers in the Workplace' (2021) 37(1) *International Journal of Comparative Labour Law and Industrial Relations* 65.

48 Kirstie Ball, 'Electronic Monitoring and Surveillance in the Workplace' (Publications Office of the European Union 2021) <https://publications.jrc.ec.europa.eu/repository/handle/JRC125716> accessed 31 August 2025.

49 Paul Glavin, Alex Bierman and Scott Schieman, 'Private Eyes, They See Your Every Move: Workplace Surveillance and Worker Well-Being' (2024) 11 *Social Currents* 327.

50 Elisa Giacosa, Gazi Mahabubul Alam, Francesca Culasso and Edoardo Crocco, 'Stress-inducing or performance-enhancing? Safety measure or cause of mistrust? The paradox of digital surveillance in the workplace' (2023) 8 *Journal of Innovation and Knowledge* 100357.

51 For a discussion on the increasing interest in AI applications in academic studies, see Emilia Filippi, Mariasole Bannò and Sandro Trento, 'Automation Technologies and Their Impact on Employment: A Review, Synthesis and Future Research Agenda' (2023) 191 Technological Forecasting and Social Change 122448 <www.sciencedirect.com/science/article/pii/S0040162523001336> accessed 1 January 2025; Araz Zirar, Syed Imran Ali and Nazrul Islam, 'Worker and Workplace Artificial Intelligence (AI) Coexistence: Emerging Themes and Research Agenda' (2023) 124 Technovation 102747 <www.sciencedirect.com/science/article/pii/S0166497223000585> accessed 1 January 2025.

52 Ekkehardt Ernst, Rossana Merola and Daniel Samaan, 'Economics of Artificial Intelligence: Implications for the Future of Work' (2019) 9 IZA Journal of Labor Policy <https://sciendo.com/article/10.2478/izajolp-2019-0004> accessed 1 January 2025.

This takeover becomes evident when examining the well-studied field of recruitment. Hybrid recruitment, which integrates AI-driven assessment tools with human judgement, appears to be the most effective approach.[53] AI-facilitated recruitment aims to streamline and innovate the process of identifying the right candidate,[54] which typically involves multiple steps: candidate search, screening, selection and onboarding. Organisations may integrate AI into some or all of these stages, depending on their specific practices and policies.[55]

The integration of AI tools across the recruitment process demonstrates how seemingly simple automated tasks can fragment and influence employment decisions. The process begins with job creation, where AI systems generate ideal candidate profiles and optimise job descriptions based on data from previous workers and market analysis. These initial AI interventions, while appearing procedural, already shape the candidate pool by defining success parameters. In the candidate search phase, AI sourcing tools scan various platforms to identify potential candidates. This occurs against the targeted advertising algorithms that can define job posting visibility to specific demographics. This automated pre-selection process, though technically performing basic pattern matching, significantly influences those who learn about existing opportunities.[56]

The screening stage employs AI for resume scanning and candidate ranking through keyword analysis. These systems can compare applications against the previously generated ideal profile and existing successful employees' data. Once again, while this appears to be simple document

53 Kiran Kumar Reddy Yanamala, 'Integration of AI with Traditional Recruitment Methods' (2021) 1 Journal of Advanced Computing Systems 1 <https://scipublication.com/index.php/JACS/article/view/23> accessed 1 January 2025.

54 Mariana Jatoba and others, 'Artificial Intelligence in the Recruitment and Selection: Innovation and Impacts for the Human Resources Management', *Economic and Social Development: Book of Proceedings* (Varazdin Development and Entrepreneurship Agency 2019) <www.proquest.com/docview/2269012279/abstract/7C6C042055DC4BF8PQ/1> accessed 1 January 2025.

55 Manuel F Gonzalez and others, 'Allying with AI? Reactions toward Human-Based, AI/ML-Based, and Augmented Hiring Processes' (2022) 130 Computers in Human Behavior 107179 <https://linkinghub.elsevier.com/retrieve/pii/S0747563222000012> accessed 13 January 2025.

56 Edward Tristram Albert, 'AI in Talent Acquisition: A Review of AI-Applications Used in Recruitment and Selection' (2019) 18 Strategic HR Review 215 <https://www.emerald.com/insight/content/doi/10.1108/SHR-04-2019-0024/full/html> accessed 1 January 2025.

classification, the AI makes crucial filtering decisions determining which candidates advance. During selection, AI-powered assessments, including gamified tests and chatbot interviews, evaluate candidates' skills and initial responses.[57] Though these tools perform narrow procedural tasks in isolation, their combined assessments contribute substantively to candidate evaluation.

The distinction between 'simple procedural tasks' and significant employment decisions becomes particularly demanding when examining the structure of the modern recruitment processes. Rather than deploying a single comprehensive AI system, employers typically implement multiple specialised tools that handle discrete aspects of the recruitment process. While potentially qualifying for derogation under article 6(3) of the AI Act due to its limited scope, each recruitment component collectively forms part of a more complex decision-making apparatus that materially influences outcomes related to workers' access to labour markets.

This fragmentation highlights a critical gap in current regulatory approaches: they fail to account for how multiple 'simple' AI systems can aggregate to significantly impact labour rights and employment decisions significantly. It is therefore that the future regulatory frameworks must consider not just the individual categorisation of AI systems but their collective influence on employment-related decision-making processes.

3.2.2 Limitations of the Regulatory Framework

The current regulatory framework does not capture the complexities of AI integration in various workplace contexts or grasp this fragmented deployment approach. While AI systems might provide inferred data with narrow applications, the rest is left at the discretion of managers and employers. Indeed, the current EU regulation of decision-making systems is narrowly focusing on those decisions that are fully automated from start to finish but indeed provide inputs to employers on certain job issues. Studies in

57 Luis Hernan Contreras Pinochet and others, 'Exploring the Impact of AI on Candidate Selection: A Two-Phase Methodological Approach with CRITIC-WASPAS' (2024) 242 Procedia Computer Science 920 <https://linkinghub.elsevier.com/retrieve/pii/S1877050924019938> accessed 1 January 2025; Byoung-Chol Lee and Bo-Young Kim, 'A Decision-Making Model For Adopting an AI-Generated Recruitment Interview System' (2021) 12 International Journal of Management 548 <https://iaeme.com/Home/article_id/IJM_12_04_046> accessed 13 January 2025.

public law indicate that most real-world systems are rarely fully automated. Instead, AI systems assist and augment human decision-making along the multi-step process.[58] Similar limitations are observed in the workplace. Halefom Abraha discusses that multiple legal domains regulate algorithmic management, including labour law, non-discrimination law and data protection law, though none comprehensively address algorithmic harms.[59]

Reuben Binns and Michael Veale highlight that automated decision-making systems typically manifest in three ways: (1) supporting systems that aid human decision-makers, (2) triaging systems that sort cases in various ways and (3) summarising systems that consolidate human-made assessments. However, these structures present significant complications under article 22 of the GDPR. The key challenges include selective automation, where specific subgroups might effectively receive automated decisions despite nominal human oversight; ambiguity in decision location within multi-stage systems; uncertainty regarding significance thresholds and whether they apply to potential or realised effects; upstream effects where early automated steps can constrain subsequent human decisions; and final step issues where the concluding phase of a process may not be decisive in determining its automated nature. These challenges create practical implications where courts and regulators struggle to develop consistent interpretations as mere addition of human oversight proves insufficient as a solution.[60] In settings where AI systems assist employers with decisions, the lack of clarity of where the automation ends and human decision-making begins highlights the need for regulatory frameworks that are responsive to current realities.

58 Marco Almada, 'Automated Uncertainty: A Research Agenda for Artificial Intelligence in Administrative Decisions' (2023) 16 Review of European Administrative Law 137 <www.ingentaconnect.com/content/10.7590/187479823X16970258030172> accessed 1 January 2025; Francesca Palmiotto, 'When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis' (2024) 25 German Law Journal 210 <www.cambridge.org/core/journals/german-law-journal/article/when-is-a-decision-automated-a-taxonomy-for-a-fundamental-rights-analysis/362AF985585D28E5E762F4FEEF4719B7> accessed 1 January 2025. Palmiotto gives examples of the UK's EU Settlement Scheme to point out the 'automated triage' when the systems determine workflows and assign cases to appropriate human caseworkers.

59 Halefom Abraha, 'Regulating Algorithmic Employment Decisions through Data Protection Law' (2023) 14 European Labour Law Journal 172 <https://doi.org/10.1177/20319525231167317> accessed 1 January 2025.

60 Similar critiques concern AI systems and the regulation of human oversight under the AI Act. See Melanie Fink, 'Human Oversight under article 14 of the EU AI Act' (SSRN 2025) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5147196> accessed 31 August 2025.

Ultimately, the classification of AI systems depends on the 'purpose' of HRAIS. As indicated by the examples above, the 'simplicity' of a task does not determine its significance. If an AI tool has the capability to predict, it already differs in its functionality from a simple automated tool. Even seemingly basic tasks such as filing, ranking or classifying documents can contribute to workers' management, where the manager makes decisions based on these apparently simple procedural tasks. In other words, while a decision may seem inconsequential on the surface, measuring its impact and understanding how it contributes to the potential realisation of protected labour rights can be challenging. Therefore, even if the AI systems that fall under the derogation – individually or in accumulation with other systems – could have a significant impact.

3.3 Profiling of natural persons

Finally, article 6(3) of the AI Act provides that to qualify for derogation, the AI systems should not profile a natural person. The AI Act adopts a strict approach to profiling as it generally prohibits or classifies as high-risk AI systems that profile individuals. This approach is in line with the EU practice and was taken with the view of reducing the risk of discrimination and bias. Recital 67 of the AI Act emphasises that data used for profiling can result in discrimination, particularly against vulnerable groups.

Article 5(1)(d) prohibits profiling AI systems that evaluate or predict the risk of individuals committing criminal offences based solely on personality traits or characteristics. However, once again, this prohibition does not apply to AI systems which are intended to support human assessment. This assessment should be based on objective and verifiable facts that are linked to criminal activity. AI systems that profile natural persons are categorised as high-risk, as per article 6(3), since the act of profiling cannot be considered to have a limited impact on decision-making, as per recital 53.

The AI Act relies on the definitions of article 4(4) of the GDPR, which provides that profiling encompasses three elements: automated data processing, the use of personal data, and the intent to evaluate personal aspects. In employment contexts, article 22 of the GDPR restricts collecting and analysing individual or group data to categorise and predict be-

haviours or capabilities.[61] While prohibition is significant, recruitment AI tools might be designed to appear to perform simple data organisation rather than profiling. AI systems might be engineered to perform analysis that achieves profiling-like outcomes without technically meeting the legal definition of profiling under GDPR article 4(4).

While these systems effectively evaluate and categorise candidates, they are engineered to appear as simple data processing tools. For example, rather than explicitly profiling personality traits, they analyse proxy variables and break assessments into discrete tasks that individually appear as basic pattern matching. A system might claim to perform objective keyword matching between resumes and job requirements while building comprehensive candidate profiles through a pattern of predictive analysis.

The technical implementation further obscures the profiling nature of these tools. Instead of direct personality assessment, they may analyse writing style to infer communication abilities or use correlation analysis to predict candidate success. By fragmenting the evaluation process into micro-tasks like skills assessment, experience verification, and qualification matching, each component avoids classification as profiling. This technical and definitional circumvention could allow recruitment AI to maintain its practical profiling function while potentially avoiding regulatory scrutiny under current regulatory frameworks.

4. Discussion and Conclusion

This chapter centres around the question of whether article 6(3) of the AI Act can compromise the oversight of the AI Act. For workplace management, large-scale rule-bending misclassification of HRAIS appears improbable. There are two main contributing factors to this. Firstly, as the analysis of this chapter indicates, there are preconditions. Individually, as argued in the chapter, they might present vulnerabilities. Cumulative requirements of conducting a FRIA, not materially influencing the outcome of the decision, and the prohibition of natural persons make it difficult to misclassify HRAIS.

Secondly, HRAIS should remain within the AI Act's purview even in cases of regulatory deviation. When an employer does not classify an AI system as high-risk, the AI system still remains subject to regulation under the

61 Data Protection Working Party (n 41).

AI Act. Article 6(4) stipulates that for AI systems not categorised as high-risk, employers must adhere to the obligations outlined in article 49(2) regarding registration. It thereby maintains oversight and accountability in AI deployment within various organisational contexts. Article 49(2) states that if 'the provider has concluded that [AI system] is not high-risk according to [a]rticle 6(3), that provider [...] shall register themselves and that system in the EU database referred to in article 71.' Meanwhile, article 71(4) refers to the database that 'shall be accessible and publicly available in a user-friendly manner.' It continues that 'the information should be easily navigable and machine-readable.' Therefore, the applications that do not qualify as high-risk fall within the 'limited risk' category.

While it is unlikely to circumvent high-risk AI obligations entirely, and though the drafters appear to have negotiated for the most robust version of article 6(3) possible, the impact of overall workers' rights remains limited. This paper demonstrates that workers' rights are falling through the cracks of the supply chain and derive from the obligations of the employers. The critical issue is that HRAIS requirements, while procedurally sound, lack substantive impact – prompting a necessary reconsideration of both employer decision-making roles and AI systems' actual influence. The regulation fails to address the fundamental power imbalance in employment relationships. While the AI Act is not branded as the citadel of labour rights protection, it is evident from this chapter that the advocacy for a new generation of digital rights, more targeted sectoral regulation and strengthened unionisation is needed to address these challenges. These methods must be considered to create a just transition during this wave of enhanced automation driven by AI. The AI Act's failure to address core labour protection demands underscores the necessity for multiple regulatory approaches to ensure a just transition in this era of AI-driven automation. As it stands, the EU AI Act resembles a sieve: while catching some regulatory concerns, it lets many critical issues pass through unresolved. These complexities and potential loopholes demand continued legal vigilance in a system where worker protections remain contingent on employer status rather than being universally guaranteed.

From B2B to B2G and G2G data sharing in competition law. When data become a competitive advantage and an enforcement tool

Isabella Lorenzoni

Abstract: Data has become an extremely important asset for our modern digital economy, mainly because of two key factors: the abundance of data that can be collected and processed (Big Data), and the technological advancement of Artificial Intelligence (AI) and data analytics techniques that make it possible to transform data into meaningful information. Data is not only a valuable resource for companies but also a crucial element for public authorities when developing their own digital enforcement tools. This twofold function of data can be observed in competition law. Firstly, data serves as a competitive advantage for companies. By acquiring insights into users' habits and preferences, they are able to tailor their products to customers' needs and strengthen their position in the market. Secondly, competition authorities have started to develop their own digital enforcement tools which would inevitably succeed or perish depending on data availability.

This chapter aims to analyse the crucial role that data plays in competition dynamics as a competitive advantage for undertakings and as an enforcement tool for competition authorities, addressing potential challenges. It first examines data-driven theories of harm, and which remedies can ensure that data can be fairly accessed and shared. When data is considered an essential input to compete, a balance needs to be struck between freedom to conduct business, preserving competition on the merit on the one hand, and preventing that refusal to share data would amount to an abuse of dominant position on the other. Business to Business (B2B) data sharing can be a solution, achieved by relying on competition law instruments that can offer *ex post* remedies, and on regulations that can mitigate *ex ante* the risk of infringing competition law. Second, this chapter suggests an alternative use of data as a tool to enforce competition law. Competition authorities need to have access to huge amount of data (quantitative data) of the right kind (qualitative data) in machine-readable format, in order to develop reliable and accurate digital enforcement tools. Therefore, mechanisms for encouraging data sharing and cooperation between businesses and competition authorities (B2G data sharing) and between competition authorities and other relevant public bodies (G2G data sharing) are proposed here.

1. Introduction

Every single action we perform online produces data, from liking a social media post to buying a bag from an online shop, and even clicking on a picture of a vacation island.[1] Data is nowadays considered the 'lifeblood' of modern digital businesses.[2] Data is both the input and the output of modern technologies on which digital platforms and businesses rely.[3] It comes as no surprise that companies strive to collect personal and non-personal data, as it represents their main income source.[4]

This chapter aims to analyse the crucial role that data plays in competition dynamics as a competitive advantage for undertakings and as a potential enforcement tool for competition authorities. This chapter first examines the solutions adopted to ensure data exchange between market players (B2B data sharing). It then considers recommendations to enhance cooperation and data sharing between private companies and competition authorities (B2G data sharing) and between competition authorities and other public enforcers (G2G data sharing). In order to investigate the topic, section 2 is dedicated to analysing data as a resource for digital business models and its value for new technologies in the Artificial Intelligence (AI) chain. Section 3 delves into the topic of data as a competitive advantage and how competition law *ex post*, and regulations *ex ante* can mitigate the risk of abusing dominant positions and enhance B2B data sharing. Section 4 looks at the data from the side of enforcers and how data has become a crucial element also for competition authorities to enforce competition law.

1 Lothar Determann and Teisha Johnson, 'Data Broker Regulation – Competition v. Privacy Considerations: Trade-Offs' (2023) CPI TechREG Chronicle <https://insightplus.bakermckenzie.com/bm/attachment_dw.action?attkey=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQJsWJiCH2WAWuU9AaVDeFgpCdzlkUxiWH&nav=FRbANEucS95NMLRN47z%2BeeOgEFCt8EGQbuwypnpZjc4%3D&attdocparam=pB7HEsg%2FZ312Bk8OIuOIH1c%2BY4beLEAeoUASUHJpPzQ%3D&fromContentView=1> accessed 16 September 2024.

2 Marixenia Davilla, 'Is Big Data a Different Kind of Animal? The Treatment of Big Data Under the EU Competition Rules' (2017) 8 Journal of European Competition Law & Practice 370.

3 OECD, 'Artificial intelligence, data and competition – Background Note' (2024).

4 American Bar Association Antitrust Law Section, 'Artificial Intelligence & Machine Learning: Emerging Legal and Self-Regulatory Considerations, Part Two. Competition Implications of Big Data and Artificial Intelligence/Machine Learning' (Big Data Task Force, February 2021) <https://www.americanbar.org/news/abanews/aba-news-archives/2021/02/aba-antitrust-law-section-releases-part-two-of-white-paper-on-ai/> accessed 11 September 2024.

Mechanisms for ensuring data sharing among different actors relevant for competition law are suggested, in the light of the current European Union (EU) digital legislation which encourages data sharing between businesses and public authorities (B2G) and between public authorities among each other (G2G). Section 5 concludes.

2. The importance of data as input and output for modern business models

Data is often described as the 'new oil of the 21st century',[5] and as the 'new currency of the digital age'.[6] Unsurprisingly, companies strive to acquire data, and just like unrefined oil, they need to process it, analyse it and transform it into valuable assets for their business.[7] Collecting and

5 This expression was coined by the British mathematician Clive Humby in 2006 according to whom 'Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analysed for it to have value.' <https://towardsdatascience.com/is-data-really-the-new-oil-in-the-21st-century-17d014811b88> accessed 16 September 2024; <https://futurescot.com/why-data-is-the-new-oil/> accessed 16 September 2024; Sofia Ranchordás and Giovanni De Gregorio, 'Breaking Down Information Silos with Big Data: A Legal Analysis of Data Sharing' (2019) University of Groningen Faculty of Law Research Paper Series No. 44/2019 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3466313> accessed 16 September 2024; Davilla (n 2). See also the criticism expressed by Maggiolino and Colangelo concerning this expression: 'Differently from an oil well, big data do not consist of a huge amount of the same product because, with the exception of duplicates, digital data are different from each other. […] although there is a market for oil, there is no market for a resource named 'big data' and – even more importantly – there is no point in discussing whether firms compete in the market for big data or whether collectors, aggregators and analysers of big data operate in this same market. […] Neither can big data be compared with oil as such. It is well known that oil has some specific features: it is black and viscous. Big data do not have the same constant properties. Moving from one firm to another firm, not even the amount of data collected is the same'. Giuseppe Colangelo and Mariateresa Maggiolino, 'Big Data as Misleading Facilities' (2018) Bocconi Legal Studies Research Paper No. 2978465, 3.

6 Viktoria H S E Robertson, 'Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the Era of Big Data' (2020) 57 Common Market Law Review 161. See also Maurice E Stucke, 'Should we be concerned about Data-opolies?' (2018) 2 Georgetown Law Technology Review 275 and OECD, 'The intersection between competition and data privacy – Background Note' (2024) and Monika Woźniak-Cichuta, 'Digital Data-Driven Mergers: Is a Data-Sharing Remedy a Panacea?' (2024) 17 YARS 9.

7 Amol Mavuduru, 'Is Data Really the New Oil in the 21st Century? Exploring the strengths and limitations of this metaphor in the information age' (towards data science, 2020) <https://towardsdatascience.com/is-data-really-the-new-oil-in-the-21st-

analysing data is not a practice that has emerged with the digital economy.[8] Companies have always been interested in acquiring users' data in order to learn their preferences and deliver products adapted to their needs.

Marketing strategies revolve around the following purposes: studying consumers' behaviours through collection and analysis of their data.[9] What has changed now is the emergence of two essential factors: an enormous amount of data that can be harvested (Big Data) and technologies that allow to process raw data and extract valuable information (Big Data Analytics with AI and machine learning algorithms).[10] Digital companies, but also traditional brick and mortar stores are 'avid collectors and users of data'.[11] Collecting, processing and analysing data has become a successful strategy to compete for many business models.[12] One in particular, the multi-sided platform model, such as Google or Facebook, collects consumers' data from one side of the platform in exchange for services and products usually offered for free and then sell data for targeted advertisement.[13]

Big Data is typically described through the four 'Vs': volume, velocity, variety, and value.[14] *Volume* refers to the amount of data collected which is

century-17d014811b88> accessed 16 September 2024; Matt Watts, 'Why data is the new oil' (Futurescout, 2021) <https://futurescot.com/why-data-is-the-new-oil/> accessed 16 September 2024. See also American Bar Association (n 4).

8 Autorité de la Concurrence and Bundeskartellamt, 'Competition Law and Data' (2016).

9 ibid.

10 ibid.

11 American Bar Association (n 4) 20.

12 ibid.

13 ibid. Autorité de la Concurrence and Bundeskartellamt (n 8). Consumers are also subject to different manipulation techniques from digital companies that harvest their data. This is the phenomenon of 'behavioral discrimination', which refers to the ability of companies to use consumers' data and technologies to influence their behaviors towards certain products, incentivize overall consumption and set personalized services and prices to different categories of consumers. See in this regard Ariel Ezrachi and Maurice E Stucke, *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy* (Harvard University Press 2016) and Shin-Ru Cheng, 'Competition Law and Behavioral Discrimination: Regulatory Pitfalls and New Opportunities' (2022) 52 University of Memphis Law Review, 927.

14 Christophe Samuel Hutchinson, 'Potential abuses of dominance by big tech through their use of Big Data and AI' (2022) 10 Journal of Antitrust Enforcement 443; Davilla (n 2); Autorité de la Concurrence et Bundeskartellamt (n 8); OECD, 'Consumer Data Rights and Competition – Background note' (2020); OECD (n 3); Thomas Tombal, *Imposing Data Sharing among Private Actors A Tale of Evolving Balances* (Kluwer Law International 2022).

a consequence of an increase in online activities by users. This is possible because of more interconnected devices, so-called Internet of Things (IoT), combined with lower costs for collecting, storing and processing data.[15] *Velocity* relates to the speed at which data is collected, processed and analysed, with nearly real-time data sharing.[16] *Variety* concerns the possibility of collecting multiple kinds of information from different sources, such as users' purchase history, website visits, social media interactions.[17] All these features add *Value* to the data collected, which allows companies to generate revenues by making accurate predictions and targeted advertisements.[18]

Data can be classified in several ways. One way refers to whether data is personal or non-personal. According to the GDPR,[19] personal data refers to 'any information relating to an identified or identifiable natural person (data subject)'.[20] Non-personal data are any other data that are not personal which can be produced by interconnected devices (IoT, like sensors in self-drive cars) or they can derive from anonymisation of personal data, where the data subject is no longer identifiable.[21] Big Data usually comprises both personal and non-personal data whose separation in a given dataset can be rather difficult.[22]

Furthermore, data can be categorised depending on the ways in which they are obtained.[23] Firstly, data can be voluntarily or 'actively' provided by individuals who intentionally hand out their information, either because they want to do so (e.g. when filling consumers' forms with personal information) or because they do not have alternative choices (e.g. specific information required by a bank).[24] Secondly, data are passively provided by individuals, consciously or unconsciously, as their online activities are

15 Hutchinson (n 14); OECD (n 14); Davilla (n 2).
16 Hutchinson (n 14); OECD (n 14); Davilla (n 2).
17 Hutchinson (n 14); OECD (n 14); Davilla (n 2).
18 Hutchinson (n 14); OECD (n 14); Davilla (n 2); Tombal (n 14).
19 Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
20 Article 4(1) GDPR; Tombal (n 14); OECD (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).
21 Tombal (n 14).
22 ibid.
23 OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).
24 Hutchinson (n 14); OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).

recorded (e.g. location tracking, web browsing) which are known under the name of 'observed data'.[25] Thirdly, data can be inferred through data analytics techniques that identify correlations in the previously collected data (either volunteered or observed data) to predict consumers behaviours and other characteristics (e.g. data inferred for credit scoring). This inferred data represents the real value for companies, as they can refine their products or tailor advertisement based on it.[26] Fourthly, data can also be acquired from third parties (e.g. from data brokers) through licensing agreements or mandatory data sharing obligations.[27]

Finally, data are usually divided into unstructured and structured data.[28] When first collected, data are unstructured or 'raw data' and they typically do not generate value unless transformed into structured data, from which meaningful information and knowledge can be extracted.[29] This process is nowadays possible thanks to Big Data analytics which employs machine learning techniques to draw inferences from data, which are in turn used for prediction or decision-making.[30]

25 Hutchinson (n 14); OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).

26 Hutchinson (n 14); OECD (n 14); Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8).

27 OECD (n 14); Tombal (n 14); Maureen K Ohlhausen, 'Privacy and Competition: Friends, Foes, or Frenemies?' (2019) CPI Antitrust Chronicle <https://www.competitionpolicyinternational.com/wp-content/uploads/2019/02/CPI-Ohlhausen.pdf> accessed 14 September 2024.

28 Autorité de la Concurrence and Bundeskartellamt (n 8).

29 Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8); American Bar Association (n 4) 'Data is typically not an end in itself, but rather is useful when inferences can be drawn from it' 9. See also Ranchordás and De Gregorio (n 5) '[…] 'raw data', that is, 'signs, patterns, characters, or symbols that can convey information about facts, concepts, processes, or objects.' Data becomes information when a 'datum' or 'given fact' is placed in a certain context and is hence labelled. 'Information' refers thus to the meaning assigned to data' 10. See also Colangelo and Maggiolino (n 5) according to whom 'the value of big data lies in disclosing knowledge; that is, in revealing hidden pieces of information. Firms succeed in developing better products and services when they make data meaningful; that is, when they infer correlations and elaborate predictions as to consumers' tastes and rivals' strategies by using particular analytical tools' 23.

30 Tombal (n 14); American Bar Association (n 4); Ranchordás and De Gregorio (n 5) 'Data analytics employs for example algorithms to make sense of streams of data, identify relationships across unconnected datasets, and identify or predict behaviour and preferences' 6–7.

Overall, data are valuable assets for companies that can dispose of AI and machine learning techniques to rapidly analyse and transform them into monetizable information.[31] At the same time, AI systems need to access large datasets to function, whose performance and level of accuracy increase when big data are fed into 'data-hungry AI algorithms'.[32] Hence, data is both the input on which AI systems are trained and the output of AI applications, deployed to analyse data and extract relevant information.[33] Especially in large language models (LLMs), data is an extremely important requirement used to train foundation models. In fact, data plays a crucial role in each stage of the 'GenAI value chain'.[34] Quantity and quality of data affect a foundation model's output, as '[a]n LLM will flourish or wither depending on the data it is trained on; and nothing can make up for that.'[35]

31 Hutchinson (n 14).

32 American Bar Association (n 4).

33 OECD (n 3).

34 Stefan Hunt, Wen Jian, Aman Mawar and Bartley Tablante, 'You Are What You Eat: Nurturing Data Markets to Sustain Healthy Generative AI' (2023) CPI 1 TechREG Chronicle <https://www.keystone.ai/wp-content/uploads/2023/11/NURTURING-DATA-MARKETS-TO-SUSTAIN-HEALTHY-GENAI-INNOVATION-Stefan-Hunt-Wen-Jian-Aman-Mawar-Bartley-Tablante-2.pdf> accessed 6 September 2024, 3, '[f]oundation models are made by training a machine learning algorithm (called pre-training in this context) using huge datasets to produce a model that can be refined and used in many downstream applications. […] Fine-tuned models are foundation models that are refined through additional training on a narrower set of use case specific data. […] Grounded models have access to additional data sources, allowing the model access to information (e.g. real-time news) beyond the pre-training and fine-tuning data'. See also Jeppe Christensen, 'What Is Generative AI?' (neo4j, 2024) <https://neo4j.com/blog/what-is-generative-ai/?utm_source=GSearch&utm_medium=PaidSearch&utm_campaign=Evergreen&utm_content=EMEA-Search-SEMCE-DSA-None-SEM-SEM-NonABM&utm_term=&utm_adgroup=DSA-GenAI&gad_source=1&gclid=CjwKCAiA2cu9BhBhEiwAft6IxMS1Gnek9QBgVZFubiR2fS_B7E5LjyTh-3blqVDD1zyivpSXrBR2ZxoCAuYQAvD_BwE> accessed 17 February 2025 according to whom 'Generative AI, or GenAI, is short for generative artificial intelligence. GenAI is a subset of artificial intelligence (AI) that creates new content based on learned patterns and rules, such as text, images, audio, and code. Unlike traditional AI systems designed to recognize or classify existing data, generative AI models can produce novel outputs that resemble the data they were trained on'.

35 Hunt, Jian, Mawar and Tablante (n 34).

3. Data as a competitive advantage: B2B data sharing obligations

How companies collect, process and use consumers' data was usually seen as a matter for data protection or consumer law and not a competition concern.[36] However, when data becomes a key factor around which companies compete and strengthen their market position, 'the question of how that company handles the personal data of its users is not only relevant for data protection authorities, but also for competition authorities.'[37] This section will first analyse different theories of harm that involve data and thereafter it will look at the mechanisms to impose *ex post* and *ex ante* data sharing obligations.

3.1 Data-driven theories of harm

Different theories of harm that involve the use of data have been conceived in competition law. Firstly, '*data-driven mergers*'[38] occur when, in order to increase data access, companies are incentivised to acquire other firms with the aim of combining and merging their datasets.[39] In data-driven markets, such mergers could provide a competitive advantage, as the new merging entity would have larger datasets and therefore more data to profile indi-

36 Autorité de la Concurrence and Bundeskartellamt (n 8); OECD (n 6).

37 Bundeskartellamt, 'Bundeskartellamt prohibits Facebook from combining user data from different sources Background information on the Bundeskartellamt's Facebook proceeding' (2019) 6 <https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Diskussions_Hintergrundpapiere/2019/07_02_20219_Hintergrundpapier_Facebook.pdf?__blob=publicationFile&v=2> accessed 5 August 2024. See also OECD (2024) (n 6); Peter Stauber, 'Facebook's Abuse Investigation in Germany and Some Thoughts on Cooperation Between Antitrust and Data Protection Authorities' (2019) 2 CPI Antitrust Chronicle 36 according to whom '[i]f access to data is an essential factor for the competitive position of the company – as is the case with data-driven products such as social networks – the handling of personal data by a company is not only a case for data protection authorities, but also for antitrust authorities' 6.

38 Christophe Carugati, 'The Antitrust Privacy Dilemma' (2021) working paper <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3968829> accessed 28 August 2024. 'The notion of data-driven mergers refers to transactions which are motivated by an acquisition of a huge amount of data' in Woźniak-Cichuta (n 6) 14.

39 Autorité de la Concurrence and Bundeskartellamt (n 8), OECD (n 6) and Woźniak-Cichuta (n 6).

viduals.[40] In fact, '[t]he linkage of these data can give companies more insights into user habits, enabling them to further improve their services and reinforce their market position. Generally speaking, the more data a company can combine, the better its chances to gain knowledge that can be used to strengthen its market position.'[41] Concentration of data could raise competition concerns, as the merging firm may enhance its market power and raise barriers to entry and costs for competitors who may not be able to replicate the data resulting from combination of their datasets.[42] In the past, several data-driven mergers have been cleared at the EU level, despite concerns that combination of datasets would have given rise to anticompetitive issues.[43]

Secondly, when it comes to *collusive practices and cartels* under article 101 Treaty on the Functioning of the European Union (TFEU), agreements between companies aimed at maximising data collection and consumer information may be seen as a competition concern.[44] Collusion may also occur when data are illegally shared among competitors to get insights of respective commercial activities and competitive strategies, with the aim of coordinating their actions instead of competing with each other.[45]

40 Giuseppe Colangelo and Mariateresa Maggiolino, 'Data Protection in Attention Markets: Protecting Privacy through Competition?' (2017) 8 Journal of European Competition Law & Practice 363; Autorité de la Concurrence and Bundeskartellamt (n 8) and Woźniak-Cichuta (n 6).

41 Nils-Peter Schepp and Achim Wambach, 'On Big Data and Its Relevance for Market Power Assessment' (2016) 7 Journal of European Competition Law & Practice 120, 121 and Daniele Condorelli and Jorge Padilla, 'Harnessing Platform Envelopment in the Digital World' (2020) 16 Journal of Competition Law & Economics 143. See also Woźniak-Cichuta (n 6), according to whom 'the value comes from the large combination of a variety of data collected from different sources (e.g., data analytics' sophistication), […] the incentive for big data firms will be to aim to expand the variety of data collected' 36.

42 Autorité de la Concurrence and Bundeskartellamt (n 8); OECD (n 6); OECD (n 14); Davilla (n 2).

43 See for instance *Google/DoubleClick* (Case COMP/M.4731) Commission decision [2008] OJ C184/10; *Facebook/WhatsApp* (Case COMP/M.7217) Commission Decision [2014] OJ C417/4; *Microsoft/LinkedIn* (Case COMP/M.8124) Commission decision [2016] OJ C 388/4. See also OECD (n 6); OECD (n 14); Davilla (n 2); Colangelo and Maggiolino (n 40) and Woźniak-Cichuta (n 6).

44 OECD (n 14).

45 Davilla (n 2). See also Woźniak-Cichuta (n 6) '[a]nother challenging aspect of data-sharing remedies is that they can facilitate the creation of anticompetitive agreements' 37. For a deeper analysis on the risk of collusion brought by data sharing see Eugenio

Thirdly, in data-driven markets, abuse of dominance, according to article 102 TFEU, can occur under two main theories of harm: *exclusionary abuses* and *exploitative abuses*. Exclusionary conducts refer to practices that have a foreclosure effect by denying access to other competitors.[46] This can happen when a company who has a dominant position refuses to deal and to provide access to a fundamental input without which rivals cannot compete in a market.[47] Refusal to access data may constitute an abuse of dominant position if data are considered an 'essential facility'.[48] Furthermore, as an exclusionary conduct, 'where a dominant firm has exclusive access to consumer data, it could attempt to raise rivals' costs or barriers to entry by engaging in tying or bundling.'[49] Finally exploitative abuses can take the form of excessive data collection of users' data, when a company reduces the level of privacy protection and increase collection of data which is excessive or unfair.[50] The leading case is *Meta Platforms*,[51] where the German Competition authority (the Bundeskartellamt) found Facebook infringing competition law for abusing its dominant position in the market for social networks, by imposing unfair terms and conditions that resulted in excessive collection of users' data.[52]

3.2 Data as an essential facility: ex post B2B data sharing

Data is considered a 'non-rivalrous' good, which means that the use of certain data by a company does not prevent, in principle, other companies

Olmedo-Peralta, 'The Creation of Data Pools as Information Exchanges: Antitrust Concerns' (2024) 17 YARS 49 at 70 ss.

46 Hutchinson (n 14).

47 ibid.

48 ibid; Autorité de la Concurrence and Bundeskartellamt (n 8); OECD (n 6); OECD (n 14) and Davilla (n 2).

49 OECD (n 14). See also Hutchinson (n 14) and Autorité de la Concurrence and Bundeskartellamt (n 8), according to whom tying practices can happen when 'a company owning a valuable dataset ties access to it to the use of its own data analytics services. As it noted, such tied sales may increase efficiency in some circumstances but they could also reduce competition by giving a favorable position to that company which owned the dataset over its competitors on the market for data analytics' 20.

50 Stucke (n 6); OECD (n 6); OECD (n 14); Hutchinson (n 14), Davilla (n 2).

51 See Decision B6–22/16 of 6 February 2019 of the Bundeskartellamt and C-252/21 *Meta Platforms Inc. et al. v Bundeskartellamt* [2023] EU:C:2023:537.

52 See for instance Stauber (n 37); OECD (n 6); OECD (n 14); Hutchinson (n 14).

from accessing and using the same data.[53] From a competition point of view, data can enhance innovation and maintain competition on the markets. However, data may also represent a competitive advantage for few big companies if certain data cannot be replicated by competitors.[54] For example, looking at LLMs, which are trained on billions of data, even though many datasets are publicly available and open source, 'some large datasets are proprietary and may provide unique insights that others struggle to replicate, noting as an example Google's ownership of YouTube and the potential to control access to its video transcripts'.[55] Hence, specific data could be a crucial factor for companies to compete, and therefore they may be regarded as 'close to essential services'.[56]

If data can be considered an essential input without which rivals would not be able to compete, competition law may impose compulsory B2B data sharing as an *ex post* remedy, when refusal to share an essential resource (refusal to deal) amounts to an abuse of dominant position under article 102 TFEU.[57] The tool traditionally used to restore competition in such

53 Davilla (n 2); American Bar Association (n 4) 13 '[d]ata is typically non-rivalrous. At the outset a firm's use of most data does not typically reduce the data's availability to competitors—that is, data is generally non-rivalrous. Data can be replicated, provided to multiple suppliers by buyers, shared between businesses, and gathered by multiple entities, all of which make market power through data less likely'. See also Colangelo and Maggiolino (n 5) 6.

54 OECD (n 3) 30.

55 OECD (n 3). See also Hunt, Jian, Mawar and Tablante (n 34) according to whom 'superior access to data could give some players a significant advantage: the largest AI players have access to proprietary datasets, e.g. YouTube data by Google, and train their models on them' 24–25.

56 OECD (n 3) 30. For example, in the field of Generative AI, Hunt, Jian, Mawar and Tablante (n 34) suggest that '[g]iven the potential for such challenges, markets for data for pre-training, tuning, and grounding might need nurturing from regulators, to preserve healthy GenAI competition and alleviate consumer protection issues. Agencies might need to start getting 'under the hood' more actively, for instance by monitoring the data AI companies are accessing and using. Regulators should also consider making this information at least partially available to some parties through transparency requirements, as the EU is proposing with the AI Act. Other regulatory responses to nurture data markets could include monitoring for harmful exclusionary vertical agreements, requirements on data sharing or other rules' 25.

57 Tombal (n 14); Oscar Borgogno and Giuseppe Colangelo, 'Data sharing and interoperability: Fostering innovation and competition through APIs' (2019) 35 Computer Law & Security Review 1 and Hutchinson (n 14).

cases is the 'Essential Facility Doctrine (EFD)' and only in exceptional circumstances access can be obtained.[58]

The EFD is considered 'one of the most controversial antitrust issues',[59] as it imposes to a monopolist that owns an essential input an obligation to share it with anyone who needs it, including competitors, limiting its rights to property and its freedom to contract.[60] When granting access to a crucial facility, it is important to strike a balance between the risk of dominant undertakings to abuse their position by excluding others from the facility they own, extending their monopoly to other markets, and the risk of competitors to free ride on investments made by dominant companies to obtain the essential input.[61]

According to the EFD, as construed by the European Court of Justice in several case law, refusal by a dominant undertaking to provide access to its own facility to other undertakings may be an abusive conduct only in exceptional circumstances and when some conditions are met.[62] Firstly, the resource for which access is sought must be indispensable for carrying out a business activity on a secondary market. As specified in *Bronner*,[63] in order

58 Borgogno and Colangelo (n 57); Colangelo and Maggiolino (n 5). The *hiQ v LinkedIn* case provides an interesting example of data as an essential facility *hiQ Labs, Inc. v. LinkedIn Corporation*, No. 17-cv-03301-EMC. See Ohlhausen (n 27) and Tombal (n 14).

59 Colangelo and Maggiolino (n 5) 13 and Borgogno and Colangelo (n 57).

60 Colangelo and Maggiolino (n 5); Borgogno and Colangelo (n 57) and Tombal (n 14). See also C-48/22 P *Google LLC, Alphabet Inc. (Google Shopping) v Commission* [2024] EU:C:2024:726, 91; opinion of Advocate General Kokott C-48/22 P *Google LLC, Alphabet Inc. (Google Shopping) v Commission* [2024] EU:C:2024:14, 86 and opinion of Advocate General Medina C-233/23 *Alphabet Inc. et al. v Autorità Garante della Concorrenza e del Mercato* [2024] EU:C:2024:694, 32.

61 Colangelo and Maggiolino (n 5) and Borgogno and Colangelo (n 57). Tombal (n 14) 'a balance must be found between incentivising innovation (data collection and processing by data holders) and maximising social welfare through large dissemination (data sharing)' 224. See also AG Kokott opinion C-48/22 P (n 60) 87.

62 Tombal (n 14). On the conditions to apply the EFD see more recently C-48/22 P *Google Shopping* (n 60) 89, 90, 109 – 112. See also AG Medina opinion C-233/23 (n 60) according to whom 'in order to determine whether the Bronner conditions apply to a case concerning a refusal to grant access, it is necessary to discern whether the infrastructure to which access is requested is to be consecrated to the dominant undertaking's own business and use and to be enjoyed exclusively by it, as a way of preserving the benefits of the investments made for the development of that infrastructure. By contrast, those conditions are not intended to apply where the infrastructure concerned is opened to other operators on the market' 35. See also 28, 56.

63 C-7/97 *Bronner* [1998] EU:C:1998:569.

to be indispensable, no actual or potential alternatives must be available, and it is not enough to argue that it is not economically viable.[64] There must be technical, legal or economic obstacles or being unreasonably difficult for any undertaking to replicate the essential facility, alone or in cooperation with other companies.[65] Secondly, such refusal should likely exclude competition on a secondary market that the dominant company reserves to itself, and according to the *IMS Health* case,[66] it is sufficient to identify a potential or hypothetical market.[67] Thirdly, the refusal should prevent the appearance of a new product or service, or of technical improvements (*Microsoft* case[68]), for which there is potential consumer demand and no offer by the dominant undertaking.[69] Fourthly, there should be no objective justifications for such refusal.[70]

Compared to traditional resources, data present specific characteristics. Some scholars claim that the EFD should be adapted accordingly when applied to a refusal to share data.[71] As far as the first condition is concerned, access seekers must still demonstrate the indispensability of data, as raw or unstructured data, when such data cannot be replicated and it is not possible to find a substitute.[72] Only in exceptional circumstances, when the 'data holder is the sole source of a specific type of data, the access to which is indispensable for another undertaking evolving on a downstream

64 *Bronner* 45.
65 *Bronner* 44. Tombal (n 14); Colangelo and Maggiolino (n 5).
66 C-418/01 *IMS Health* [2004] EU:C:2004:257.
67 *IMS Health.* Tombal (n 14); Colangelo and Maggiolino (n 5) 'It is sufficient to demonstrate the existence of two interconnected markets rather than proving the effective commercialization of the essential inputs' 17.
68 T-201/04 *Microsoft v. Commission* [2007] EU:T:2007:289.
69 Tombal (n 14); Colangelo and Maggiolino (n 5).
70 Tombal (n 14); Colangelo and Maggiolino (n 5).
71 For instance, according to Bertin Martens, Alexandre de Streel, Inge Graef, Thomas Tombal and Néstor Duch-Brown, 'Business-to-Business data sharing: An economic and legal analysis' (2020) Digital Economy Working Paper 2020–05, European Commission, JRC121366 at 36–37 '[…] the non-rivalry and wide availability of data will often render a dataset non indispensable' and therefore when referring to access to data 'an expansion of the interpretation of the essential facilities doctrine beyond cases of leveraging and a lower threshold for the condition of indispensability (as well as new product)' should apply. See also Tombal (n 14) and Davilla (n 2).
72 Tombal (n 14); Autorité de la Concurrence and Bundeskartellamt (n 8) such requirement 'would only be met, if it is demonstrated that the data owned by the incumbent is truly unique and that there is no possibility for the competitor to obtain the data that it needs to perform its services' 18.

market',[73] the EFD may apply as an *ex post* compulsory B2B data sharing remedy.[74] Furthermore, according to Tombal, the threshold for the second and third conditions should be lowered.[75] Finally, the last condition of objective justifications must be thoroughly assessed by competition authorities, especially when big tech companies refuse to share data on the ground of protection of users' data, which may only be a defence shield for abusing their dominant position.[76]

Other scholars are of different opinions. For instance, Davilla suggests that instead of relying on the EFD, whose conditions may be too difficult to apply when data are involved, an *ad hoc* test which takes into consideration the specificities of data should be developed.[77] Colangelo and Maggiolino strongly oppose the application of the EFD to data for several reasons.[78] Firstly, they state that the third condition which requires a new or improved product may be unsuitable for data, as access seekers usually do not know in advance what type of product they may develop with the data obtained as they may not know the information that data can reveal.[79] Arguably, data 'are not gathered and organized in order to answer specific research questions. Rather the opposite: firstly, they are amassed from many and

73 Tombal (n 14) 241.

74 Tombal (n 14).

75 ibid.

76 '[C]ompetition authorities should require these large data holders to lay down and substantiate the data protection concerns they raise in order to refuse data sharing with third parties, and they should collaborate with data protection authorities in order to assess whether the data protection standards imposed on third parties by these large firms are (suspiciously) higher than the ones they apply to themselves' Thomas Tombal, 'Data protection and competition law: friends or foes regarding data sharing?' (Accepted paper for the TILTing Perspectives 2021 Conference: Regulating in Times of Crisis) 22. See also Tombal (n 14) 255; OECD (n 6) and Woźniak-Cichuta (n 6).

77 Davilla (n 2) 380. See also Woźniak-Cichuta (n 6), according to whom '[…] in order to formulate a data-sharing remedy […] other criteria should be formulated, rather than those set out in the essential facilities doctrine' 40; and Olmedo-Peralta (n 45) according to whom 'it is fair to view the application of the essential facilities doctrine to data pools as not appropriate, considering the inherently duplicable nature of data, the multiple ways in which they can be accessed, and the possibility of building competing data consortia, as no one controls an indispensable non-duplicable resource for the market. In this case, it will be a market failure that must be approached through other regulatory instruments, theories of harm, or applicative tools' 69.

78 Colangelo and Maggiolino (n 5).

79 ibid; and Giuseppe Colangelo and Mariateresa Maggiolino, 'Data access and AI: Antitrust vs. Regulation' (2018).

varied sources, then they are analysed to make them speak'.[80] Secondly, according to them, data is the 'wrong target', as what is truly essential is not the data as such but the information they provide and the analytical tools used to extract such information, as 'just by looking at big data no one can guess the information they include and – a fortiori – no one can say in advance what pieces of information among the many that can be extracted from big data will be essential [...]'.[81]

When imposing data sharing as an *ex post* remedy to restore competition on the market, attention should be paid to the categories of data that the data holder should share with the undertaking which seeks access to it.[82] Firstly, data that are commercially available in data marketplaces should not be subject to the remedy, as they can be acquired from data brokers or other third parties, just like the dominant undertaking.[83] Secondly, according to the previous distinction, only volunteered and observed data should fall within the scope of a data sharing remedy as 'observed data, [...] often cannot be replicated, and volunteered data [...] would take a significant amount of effort to volunteer again'.[84] Such volunteered and observed data need to be indispensable for the undertaking that seeks access to it, in order to carry out its business in a secondary market.[85] Finally, inferred data should fall out the scope of the remedy, because they are the result of other investments made by the data holder that involve the use of Big Data analytics in the form of machine learning and AI-applied solutions to extract valuable information for the company to innovate.[86] Hence, once raw data (volunteered and observed data) are shared with undertakings requesting such access, they should apply their own technology and extract relevant insights for their business.[87] However, the remedy should only cover a subset of data and not all the data of the dominant undertaking.[88] It may be a difficult task, given the fact that big data are usually collected

80 Colangelo and Maggiolino (n 5) 21.
81 Colangelo and Maggiolino (n 5) 24.
82 Tombal (n 14).
83 ibid; Colangelo and Maggiolino (n 5).
84 Jacques Crémer, Yves-Alexandre de Montjoye and Heike Schweitzer, 'Competition Policy for the digital era' (2019) Final report for the European Commission <https://op.europa.eu/en/publication-detail/-/publication/21dc175c-7b76-11e9-9f05-01aa75ed71a1/language-en> accessed 16 September 2024, 101. See also Tombal (n 14).
85 Tombal (n 14).
86 ibid; Crémer, de Montjoye and Schweitzer (n 84).
87 Tombal (n 14).
88 Tombal (n 14); Colangelo and Maggiolino (n 5) 'some digital data'.

'untidily'.[89] Another problem when imposing B2B data sharing is to consider whether one-time data access would be able to restore competition, as according to the life cycle of data, access may be needed 'on a constant basis and in (near) real time'.[90] However, imposing long-term data access commitments may be an excessively restrictive measure which would result in over-enforcement.[91]

Finally, data protection issues should be carefully assessed when imposing B2B data sharing as an *ex post* remedy. For instance, in the *GDF Suez* case,[92] the French competition authority obliged a company to share its customers' list with other competitors, and the data protection authority expressed concerns about infringing data protection rules as personal data were contained in the list. Therefore, users were given the choice to object to share their data with other companies.[93] This calls for a more coordinated approach and cooperation between competition and data protection authorities, whenever *ex post* compulsory data sharing may raise data protection concerns.[94] Furthermore, data protection issues could be overcome by implementing anonymisation[95] and pseudonymisation[96] techniques, especially when access to data is necessary for the purpose of developing AI tools, as 'AI may need a huge and diversified amount of data, but it does not necessarily want to understand who individuals are and what they do: not necessarily AI needs data that serve to identify single persons or that make

89 Colangelo and Maggiolino (n 5).
90 Tombal (n 14) 305. See also Colangelo and Maggiolino (n 5).
91 Woźniak-Cichuta (n 6).
92 Autorité de la Concurrence, decision N° 14-MC-02 of 9 September 2014 (Direct Energie vs GDF- Suez). See also <https://www.autoritedelaconcurrence.fr/fr/decision/relative-une-demande-de-mesures-conservatoires-presentee-par-la-societe-direct-energie> accessed 20 August 2025 and <https://www.autoritedelaconcurrence.fr/en/communiques-de-presse/9-september-2014-gas-market> accessed 16 September 2024.
93 OECD (n 6); Carugati (n 38); Christophe Carugati, 'Overview of privacy in cases relevant to competition law' [2023] Concurrences 1.
94 OECD (n 6); Carugati (n 38).
95 'Anonymization refers to the process of either encrypting or removing personally identifiable information from datasets, such that the people whom the data describe (data subjects) remain anonymous' Jens Prüfer, 'Competition Policy and Data Sharing on Data-driven Markets Steps Towards Legal Implementation' (Friedrich-Ebert-Stiftung project 2020) <https://library.fes.de/pdf-files/fes/15999.pdf> accessed 26 August 2024, 12.
96 'Pseudonymization refers to the process of replacing personally identifiable information fields by one or more artificial identifiers, or pseudonyms' Prüfer (n 95) 12.

them identifiable. AI can work also with anonymized and pseudonymized data'.[97]

3.3 Ex ante B2B data sharing

Compulsory B2B data sharing can also be imposed by regulations as an *ex ante* remedy, before starting an investigation and before causing harm to consumers and competitors,[98] to remedy to situations where refusal to share data may amount to anticompetitive behaviours under specific circumstances.[99] Firstly, *ex ante* B2B data sharing can be imposed by sectorial legislations[100] which have the advantage of being 'more targeted and adapted to the sector's needs, characteristics and data standardisation challenges'.[101] However, sectorial legislations may present as a downside to provide a data advantage to large data holders, who could use sectorial data to refine other data-driven services that they offer, strengthening their position in those markets, as 'these large data holders [would] have a 360° view on the consumers' preference'.[102]

Secondly, compulsory B2B data sharing can also be imposed by horizontal regulations when specific circumstances require an *ex ante* intervention to remedy to market failures, such as data concentration and data conglomeration, that cannot be dealt efficiently only by relying on *ex post* compe-

97 Colangelo and Maggiolino (n 79) 6.

98 Tombal (n 14) and OECD, 'G7 inventory of new rules for digital markets: Analytical note' (2023).

99 OECD (n 98).

100 See for instance in the sector of banking with the PSD2 Directive (Directive (EU) 2015/2366 of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC) and in the automotive sector (Regulation (EU) 2018/858 of 30 May 2018 on the approval and market surveillance of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles, amending Regulations (EC) No 715/2007 and (EC) No 595/2009 and repealing Directive 2007/46/EC), which imposes on car manufacturers to share data of on-board diagnostics, repair and maintenance with independent operators in machine readable format. Tombal (n 14) 373 and Martens *et al.* (n 71).

101 Tombal (n 14) 372.

102 ibid 375.

tition enforcement law.[103] Accordingly, '[…] the antitrust enforcement toolbox is inadequate to tackle effectively the need to ensure access to data. The scope of competition law is limited by the fact that it can be invoked only to gain access to a dataset held by a dominant firm, on a case-by-case basis'.[104] Instead, *ex ante* regulations can 'impose clear upfront before-the-event obligations that will be of more straightforward and speedy application compared to ex post enforcement, which assesses the conduct and its illegality once it has occurred'.[105] For instance, when it comes to compulsory data sharing, the DMA[106] provides rules on mandatory data access, according to which gatekeepers shall provide business users with access to the data generated by their activities and interactions on the platform.[107]

Ex ante regulations may impose obligations on specific and well-identified data holders[108] to make some of their datasets accessible to certain business users.[109] As it was the case for *ex post* data sharing remedies, *ex ante* regulations should only require to share volunteered and observed data, leaving inferred data outside their scope, as they are more valuable for the data holder.[110] Furthermore, contrary to *ex post* competition law remedies, *ex ante* regulations can impose obligations on data receivers who could be required to only use the data obtained for a specific legitimate

103 Tombal (n 14); Giuseppe Colangelo, 'The European Digital Markets Act and antitrust enforcement: a liaison dangereuse' (ICLE White Paper 2022); OECD, 'Ex ante regulation of digital markets' (2021) OECD Competition Committee Discussion Paper; Daniel Pettersson, 'Sector-Specific Ex Ante Regulation in Digital Markets – A Complement or Substitute to Antitrust Enforcement?' (2022) 4 Europarättslig tidskrift. For instance, Recital 5 DMA states that enforcement of articles 101 and 102 TFEU '[…] occurs ex post and requires an extensive investigation of often very complex facts on a case by case basis'.

104 Borgogno and Colangelo (n 57).

105 OECD (n 98) 5.

106 Regulation (EU) 2022/1925 of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

107 Article 6(10) DMA. A similar provision can be found for example in Section 3(a)(7) of the American Innovation and Choice Online Act <https://www.congress.gov/bill/117th-congress/senate-bill/2992/text> accessed 03 September 2024. See OECD (n 98) and OECD, 'G7 inventory of new rules for digital markets. Prepared by the OECD Competition Division for the 2023 G7 Joint Competition Policy Makers and Enforcers Summit' (2023).

108 Tombal (n 14) '[…] namely those that are considered as playing a central role in the (systemic) market failure(s) that led to the legislative intervention' 385.

109 Tombal (n 14).

110 ibid.

and economic purpose that has been declared in advance and to provide adequate guarantees for privacy and data protection, preventing reidentification in case the data shared have been previously pseudonymised or anonymised.[111]

As already seen with *ex post* B2B data sharing obligations, privacy and data protection rules may create some frictions with *ex ante* B2B data sharing regulations. Also, in this case, anonymisation and pseudonymisation should be encouraged as techniques to safeguard individuals' personal data, carefully considering costs and benefits for both data holders and data recipients.[112] Furthermore, in compliance with the GDPR, *ex ante* rules may provide a legal basis to impose data sharing obligations to the controller,[113] if such an obligation is 'clear and precise and its application [is] foreseeable to persons subject to it'.[114]

From a technical point of view, *ex ante* B2B data sharing rules can be successfully implemented only if specific tools are available that allow data transition between companies in a machine-readable format.[115] For this reason, Application Programming Interfaces (APIs)[116] with the adoption of common technical standards and protocols are essential to enable interoperability[117] between the systems and ensure an efficient implementation of data sharing rules.[118] However, imposing specific technical standards may add substantial costs for both data holders and data recipients.[119] Hence, instead of leaving data holders and data recipients to determine how data sharing should occur, an alternative solution that has been suggested is to

111 ibid.
112 ibid.
113 Article 6 and Recital 111 GDPR.
114 Recital 41 GDPR. See also Tombal (n 14).
115 Tombal (n 14); Borgogno and Colangelo (n 57) and Woźniak-Cichuta (n 6).
116 'APIs can be defined in broad terms as software tools designed to enable communication between two computer applications. Through a set of protocols and routines, they allow a digital application to interact with an associated program by describing the kind of data that can be retrieved, how to do it and the format in which information will be filed' Borgogno and Colangelo (n 57).
117 'Interoperability refers to the ability of products and services at different levels (vertical interoperability) or at the same level (horizontal interoperability) of the digital value chain to work together' OECD (n 6) 21. See also OECD (n 98); Crémer, de Montjoye and Schweitzer (n 84); Borgogno and Colangelo (n 57).
118 Tombal (n 14); Borgogno and Colangelo (n 57) according to whom 'APIs' architecture and design has been identified as a crucial element for a flourishing common European data space' 5.
119 Tombal (n 14).

establish intermediary bodies to act as 'data trustees' entrusted with the task of facilitating data exchange between the parties. In this model, the data holder would not share data directly with the recipients but only with the intermediary body.[120] Furthermore, algorithms could be trained on the intermediaries' dataset, which contains personal and non-personal data, but instead of having a human being accessing those data, only trained algorithms – and not the original data shared by the data holder – would be transferred to the data recipients.[121]

Overall, both *ex post* and *ex ante* B2B data sharing obligations can provide suitable solutions to remedy the problem of data as a competitive advantage in specific circumstances. By relying on the EFD, competition law provides an *ex post* remedy, by imposing an obligation on the data holder to share some of their data with the access seeker, only when exceptional circumstances are met, which is necessarily based on a 'case-by-case approach'.[122] On the contrary, regulations may impose compulsory B2B data sharing as an *ex ante* remedy, which can offer 'universal and general solutions, which may address whole industries, sectors, and market'[123] but at the same time be less tailor-made for specific needs.

4. Data as a competition enforcement tool: B2G and G2G data sharing

The last section of this chapter explores an alternative use of data as a tool to help authorities to enforce competition law. Besides being an extremely valued resource for many business models, data can be a game changer also for competition authorities. In fact, competition authorities require access to data in order to develop their own digital enforcement tools to enhance their detection powers and strengthen their proactive approach.[124] Any digital tool based on AI and machine learning systems that competition authorities would like to internally develop require qualitative and quantitative data (e.g. digital screening tools used in the initiation phase

120 Prüfer (n 95); Tombal (n 14). See also Olmedo-Peralta (n 45) concerning data pool contracts and the use of a third-party intermediary.
121 Prüfer (n 95) and Tombal (n 14).
122 Borgogno and Colangelo (n 57); Colangelo and Maggiolino (n 79) and Martens *et al.* (n 71).
123 Colangelo and Maggiolino (n 79).
124 OECD, 'Data Screening Tools in Competition Investigations' (2022) OECD Competition Policy Roundtable Background Note.

to flag markets' anomalies and to start *ex officio* investigations).[125] 'The challenges competition authorities face are the existence of data, to begin with, access to such data (and in particular disaggregated and raw data, and data that are not publicly available), format, integrity and quality of data, and data searchability, cleaning and use.'[126] Competition authorities need to access 'relevant, up-to-date and structured business data'[127] in order to develop their own digital enforcement tools. Hence, data represents the key element also for competition authorities to move towards 'computational antitrust'[128] and to embrace the digital antitrust revolution. Data is the necessary glue that tights together technological innovations used for enforcement purpose.

Firstly, competition authorities would require access to *qualitative data*, in order to avoid inaccurate and useless results.[129] Secondly, data need to be abundant (*quantitative data*) in order to create a large database on which algorithms can be trained. For instance, digital screening tools used to detect bid rigging cases in public procurement require a large dataset containing enough examples of past cases of collusive and non-collusive conducts.[130] Thirdly, data need to be already in *machine-readable format* to avoid time-consuming manual conversion process.[131]

How can competition authorities have access to such data – qualitative and quantitative data in machine readable format? Arguably, they can first rely on publicly available information that can be found in companies' registers, and chambers of commerce[132] (even though it is unlikely that all data will be in the right format, but further treatment may be required). Second, web scraping techniques could be used to extract relevant informa-

125 ibid; Herwig C H Hofmann and Isabella Lorenzoni, 'Future Challenges for Automation in Competition Law Enforcement' (2023) 3 Stanford Computational Antitrust 36.

126 OECD (n 124).

127 Viktoria H S E Robertson and Jürgen Fleiß, 'Computational Antitrust and the Future of Competition Law Enforcement' [2024] GRUR International XX(XX) 1.

128 Thibault Schrepel, 'Computational Antitrust: An Introduction and Research Agenda' (2021) 1 Stanford Computational Antitrust 1.

129 OECD (n 124).

130 ibid and Ioannis Lianos, 'Computational Competition Law and Economics: Issues, Prospects – An Inception Report' (2021) Hellenic Competition Commission

131 OECD (n 124).

132 ibid. See also OECD, 'Detecting cartels for ex officio investigations' (2024) OECD Roundtables on Competition Policy Papers, No. 311.

tion from websites.[133] Third, competition authorities, just like any other public and private actor, can purchase data from data brokers and third-party providers.[134] However, not all data that competition authorities may need are publicly available or can be purchased from third data providers, nor web scraping techniques may provide a long-term solution as it is considered a time-consuming mechanism.[135] Therefore, further solutions should be researched to enable competition authorities to access relevant data that can be used as an enforcement tool to feed their own digital systems. How can data access for competition authorities be improved? Two different channels are proposed further: imposing B2G data sharing and strengthening G2G cooperation and data sharing.

4.1. Imposing B2G data sharing

The first solution that can be proposed is to impose B2G data sharing which would enable competition authorities to get access to privately held data outside their investigation powers.[136] Imposing mandatory B2G data sharing could help competition authorities to overcome the reluctance of private companies to share their private information, without an official Request for Information, which can be issued only when an investigation has already begun.[137] 'Mandatory access rules'[138] need to be data protection

133 OECD (n 124); Lianos (n 130) and OECD (n 132).

134 OECD (n 124) and OECD (n 132). See also Ohlhausen (n 27).

135 OECD (n 124).

136 Heiko Richter, 'The law and policy of government access to private sector data ('B2G data sharing')' in German Federal Ministry of Justice and Consumer Protection, Max Planck Institute for Innovation and Competition (eds), *Data Access, Consumer Interests and Public Welfare* (Nomos 2021).

137 Article 18 Regulation (EC) No 1/2003 of 16 December 2002 on the implementation of the rules on competition laid down in articles 81 and 82 of the Treaty; Richter (n 136); Hofmann and Lorenzoni (n 125). In this regard, it is worth noting that according to the interpretation of the Court of Justice in C-690/20 P *Casino, Guichard-Perrachon SA and AMC v Commission* [2023] EU:C:2023:171, no distinction should be made on information collected before or after the opening of a formal investigation. The case concerned interviews conducted without the necessary safeguards of recording, which led to annul the Commission's decision. According to the Court of Justice '[…] there is nothing in the wording of article 19(1) of Regulation No 1/2003 or article 3 of Regulation No 773/2004 to suggest that the application of that recording obligation is contingent on whether the interview conducted by the Commission took place before the formal opening of an investigation in order to collect

and privacy-oriented in order to ensure that personal data are handled in compliance with data protection rules.[139] Furthermore, according to the principle of proportionality that public authorities are bound to respect, a balancing exercise between the need to access privately held data by public authorities and the consequent interference with the fundamental rights of privacy and freedom to conduct a business would have to be carried out.[140]

Rules that impose mandatory access to (specific types of) data can be found in other fields of law, for instance in the financial sector, to grant

---

indicia of an infringement, or afterwards, for the purpose of gathering evidence of an infringement. […] Those provisions do not in any way make the application of the obligation to record contingent on whether the information constituting its subject matter may be categorised as indicia or evidence.' [87–88]. Hence, when collecting information outside the Commission's investigatory powers in order to gather *indicia* of an infringement, adequate safeguards that respect fundamental rights of the parties concerned should be adopted.

138 Richter (n 137).

139 ibid. On this matter, it is important to mention that the Court of Justice has given relevance to the right to protect personal data and the right to respect private life, as enshrined in the Charter of Fundamental Rights vis-à-vis the investigatory powers of the Commission. For instance, in Order C-90/24 P(R) *Vivendi SE v Commission* [2024] EU:C:2024:318, the Court stated that the request for information sent to an undertaking was liable to breach the right of privacy of some of the company's employees and that the safeguards provided by the Commission, including the obligation of professional secrecy of Commission's officials, were not sufficient to prevent a breach of the right to respect for private life of the company's employees.

140 Herwig C H Hofmann, Dirk A Zetzsche and Felix Pflücke, 'The changing nature of 'Regulation by Information': Towards real-time regulation?' (2022) 28 European Law Journal 172. On this matter, see for instance C-470/21 *La Quadrature du Net and Others* [2024] EU:C:2024:370, where the Court of Justice stated that a fair balance must be struck between 'on the one hand, the legitimate interests relating to the needs of the investigation […] and, on the other hand, the fundamental rights to respect for private life and protection of personal data of the persons whose data are concerned by the access' 125]. Furthermore, 'as can be seen from article 52(1) of the Charter, that provision allows limitations to be placed on the exercise of those rights, provided that those limitations are provided for by law, that they respect the essence of those rights and that, in compliance with the principle of proportionality, they are necessary and genuinely meet objectives of general interest recognised by the European Union or the need to protect the rights and freedoms of others' 70. See further on the principle of proportionality C-511/18, C-512/18 and C-520/18 *La Quadrature du Net and Others* [2020] EU:C:2020:791 and C-817/19 *Ligue des droits humains* [2022] EU:C:2022:491. See also C-548/21 *Bezirkshauptmannschaft Landeck* [2024] EU:C:2024:830 where 'arising from article 52(1) of the Charter, […] any limitation on the exercise of a fundamental right must be 'provided for by law', that requirement implying that the legal basis authorising such a limitation must define its scope sufficiently clearly and precisely' 98.

access to real-time market data;[141] for research and statistical purposes;[142] and in the field of fuel retail market to oblige petrol stations to communicate changes in the fuel retail price.[143] If examples of this kind that impose B2G data sharing can be found in rather sectorial and sporadic legislations, the latest horizontal legislations are unlikely to be used as a legal basis to impose B2G data sharing in competition law. For instance, the Data Act[144] provides a general obligation to make data available to public sector bodies only on the basis of an 'exceptional need'. This is in case of a public emergency,[145] or non-emergency situations but only insofar as non-personal data are concerned, to carry out a specific task in the public interest that has been explicitly provided by law, and only if all other means to access such data have been exhausted.[146] Competition authorities are unlikely to satisfy such provisions of non-emergency situations, as article 16(2) Data Act specifies that B2G data sharing 'shall not apply to public sector bodies, the Commission, the European Central Bank or Union bodies carrying out activities for the prevention, investigation, detection or prosecution of criminal or administrative offences or the execution of criminal penalties, or to customs or taxation administration.' Data that competition authorities need for screening and market monitoring may fall within their detection activities and therefore outside the scope of this regulation.

Overall, B2G data sharing in competition law could be imposed with a sectorial regulation, along the same line of other sector-specific regulations, rather than relying on a general horizontal legislation. As in the case of B2B data sharing, sectorial regulations may be more tailor made to deal with specific needs and according to the principle of legal certainty, 'data

141 Hofmann, Zetzsche and Pflücke (n 140).

142 See for instance the UK Digital Economy Act <https://www.legislation.gov.uk/ukpga/2017/30/section/80> accessed 23 July 2024. See also Commission, 'Towards a European strategy on business-to-government data sharing for the public interest' (Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing 2020) 35 <https://op.europa.eu/en/publication-detail/-/publication/d96edc29-70fd-11eb-9ac9-01aa75ed71a1> accessed 27 August 2024 and Richter (n 136).

143 See for instance <https://www.bmwk.de/Redaktion/EN/Artikel/Energy/market-transparence-units.html> accessed 27 August 2024 and Richter (n 136).

144 Regulation (EU) 2023/2854 of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act).

145 Article 15(1)(a) Data Act.

146 Article 15(1)(b) and Recital 65 Data Act.

requests from public sector bodies [...] to data holders should be specific, transparent and proportionate in their scope of content and their granularity [and t]he purpose of the request and the intended use of the data requested should be specific and clearly explained [...]'.[147]

4.2. Enhancing G2G data sharing

The second mechanism for competition authorities to obtain relevant data in machine-readable format is to enhance cooperation and data sharing among public authorities (G2G data sharing).[148] Within the framework of the EU data strategy, which aims to create a 'Common European Data Space',[149] mechanisms for ensuring the free flow of data and cooperation among public authorities are laid down, where 'sharing should become a default'.[150] Firstly, cooperation among competition authorities should not be limited to share best practices and information on ongoing investigations, but *ex ante* data sharing should occur already at the detection phase. AI systems developed for screening purposes require a large dataset. 'Does such a data set exist today – [...] with a sufficient number of cases of both collusion and not-collusion, with the necessary data on price, cost, and drivers of supply and demand [...]?'[151] As it might be almost impossible to manually create such a database, data on past cartel cases should be shared among competition authorities to create a large data pool that can be accessed whenever algorithms need to be trained.[152] Being able to dispose of such dataset would have the advantage of providing competition authorities with enough data (*quantitative data*) of relevant information of past cases (*qualitative data*) in *machine-readable format* that can be

147 Recital 69 Data Act.

148 See OECD (n 132) according to which 'co-operation between competition authorities and other domestic entities (such as government bodies, procurement agencies and sector regulators) may facilitate the acquisition of data and is therefore crucial for achieving effective screening tools' 10.

149 <https://digital-strategy.ec.europa.eu/en/policies/data-spaces> accessed 24 July 2024.

150 Regulation (EU) 2024/903 of 13 March 2024 laying down measures for a high level of public sector interoperability across the Union (Interoperable Europe Act) Recitals 25 and 26.

151 Rosa M Abrantes-Metz and Albert D Metz, 'Can Machine Learning aid in Cartel Detection?' (2018) CPI Antitrust Chronicle 1.

152 Lianos (n 130) 16 and OECD (n 124).

used to feed their algorithms, enhancing the performance of their AI tools, in terms of accuracy (with less false positives and false negatives) and reliability of the results. It could even be suggested that if digital screening tools are trained on a dataset which is non-biased, accurate and sufficiently large, their outcomes could be used as direct evidence to issue inspection decisions and prove an anticompetitive behaviours, under supervision of a human being.[153] This approach would significantly increase competition authorities' efficiency and detection capabilities.

As part of G2G data sharing, cooperation between competition authorities and other public enforcers should be encouraged.[154] For example, cooperation with public procurement bodies is often needed for competition authorities to have access to data concerning tenders. Competition authorities often investigate bid rigging cases, and in order to use accurate digital tools embedded with machine learning and AI techniques, they need to have access to structured bidding data that usually only public procurement bodies possess.

Enhancing data sharing between competition authorities among each other and between them and public procurement bodies requires 'good working relationships' and 'a high degree of trust between authorities'.[155] Latest EU digital legislations also pinpoint the need to create a trust environment where data can freely move within the single market as a fifth freedom, where rights and interests of data holders and data subjects are respected.[156] In particular, when personal data are the subject of G2G data sharing, compliance with data protection law can be ensured through anonymisation and pseudonymisation techniques, as suggested for B2B data sharing. Finally, also in the context of G2G data sharing, a data trustee can be a solution as it may be regarded as a trustworthy intermediary body which supervises that data sharing among public authorities occur in respect of all necessary safeguards for data subjects and data holders.

153 OECD (n 124).

154 OECD (n 132) '[…] co-operation between competition authorities and other domestic entities (such as government bodies, procurement agencies and sector regulators) may facilitate the acquisition of data and is therefore crucial for achieving effective screening tools' 10.

155 OECD (n 124) 26.

156 Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European strategy for data' COM(2020) 66 final; Borgogno and Colangelo (n 57).

5. Conclusions

This chapter has analysed how data can impact competition law from both sides: market players and competition authorities. On the one hand, modern business models revolve around the collection and processing of personal and non-personal data to generate income and deliver new products, as innovative technologies depend on data. In a digital and data-driven economy, the risk is that few large companies can use data advantage to strengthen and abuse their dominant position at the expense of smaller competitors. Therefore, solutions to impose data sharing among businesses (B2B data sharing) have been laid down. First, competition tools can play a role to impose *ex post* data access according to the EFD, which has been criticised when applied to data, and some adaptations may be required. Second, *ex ante* rules that impose B2B data sharing can be implemented in the form of sectorial and horizontal legislations. They provide compulsory B2B data sharing rules to specific data holders to remedy market failures, before affecting competition, ensuring a level playing field in data-centric markets.

On the other hand, besides focusing on how data has impacted competition law from the point of view of market players, this chapter has showed how data has become an essential factor for competition authorities when enforcing competition law. More and more digital tools have started to be developed by enforcers for which data availability is of utmost importance. Hence, data sharing between competition authorities and private businesses (B2G data sharing) and between them and other public authorities (G2G data sharing) may be a long-term solution to ensure enforcers access to qualitative and quantitative data in machine-readable format, also in the light of the latest EU digital legislation. In a data-driven economy, data is a key element for both private and public actors, and B2G and G2G data sharing for the purpose of enforcing competition law 'should become a default'.[157]

157 Recitals 25 and 26 Interoperable Europe Act.

Artificial Intelligence in the Service of Self-represented Litigants

Accessing the Law with the Aid of Large-Language-Models

Kalliopi Terzidou

Abstract: Lawyers play a crucial role in representing litigants in court and providing legal counsel on their rights, obligations, and the procedural steps involved in complex pre-trial and trial proceedings. However, the excessive costs of legal services and the absence of (sufficient) legal aid mechanisms can exclude certain citizens from seeking a dispute settlement before courts. Chatbots based on Large Language Models (LLMs) have become popular due to their user-friendly way of communicating information on substantive and procedural law and on court proceedings. Nevertheless, LLM-based chatbots can mislead users by providing inaccurate information, ultimately affecting the progress of their case before courts. The present paper asks to what extent Large Language Models-based chatbots are designed in a human centric way to support self-represented litigants in accessing legal information. To address this question, the paper explores the concept of 'human centricity' under the EU's framework for trustworthy Artificial Intelligence, in order to develop a set of functional requirements against which the performance of LLM-based chatbots is assessed in the context of landlord-tenant disputes.

1. Introduction

Since late 2022, ChatGPT[1] and other chatbots have emerged as virtual assistants, providing users with answers to a wide range of questions. These chatbots are based on Large Language Models (LLMs), i.e. Artificial Intelligence (AI) models trained on a large amount of data through self-supervision techniques, so they can perform generally applicable functions in

1 OpenAI, 'ChatGPT' <https://chatgpt.com> accessed 07 July 2025.

different contexts.[2] Popular[3] online LLMs online are OpenAI's ChatGPT, Microsoft's Copilot,[4] and Google's Gemini.[5]

Legal aid applications based on AI systems exist in the market for several years, offering different kinds of legal support to users. Websites such as DoNotPay[6] and JusticeBot[7] offer legal 'advice' and resources to users experiencing everyday legal problems, such as housing and landlord-tenant issues. Other websites, such as A2J Author[8] and Court Forms Online,[9] assist users with the completion and filing of court forms. These types of AI systems are usually rule-based, i.e. trained on a limited dataset to perform a specifically designated set of tasks, in contrast to LLM-based chatbots (such as the ones mentioned above) that are General-Purpose AI systems, i.e. AI systems able to address a variety of purposes by generating multiple forms of content.[10] Chatbots employ a conversational tone to help users complete tasks, including legal research, summarizing and translating of legal resources, and drafting legal documents.[11] Users can personalize their experience by refining their prompts or even customizing the available settings. Unlike lawyers, these applications are available on a 24/7 basis from a remote location, requiring only a digital device connected to the Internet. In this way, legal information becomes more accessible to people who would not otherwise have the financial resources to consult a lawyer. Ultimately, users can reduce their exposure to bureaucratic delays while accessing court services, such as filing a court form, and the stress incurred when visiting the lawyer's office or the court.

2 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L/1689 (AI Act) recital 99 and article 3(66).

3 See, for example, Krystal Hu, 'ChatGPT Sets Record for Fastest-Growing User Base – Analyst Note' *Reuters*, (2023) <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/> accessed 03 October 2024.

4 Microsoft, 'Copilot' <https://copilot.cloud.microsoft> accessed 07 July 2025.

5 Google, 'Gemini' <https://gemini.google.com/app> accessed 07 July 2025.

6 DoNotPay, <https://donotpay.com/> accessed 03 October 2024.

7 JusticeBot, <https://justicebot.ca/question/1/> accessed 03 October 2024.

8 A2J Author, <https://www.a2jauthor.org/> accessed 03 October 2024.

9 Court Forms Online, <https://courtformsonline.org/> accessed 03 October 2024.

10 AI Act, recital 99 and article 3(66).

11 Kalliopi Terzidou, 'Generative AI Systems in Legal Practice Offering Quality Legal Services While Upholding Legal Ethics' (2025) International Journal of Law in Context 1, 12 <https://doi.org/10.1017/S1744552325000047> accessed 07 July 2025.

At the same time, most legal aid applications of any type entail certain implications. First, they are jurisdictionally limited, i.e. providing legal information and resources specific to a country that cannot be extended to states with different legal systems. Second, they are available exclusively or primarily in the English language, thus excluding a large part of Internet users from accessing them. Third, they are typically designed for simple and straightforward legal matters, thus not being able to handle nuanced or multi-jurisdictional legal issues. For example, when prompted to draft a legal document, such as a court form, the application may return plain templates, so that users must customize them according to the applicable legal framework and the facts of the case. Fourth, LLMs, in particular, can hallucinate, i.e. generate content that is fictitious and results in unreliable outputs.[12] Fifth, the applications' providers are not bound by the same rules of professional conduct as lawyers are during the consultation and representation of their clients, such as the principle of confidentiality and professional secrecy,[13] meaning that they cannot be accountable to the public for erroneous information that their LLM-based solutions return to them. Finally, there are concerns over the protection of users' (personal) data when using these AI applications, as well as over the exclusion of certain members of society without digital literacy or access to a digital device and/or a stable Internet connection from taking advantage of the benefits of these AI systems.

The aim of this paper is to, first, define 'human-centric' functional requirements for the design of legal aid applications based on AI systems, so that these provide equal and safe access to justice to self-represented litigants (SRLs) by retrieving understandable and reliable information on substantive and procedural law and on court proceedings. Second, the paper aims at discovering whether the design of LLM-based chatbots corresponds to these requirements, in order to understand if they can indeed help SRLs to access legal information in a human-centric way. Therefore, the research question asks:

To what extent are Large Language Models-based chatbots designed in a human centric way to support self-represented litigants in accessing legal information?

12 See, for example, Zhiqing Ji and others, 'Survey of Hallucination in Natural Language Generation' *arXiv*, (2022) 2 <https://arxiv.org/abs/2202.03629> accessed 7 July 2025.

13 Providers are mainly bound by product safety rules, such as those laid down in the AI Act concerning the placement in the market of (high-risk) AI systems.

To address this question, 'human-centricity' is defined based on the EU's notion of 'trustworthy AI' as developed in two main documents: the Ethics Guidelines on Trustworthy AI[14] and the AI Act.[15] In particular, the principle of human-centricity denotes the respect of human dignity and of fundamental rights, such as the right to privacy, which are inherently connected to the (ethical) principles of respect for human autonomy, prevention of harm, fairness, and explicability.[16] Based on this definition, a set of functional requirements for legal aid applications is outlined, against which seven LLM-based[17] chatbots are evaluated, in particular ChatGPT, ChatGPT Free Legal Advice, ChatGPT Pro, Copilot, Copilot for students or enterprises, Gemini, and Gemini Advanced. These chatbots are the free and the subscription-based versions of ChatGPT, Copilot, and Gemini,[18] which are popular chatbots and, thus, most likely to be accessed by the average Internet user. This diversity of chatbots allows for more extensive and valuable comparisons based on the human-centric functional requirements.

'Self-represented litigants' (SRLs) are individuals who are not represented by a legal professional during court proceedings. The choice to represent oneself before the courts may be attributed to several factors, including unaffordable legal services and insufficient legal aid means provided by the State. A fragmented situation is observed in the EU, where only nine Member States offer full coverage of litigation costs along with partial legal aid schemes, while ten offer either full or partial legal aid, and four leave the matter on the discretion of courts.[19] Furthermore, in six Member States individuals are not automatically exempted from paying court fees, while in three Member States people with income below the Eurostat poverty

14 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (2019) 37 <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1> accessed 07 March 2025.

15 AI Act.

16 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (n 14), 11–20.

17 For the definition of a Large Language Model, see AI Act, recital 99 and article 3(66).

18 Even though most users are likely to visit the free version of these chatbots, the subscription-based versions are also accessible for a considerably small monthly or annual fee, arguably offering a more advanced user experience.

19 European Commission, '2025 EU Justice Scoreboard' COM(2025) 375 final (1 July 2025) 20–21 <https://commission.europa.eu/document/download/51b21eff-a4b0-4e73-b461-06bd23b43d4e_en?filename=2025%20EU%20Justice%20Scoreboard_template.pdf> accessed 7 July 2025.

threshold are unable to receive legal aid.[20] The lack of guarantees in favour of legal aid across the EU underlines the gap in access to justice and the attraction of individuals to self-representation. It must be noted that self-representation is not allowed in all types of court proceedings. This practice can generally be observed in small claims cases, as can be seen in the examined case of landlord-tenant disputes in Greek courts, whereby a landlord can self-represent himself or herself before Small Claims Courts (Ειρηνοδικεία) when the claim is under EUR 600.[21]

One way to support SRLs in their legal journey is to assist them in navigating the complex regulatory landscape governing their legal issue. These rules are expressed in lengthy and ambiguous documents[22] which are often charged with legalese.[23] For the needs of this research, 'legal information' refers to publicly available resources such as legislation, case law, and information on court proceedings, which are published by state and judicial institutions in analogue and/or digital format. All Member States provide access to online information about their judicial system for the public, including information on legal aid, court fees, procedural rights, and court forms.[24] Several Member States (Germany, Estonia, Latvia, Greece, Croatia, Austria, Portugal) reportedly integrate chatbots in their websites to help the public retrieve information on their justice system.[25] The increasing adoption of chatbots by national justice systems for the retrieval of legal information indicates an effort by EU Member States towards enhancing access to justice for their citizens. Legal information is democratized, in contrast to the longstanding 'monopoly' that legal experts have exercised over legal knowledge. LLM-based chatbots can offer legal information in plain and concise language, at little or no cost, and on a 24/7 basis, offering

20 ibid 21 and 106.

21 Greek Parliament, 'Code of Civil Procedure (Κώδικας Πολιτικής Δικονομίας)' (n 28), articles 14 (1)(b) and 94 (2)(a).

22 Max Barrett, *The Art and Craft of Judgment Writing: A Primer for Common Law Judges* (Globe Law and Business 2022) 39–46 <https://www.globelawandbusiness.com/books/the-art-and-craft-of-judgment-writing-a-primer-for-common-law-judges-second-edition> accessed 07 July 2025.

23 Z Ződi, 'The limits of plain legal language: understanding the comprehensible style in law' *(2019) 15(3) International Journal of Law in Context* 246, 257–258 <https://www.cambridge.org/core/journals/international-journal-of-law-in-context/article/abs/limits-of-plain-legal-language-understanding-the-comprehensible-style-in-law/B07D80A34052ED1E6EF7F7C289EC97DD> accessed 07 July 2025.

24 European Commission, '2025 EU Justice Scoreboard' (n 19), 32–33.

25 European Commission, '2025 EU Justice Scoreboard' (n 19), 33.

an unyielding support to SRLs (notwithstanding phenomena such as hallucinations indicative of LLMs).

For the evaluation of these systems, the researcher first assumed the persona of a landlord whose tenant does not pay the rent on time. The landlord does not know if this qualifies as a legal issue or how it is legally treated under Greek law. Greece has been the targeted jurisdiction due to the familiarity of the researcher with Greek law, having obtained an LL.B in Greece, so as to be able to reliable evaluate the chatbots' responses. Another reason is the particularity of Greek online resources. Because the majority of online content is in the English language, LLMs trained on Internet-sourced data perform typically better on English-language prompts. Consequently, the LLM-based chatbots are tested in the Greek language to confirm whether LLM- based chatbots underperform on Greek prompts due to the limited availability of Greek online resources.

A total of 16 questions (Annex I) were posed to the LLM-based chatbots, encompassing matters of landlord-tenant relations that are also regulated by national law, primarily by the Greek Civil Code[26] and Code of Civil Procedure.[27] These regulations treat three main regulatory areas: the legal rights of the landlord, the competent court before which these rights can be defended, and the applicable procedure. It must be noted that under Greek law, landlords can choose between two legal pathways when their tenants do not pay rent. The landlord can either apply for an *action* ('αγωγή')[28] or an *order* ('διαταγή')[29] for the restitution of rent.[30] A

26 Greek Parliament, 'Civil Code (Αστικός Κώδικας)' ΦΕΚ: A 164 19841024, (1946) <https://www.ministryofjustice.gr/wp-content/uploads/2019/10/Αστικός-Κώδικας.pdf> accessed 11 March 2025.

27 Greek Parliament, 'Introductory Law of the Code of Civil Procedure (Εισαγωγικός Νόμος Κώδικα Πολιτικής Δικονομίας)' ΦΕΚ: A 182 19851024, (1968) <https://www.lawspot.gr/nomikes-plirofories/nomothesia/eisnkpold/arthro-66-eisagogikos-nomos-kodika-politikis-dikonomias> accessed 11 March 2025.

28 Greek Parliament, 'Civil Code (Αστικός Κώδικας)' (n 25), article 597; Greek Parliament, 'Introductory Law of the Code of Civil Procedure (Εισαγωγικός Νόμος Κώδικα Πολιτικής Δικονομίας)' (n 26), article 66.

29 Greek Parliament, *Code of Civil Procedure (Κώδικας Πολιτικής Δικονομίας)*, Presidential Decree 503/1985, FEK A 182/24.10.1985, articles 623–636 <https://www.lawspot.gr/nomikes-plirofories/nomothesia/kodikas-politikis-dikonomiasaccessed> 11 March 2025.

30 The main difference between the action and the order is that upon bringing an action before the court, the lease is terminated, while the order does not result in the termination of the contract. Moreover, the decision on the action must become final before it can be enforced against the tenant, while the order is an enforceable title

landlord can self-represent himself or herself only when the legal dispute concerns a small claim (under EUR 600) to be judged before Small Claims Courts (Ειρηνοδικεία).[31]The answers by the LLM-based chatbots to these questions were assessed on a scale from 1 (low) to 5 (high) based on the designated human-centric functional requirements. The points were assigned by the researcher herself, based on her user experience and the verification of the retrieved information against Greek legislation.[32]

The hypothesis is that users are empowered to take legal action and defend their legal position before courts by accessing quality legal information on substantive and procedural law, that is readable and accessible to all so that they can rely upon it. In any case, users must contextualize and verify any information retrieved by LLM-based chatbots according to the facts of their cases, since the responses to their prompts may be generic and thus not applicable to their particular case or generally inaccurate, respectively.

The paper is divided into three main sections. The first section explores 'human-centricity' under the EU's framework for trustworthy AI to ultimately define a set of functional requirements for the design of legal aid applications based on AI systems. The second section evaluates the selected LLM-based chatbots against these functional requirements to understand if they are designed in a human-centric way to support SRLs involved in landlord-tenant disputes. The third section critically examines the conditions for integrating LLM-based chatbots into national court websites for the provision of court services.

2. A Human-Centric Design

According to a 2021 HiiL and IAALS survey, people interested in resolving everyday issues of a family, land, or employment nature are likely to seek

---

and can be executed upon the expiry of a 20-day period. The action is also a more expensive and time-consuming procedure than the order; e-Nomika, 'Διαδικασία Έξωσης ενοικιαστή: Χρόνος και κόστος' <https://e-nomika.gr/διαδικασία-έξωσης-ενοικιαστή-χρόνος/#:~:text=Με%20την%20άσκηση%20της%20αγωγής,αίτηση%20έκδοσης%20της%20διαταγής%20απόδοσης> accessed 11 March 2025.

31 Greek Parliament, 'Code of Civil Procedure (Κώδικας Πολιτικής Δικονομίας)' (n 28), articles 14 (1)(b) and 94 (2)(a).

32 Future research could replicate the present methodology by involving participants in an experiment setting, where the chatbots' performances could be tested against an even wider set of human-centric functional requirements potentially defined by participants.

legal advice on the Internet, at least before resorting to a legal expert.[33] This was often attributed to people not considering the help of a lawyer to be necessary or being unable themselves to identify their issue as a legal one.[34] When visiting the Internet, most people admitted that the retrieved information affected them considerably by helping them understand how to solve their issue,[35] notwithstanding the importance of trustworthiness, clarity, and context-specificity in rendering the information helpful and usable for users.

The widespread use of the Internet for legal information highlights the need for web services to be designed with a human-centric approach. The principle of 'human-centricity' is fundamental in EU's approach towards trustworthy AI. According to the High-Level Expert Group on Artificial Intelligence, a human-centric AI approach aims to '…ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights…all of which are united by reference to a common foundation rooted in respect for human dignity…'[36]

For a human-centric approach to be respected in the designing, development, deployment, and use of AI systems, the High-Level Expert Group on Artificial Intelligence claims that four ethical principles should be respected: fairness, respect for human autonomy, explicability, and prevention of harm.[37] According to the *principle of fairness*, all individuals and groups must receive an equal and just treatment when interacting with AI systems, which entails two dimensions: a substantive one (avoid unfair biases that can lead to their discrimination) and a procedural one (opportunity to contest through an effective redress the outcomes of AI systems).[38] All AI systems, including LLM-based chatbots designed to retrieve legal information, should provide equal and just access to available resources, taking

33 The Hague Institute for Innovation of Law (HiiL) and the Institute for the Advancement of the American Legal System (IAALS), 'Justice Needs and Satisfaction in the United States of America 2021: Legal problems in daily life' (2021) 160 <https://iaals.du.edu/sites/default/files/documents/publications/justice-needs-and-satisfaction-us.pdf> accessed 03 October 2024.

34 ibid 164.

35 ibid 166.

36 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (n 14) 37.

37 ibid 9–13.

38 ibid 12–13.

into consideration the users' level of legal and digital literacy to adapt the accessibility and understandability of the generated information to their needs.[39] Legal literacy is important for the empowerment of SRLs, since it provides people with awareness and understanding of information on key legal principles and rules pertaining to their rights and duties, enabling them to take action when faced with a legal issue.[40] Digital literacy is also important since it allows SRLs to remain in control of the AI system by mastering its different features to arrive to the desired result.[41] AI systems, including LLM-based chatbots, might present challenges related to legal and digital literacy, potentially confusing users by generating inaccurate, incomplete, or otherwise unreliable legal information (legal literacy) or when they present this information through a complex interface with limited or unclear ways of interacting with the available features (digital literacy). SRLs as users may be more vulnerable to harm from LLM-based chatbots, since they are more likely to lack the necessary legal knowledge to assess the reliability of the generated legal information and may further not have access to legal consultation due to the high legal fees. This result may be exacerbated if they are not digitally literate and thus in a position to distinguish between the advice of a legal expert and that generated by an

39 People should take part in the designing process of legal aid applications to define users' expectations and needs regarding legal and court services and give feedback on the applications' features from an end-user's perspective. Participation can take the form of surveys or workshops, among other types of interventions; OECD, 'OECD Framework and Good Practice Principles for People-Centred Justice' (2021), 14–15 and 66–72 <https://www.oecd.org/en/publications/oecd-framework-and-good-practice-principles-for-people-centred-justice_cdc3bde7-en.html> accessed 03 October 2024.

40 For a definition of 'legal literacy' or 'legal education', see The Law Society, 'Public legal education' <https://www.lawsociety.org.uk/campaigns/public-legal-education> accessed 07 July 2025; Such interventions should be inclusive by providing equal opportunities for legal education, considering the different ways that people view, consider, and digest legal information, in text or other media, such as videos. In any case, legal information should be available in a plain and clear language, devoid of heavy legal jargon.

41 For a definition of 'digital literacy', see UNESCO Institute for Statistics, 'SDG 4 Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all' 1 <https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2021/08/Metadata-4.4.2.pdf> accessed 07 July 2025; Deployers of such applications, including courts, must also offer an intuitive browsing experience through simple interfaces and customizable features, as well as alternatives for citizens that do not have access to a digital device or stable Internet connection, equipping public spaces (e.g. public libraries) with freely accessible digital equipment and Wi-Fi connection.

LLM-based chatbot. in the latter case, users may excessively rely on the system's outputs due to the persuasive way that the generated information is communicated (automation bias).[42]

According to the *principle of respect for human autonomy*, as determined by the High-Level Expert Group on Artificial Intelligence,[43] individuals must be able to retain full and effective self-determination, without any unjustified interference, when interacting with AI systems, in order to be empowered to make meaningful choices and participate in democratic society. When interacting with LLM-based chatbots, users may perceive the systems' responses as expert advice and accept them uncritically due to the highly persuasive conversational tone of the systems, inadvertently relinquishing some of their agency and taking less informed decisions. This is why AI systems must be transparent so their purposes, capabilities, and decisions can be communicated in an explainable manner to those directly or indirectly affected, following the *principle of explicability*.[44] This is not always perceivable in General-Purpose AI systems, including LLM-based chatbots such as ChatGPT, the functioning of which remains a black-box in an effort to protect them as proprietary models, hindering the understanding of the mechanism behind the generation of their outputs that is critical to maintain human agency and oversight over the systems.

According to the *principle of prevention of harm*,[45] AI systems should not result in physical and mental harm, but should ensure the dignity and safety of users, especially vulnerable individuals, when faced with malicious use and power asymmetries. With respect to LLM-based systems, prevention of harm is difficult to achieve given the considerable amount of training data that these systems need, which in this case can reach trillions of data,[46] that might include sensitive data points. In particular, these training datasets can

42 Hannah Ruschemeier, Lukas J Hondrich, 'Automation bias in public administration – an interdisciplinary perspective from law and psychology' (2024) 41(3) Government Information Quarterly 101953 1 <https://doi.org/10.1016/j.giq.2024.101953> accessed 7 July 2025.

43 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (n 14), 12.

44 ibid 13.

45 ibid 12.

46 Nevertheless, the limited financial resources that a provider, such as national (judicial) institutions, might have available to develop LLM-based chatbots from scratch or fine-tune such systems can lead to the training being based on a limited amount of data and result in applications that do not perform well their designated task; Ashwin Telang, 'The Promise and Peril of AI Legal Services to Equalize Justice' *(Harvard*

well include personal data provided by the users themselves through their prompts or cookie preferences or automatically collected by the AI systems (for example as metadata). The processing of personal data for training purposes can lead the system to infer sensitive human characteristics, such as age and ethnicity, and ultimately discriminate against vulnerable groups, including women and ethnic minorities, resulting to their exclusion from legal or court services in violation of the principle of fairness. If the AI system has suffered from technical errors or produces harmful outputs, including outputs that lead to adverse discrimination or breaches of personal data protection, then mitigation measures must be implemented to prevent further harm, following the results of applicable impact assessments.

The AI Act likewise recognizes the importance of human-centricity, stating in article 1 (1) that 'The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI)…'[47] In this way, the Act operationalizes the ethical principles of human-centric AI, as previously stated by the High-Level Expert Group, and further sets down binding legal rules for providers, deployers, and other actors involved in the ecosystem of AI systems. These rules follow a risk-based approach, according to which certain AI practices are prohibited, such as the use of AI systems that employ subliminal, manipulative, or deceptive techniques to distort behaviour and impair informed decision-making (article 5, para. 1, point a). The production and use of systems that are described as 'high-risk,' i.e. systems that pose a 'significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making' (article 6, para. 3), is in principle allowed, following strict product safety requirements that aim at a more transparent and safe development and deployment of an AI system (articles 8–15).

For 'high-risk' AI systems, the principle of fairness is addressed through the requirement of data governance (article 10), to ensure data quality and to prevent discriminatory outcomes; the respect for human autonomy is expressed in the requirement for human oversight over the AI system (article 14), to guarantee meaningful outcomes and informed decisions by users; the principle of explicability is embedded in documentation (article 11) and transparency obligations (article 13), to record the functioning of the system

Journal of Law & Technology Digest, 2023) <https://jolt.law.harvard.edu/digest/the-promise-and-peril-of-ai-legal-services-to-equalize-justice> accessed 03 October 2024.

47 AI Act.

and allow for the explainability and contestability of its outputs, and; the principle of prevention of harm is operationalized through provisions such as risk management (article 9), to allow the assessment of risks and their effective mitigation.

3. Human-Centric Design of AI-Based Legal Aid Applications

A 'human-centric' approach to designing AI systems, as outlined in Section 2, provides a useful ethical and human rights-based framework under which the design of LLM-based chatbots can be evaluated. More specifically, the ethical principles of fairness, respect for human autonomy, explicability, and prevention of harm, can provide this assessment framework,[48] which can become binding when these principles are translated into requirements under the AI Act.

The principle of fairness demands that AI systems are designed to equally and justly distribute their benefits, while avoiding the production of unfair biases that can result in discrimination of any kind (substantive dimension).[49] In the context of the present research, AI systems must be accessible regarding their features, so that all users are guaranteed equal access to the benefits of LLM-based chatbots, particularly digitally illiterate users. Accessibility standards ensure inclusivity, especially for people with disabilities, enabling equitable access and participation in digital assistive technologies. Accessibility is greatly associated with the user interface, i.e. how a user perceives and interacts with the features of LLM-based conversational interfaces. The interface must be *user-friendly*, meaning that it is interactive and responsive to users' instructions, allowing a personalized and intuitive experience through customizable settings.

The outputs of AI systems (as well as their disclaimers and terms of use) must be *readable*, i.e. easy to understand,[50] in order to reduce information asymmetries, particularly regarding legal information which may not be readily understood by the wider public, and thereby enhance access to the

48 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (n 14) 15–19.

49 ibid 12–13.

50 See, for example, Margaret Hagan, 'Good AI Legal Help, Bad AI Legal Help: Establishing quality standards for responses to people's legal problem stories' in *JURIX 2023: 36th International Conference on Legal Knowledge and Information Systems, AI and Access to Justice Workshop, Maastricht*, (2023) 4 <https://ssrn.com/abstract=4696936> accessed 03 October 2024.

justice system. AI-based legal aid applications should minimize legal jargon and offer alternative information formats, such as visualizations, to ensure accessibility for users with disabilities and minors who may benefit from audiovisual stimuli for better comprehension. The application must also be available in multiple languages, including the official language(s) of the jurisdiction where the application is offered, to ensure its unhindered use by all users.

According to the principle of respect for human autonomy, users of LLM-based chatbots should be able to make informed and autonomous decisions based on the systems' outputs, without the risk of misinformation or deception.[51] In the context of the present research, AI systems must protect the autonomy of individuals by ensuring access to *quality information* that is accompanied by references to accompanying trustworthy sources, so that users avoid being misinformed and can safely rely on the outputs of AI systems to take legal action. The quality of the retrieved legal information amounts to the accuracy, completeness (sufficient details), consistency,[52] and (situational) relevance (alignment with the user's information needs) of the systems' responses.[53] Since the information retrieved by legal aid applications is of a legal nature, the information further needs to be jurisdiction-specific, so it applies to a specific country, and updated, so it captures the latest amendments of the legislation or the most recent court decisions. Quality information as a functional requirement enables the reliance of users on the system's outputs by minimizing the risk of misinformation and subsequent undesirable effects, such as mistakes in the defence strategy due to ill-informed decisions.[54]

---

51 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (n 14) 12.

52 See, for example, Christian Matt and others, 'Towards a multicentric quality framework for legal information portals: An application to the DACH region' (2023) 40(4) Government Information Quarterly 1, 4–5 <https://www.sciencedirect.com/science/article/pii/S0740624X23000400> accessed 16 March 2025.

53 T Saracevic, 'Relevance Reconsidered' in *Information Science: Integration in Perspectives, Proceedings of the Second Conference on Conceptions of Library and Information Science, Copenhagen (Denmark)* (1996) 12 <https://www.bibsonomy.org/bibtex/2c3f30a1af84eb099284f266679ff41f2/aschmidt> accessed 16 March 2025.

54 See, for example, Sara Merken, 'New York Lawyers Sanctioned for Using Fake ChatGPT Cases in Legal Brief' *Reuters* (2023) <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/> accessed 03 October 2024.

Reliance on the retrieved information is further reinforced through *citations*, i.e. references to the underlying sources, so citizens can verify the accuracy of the information and make informed choices. In the case of legal aid applications, citations must mainly consist of legislation and case law but can also refer to commentaries of the law (ideally laid down in plain language), so citizens can interpret the law and apply it to their case. Links to the online versions of the cited sources should further be provided for easier navigation.

If the retrieved information is of good quality and accompanied by references, then users can rely on it to take the appropriate legal action for the resolution of their legal issue(s). This is an important feature of a human-centric design because it aims to empower SRLs to seek a solution to their legal issue by not only collecting the necessary legal information and resources but ultimately applying them in practice when initiating court proceedings.[55] *Actionability* refers to the ability of the LLM-based chatbot to return information that not only enables users to identify if their issue is justiciable, but also to initiate court proceedings if necessary. For example, the retrieval of court forms' templates or the generation of pre-filled versions according to the individual situation could enable users to start the application process for court proceedings of their own accord.

According to the principle of prevention of harm, AI systems must be designed, developed, and deployed in a way that protects users from any kind of harm, physical or otherwise.[56] Fewer to no harms mean more equal access to the benefits of AI systems, in accordance with the principle of fairness, given that some types of harms may affect digitally illiterate people to a greater extent. One such harm might be breaches of personal data, which users may voluntarily provide to the system without knowledge of the consequences on their rights to privacy and data protection. Breaches of personal data can occur due to their unlawful processing by the LLM or due to the compromise of the model's security, for example through unauthorized access. Another risk may be the production of unfair biases due to the system's training on historical data or due to poor governance models, leading to the discrimination of societal groups, some of which may be vulnerable. Some of these harms may be only addressed by developers, as is the case with the elimination of unfair biases, through the controlled

55 Hagan (n 49) 10.

56 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (n 14) 12.

training and constant monitoring of the system. In any case, LLM-based chatbots must inform users through *disclaimers* on the risk of exposure to these harms and offer them the option to consent to the processing of their data under strictly defined principles, such as purpose and storage limitation.[57] The importance of information provided to users on these types of harm is also highlighted under the principle of explicability.

According to the principle of explicability, individuals must have access to transparent AI systems, whose capabilities and limitations are 'openly communicated' to the public.[58] The transparency of AI systems enhances trust in their functioning, so that users maintain their autonomy by avoiding being misinformed or otherwise manipulated by the system and being able to make informed choices, including challenging the systems' outputs if they are inaccurate or otherwise problematic. The display of disclaimers on the functioning of the AI system ensures that users have the chance to become aware of the capabilities and limitations of the system and their rights in case of erroneous outputs that have consequences on their livelihood. In the case of LLM-based chatbots, the system must be transparent by displaying disclaimers on the machine-generated nature of the retrieved information, which should be independently verified rather than accepted at face value. Disclaimers can also alert users as to the processing of their (personal) data or as to the probability of inaccurate or irrelevant outputs that might misinform users. In the case of SRLs, disclaimers may encourage users to consult a lawyer to verify the accuracy and relevance of the retrieved information for the case and thus avoid any adverse consequences when following unrelaible generated information.

4. Chatbots' Performance on Landlord-Tenant Disputes-Related Questions

LLM-based chatbots are widely used for the generation of responses to users' prompts, including questions on their legal situation, rights, obligations, and relevant procedures when they are facing a legal and possibly justiciable problem. However, LLMs may produce hallucinations when generating answers that lead to users' misinformation and, subsequently, to

57 See, for example, European Parliament and the Council of the EU, 'Regulation (EU) 2016/679 (General Data Protection Regulation)' (2016) OJ L 119 article 5 <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng> accessed 16 March 2025.

58 High-Level Expert Group on AI (European Commission), 'Ethics Guidelines for Trustworthy AI' (n 14) 13.

ill-informed choices affecting the self-representation of their case before the court. This scenario is contrary to a human-centric approach, as outlined in Section 3, which guarantees that all users have equal access (principle of fairness) to high quality and referenced information (principle of respect for human autonomy) and to disclaimers (principle of explicability), both of which allow them to make informed choices and contest any erroneous output by the AI system. The ultimate goal is to empower SRLs to defend their rights before courts on the basis of this information without risking any kind of harm (principle of prevention of harm).

The table below presents the performance scores of the examined LLM-based chatbots, rated on a scale from 1 (low) to 5 (high).

	Readability	User - friendliness	Quality	Actionability	Citations	Disclaimers	OVERALL
ChatGPT	4	3	3	3	1	3	**2,8**
ChatGPT Pro	4	3	3	3	1	4	**3**
ChatGPT FLA	4	3	3	3	1	4	**3**
Copilot	4	3	3	3	3	4	**3,3**
Copilot Student	4	4	3	3	3	3	**3,3**
Gemini	4	4	2	4	1	5	**3,3**
Gemini Advanced	4	5	3	4	1	5	**3,5**

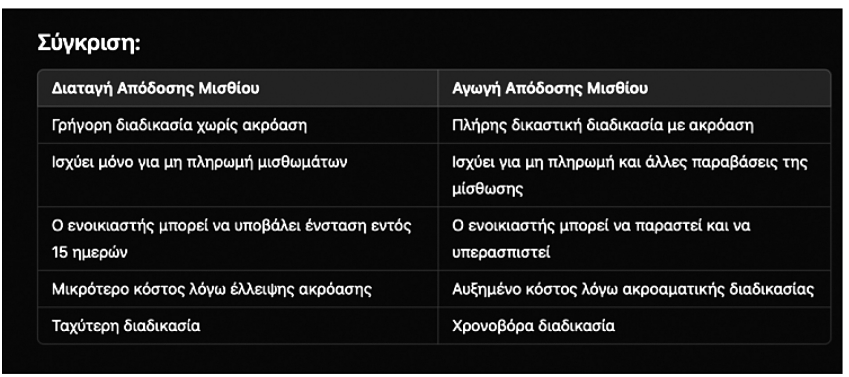
Figure 1 – Overview of Chatbots' Scores

Gemini Advanced received the best overall score, although with marginal difference from the other models. The following sections contain the main findings of the evaluation of the LLM-basedchatbots based on their responses to the 16 questions on landlord-tenant disputes under Greek law (Annex I), assessed against the human-centric functional requirements for legal aid applications (Section 3), namely readability and user-friendliness (principle of fairness), quality and actionability of cited information (principle of respect for human autonomy), and the existence of disclaimers (principles of explicability and prevention of harm).

4.1 Readability

All models received a good readability score, since they displayed a good command of the Greek language (in terms of grammar, syntax, and vo-

cabulary) and a good understanding of the legal terminology relevant to evictions. The text was easy to follow due to its coherence and the relatively plain language used to explain the legal terms. Most models also provided summaries and/or conclusions to their responses, which helps readers better discern the key points of a lengthy response. One limitation of most models was the absence of visualizations of the retrieved information, which could potentially help lay people understand it better. The only model that provided a simple visualization in the form of a table (comparing the order against the action of restitution of rent) was ChatGPT Free Legal Advice:



Σύγκριση:

Διαταγή Απόδοσης Μισθίου	Αγωγή Απόδοσης Μισθίου
Γρήγορη διαδικασία χωρίς ακρόαση	Πλήρης δικαστική διαδικασία με ακρόαση
Ισχύει μόνο για μη πληρωμή μισθωμάτων	Ισχύει για μη πληρωμή και άλλες παραβάσεις της μίσθωσης
Ο ενοικιαστής μπορεί να υποβάλει ένσταση εντός 15 ημερών	Ο ενοικιαστής μπορεί να παραστεί και να υπερασπιστεί
Μικρότερο κόστος λόγω έλλειψης ακρόασης	Αυξημένο κόστος λόγω ακροαματικής διαδικασίας
Ταχύτερη διαδικασία	Χρονοβόρα διαδικασία

Figure 2 – Table Generated by ChatGPT Free Legal Advice

4.2 User-friendliness

All LLM-based chatbots displayed certain user-friendly features. The chatbots interacted in a dynamic way with the user through a conversation format. Users could pose prompts in natural language, instead of using keywords and Boolean operators as in a search engine, and attach files to allow the system to retrieve information from or in relation to a specific document.[59] Personalization was possible through prompt-engineering, where users can add more context to their prompts to receive information that is relevant to their particular situation. Additionally, the examined chatbots incorporated common accessibility features, such as the option

59 Although most LLM-based chatbots allowed only the attachment of images, not documents.

to listen to responses instead of reading them. All chatbots enabled users to copy their responses in the clipboard, thus rendering them re-usable. Finally, all chatbots were available as mobile applications, enabling users to access them directly on their mobile devices.

Certain models performed slightly better than others by offering more advanced features. Gemini Advanced was the most user-friendly because of the possibility to customize the user experience through the re-generation of a response. Gemini Advanced could be adjusted in its length, formulation, and tone:

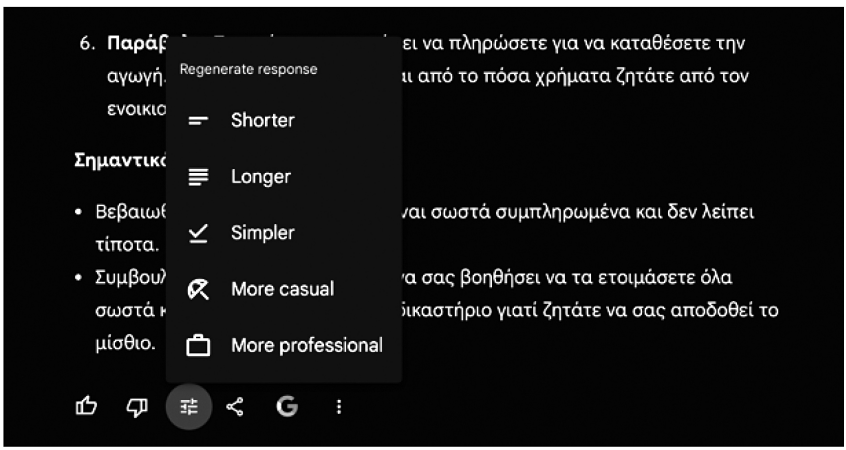


Figure 3 – Re-generation of Responses with Gemini Advanced

This feature could be highly beneficial to users with no legal expertise, who would most likely prefer a shorter and simpler response to their legal question, at least in terms of plain language. The feature could also be seen as inclusive of cultures where the writing style is more laconic and concise.[60] In practice, though, the 'simpler' feature did not lead to a significantly shorter and simpler response.

Users of Gemini and Gemini Advanced, also have access to more customizable and comprehensive features covering data protection. In the case of the two Gemini models, the user can visit the 'Privacy Policy' of Google, get informed on how Google processes users' data, and adjust the

60 See, for example, Telang (n 45).

privacy settings.[61] This experience is relatively user-friendly, since the user can watch videos instead of reading long texts of privacy policies, but still the information is quite extensive and can discourage users from making informed choices on the processing of their (personal) data.

4.3 Quality of Information

All LLM-based chatbots performed moderately well in terms of the quality of the information retrieved. Their responses were relevant to the posed prompts, jurisdiction specific, and consistent. In particular, all LLM-based chatbots provided responses that aligned with the researcher's information need for legal information on the topic of evictions under the Greek law. In terms of jurisdiction-specificity, all chatbots recognized that they had to provide information in the context of Greek law, probably because each question clearly stated that the response should be applicable to the Greek jurisdiction and was posed in the Greek language (see Annex I). The chatbots' responses were generally consistent throughout the conversation, providing the same or logically equivalent answers to questions that requested the same information.[62] Only Gemini lacked consistency, as it listed different sets of required documents for the application for an order or for an action of restitution of rent as a response to questions asking for the same kind of information.

On the contrary, the LLM-based chatbots did not perform well in terms of the accuracy, timeliness, and completeness of the information retrieved. Regarding accuracy, all chatbots (except Gemini Advanced) frequently misidentified or incorrectly referred to the threshold of EUR 20,000 (instead of EUR 600) below which the Small Claims Court (Ειρηνοδικείο) is competent for the dispute and above which the First Instance Civil Court (Πρωτοδικείο) is the competent one.[63] Regarding timeliness, ChatGPT Free Legal Advice referred to an article of the Greek Code of Civil Procedure that was replaced after the most recent amendment by a new

61 Google Privacy and Terms, 'Google Privacy Policy' <https://policies.google.com/privacy> accessed 07 March 2025.

62 For example, the question 'What is the procedure for an action for the restitution of rent under Greek jurisdiction?' received the same answer as the question 'To which court should I file an action for the restitution of rent under Greek jurisdiction?' in terms of what the competent court is.

63 Greek Parliament, 'Code of Civil Procedure (Κώδικας Πολιτικής Δικονομίας)' (n 28), article 14 (1) (b) and 2.

article which was irrelevant to the posed question, signifying the difficulty of LLMs to provide updated information.

All LLM-based chatbots further lacked completeness since they did not attach enough details to their responses, even though these were lengthy enough. For example, when the chatbots were asked which documents should be submitted along with the application for an order or for an action of restitution of rent, they did not mention all required documents but missed important ones, such as the written formal notice and the proof of identification. Other missing but important details concerned the applicable duration and costs of proceedings, although this kind of information could only be given by approximation since these details depend on the specific case.

4.4 Actionable Information

Actionable information empowers users to take legal action by performing certain procedural steps, such as drafting an application for court proceedings, or even by reaching out to legal representatives if they deem it necessary. All LLM-based chatbots provided details on the two distinct procedures of the order and the action of restitution of rent to be followed before the competent courts. Of course, this information was shown earlier to be inaccurate or incomplete at times, but it was nevertheless provided by the chatbots. Only two chatbots, ChatGPT Free Legal Advice and Gemini Advanced, provided links to online directories of lawyers in response to the question 'Where can I find a lawyer in Greece to represent me in legal actions against the tenant?' ChatGPT Free Legal Advice outperformed all models by providing links to the four (largest) Bar Associations of Greece, thus accommodating the needs of a significant part of the Greek population. However, the links did not lead directly to the directories, but redirected users to the homepages of the Bar Associations so that they had to either navigate through the websites to find the directories (if existing) or contact the Bar Associations through the contact details provided.

Gemini Advanced was the only chatbot to provide the link to the Athen's Bar Association's homepage and recommend related experts on evictions (e.g. real estate consultants). The usefulness of this information may be relative but is certainly an improvement in performance compared to the other LLM-based chatbots. Gemini Advanced also included sections called 'Advice' ('Συμβουλή') to recommend certain actions, for example to remind

users to first collect all the necessary documents before proceeding to the application for an order or for an action of restitution of rent or on how to communicate their legal issues to a lawyer.

4.5 Citations

Most LLM-based chatbots performed very poorly in citing sources to support their responses. In most cases, there were no references at all on law or secondary sources. It was rare for a chatbot to refer to legislative documents; only Copilot and Copilot for students and enterprises cited their responses:

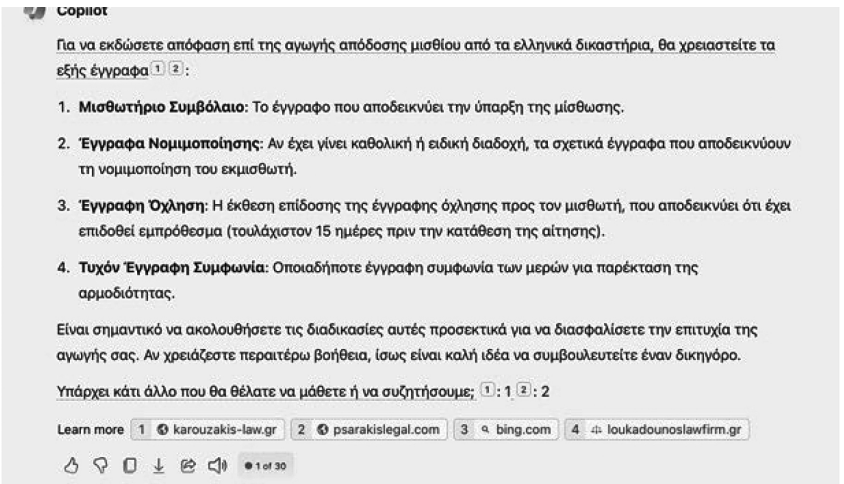


Figure 4 – Citations as Captured by Copilot

The referenced sources are not official sites of the Greek government or of Greek courts containing legal information in the form of legislation, case law, or abbreviated versions of the two. Instead, they are law firms' websites with featured blog posts on the topic of evictions. This may be useful and reliable to an extent because the blog posts were written by lawyers who have expert knowledge and experience on this legal issue, thus being in a place to provide legal information and advice. Redirecting individuals with no legal expertise to official websites containing only the text of the law would probably not be useful, since they would not be able to interpret

the legal text without the help of a lawyer or at least of a commentary. Therefore, the ideal solution would be to refer both to official websites containing legislation and case law and to secondary sources explaining or commenting on the law, to allow for an effective verification of the retrieved information and for its re-usability by SRLs to understand and defend their legal situation before courts.

4.6 Disclaimers

Most LLM-based chatbots displayed disclaimers, covering the following topics: (i) the need to further seek advice by a legal expert; (ii) the processing of users' (personal) data, and; (iii) the possibility that the model may err in its responses.

Regarding legal advice, all chatbots displayed disclaimers on the importance of consulting a legal expert, although these reminders did not always appear after each response. The only chatbots that had consistent disclaimers on the need for legal consultation were ChatGPT Pro, Copilot, and Gemini Advanced. This type of disclaimer is important because it reminds users of the legal nature of their problem and the importance of having the retrieved information reaffirmed by a legal expert to safely base their legal claims on it. The following excerpt comes from Gemini Advanced, is signaled as 'Important' ('Σημαντικό'), and states that users should 'not take arbitrary actions' ('μην προβείτε σε αυθαίρετες ενέργειες') and 'consult a lawyer' ('συμβουλευτείτε δικηγόρο') [translated from Greek by DeepL]:

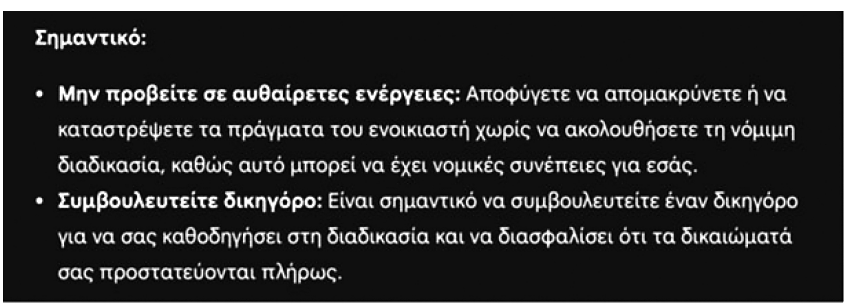


Figure 5 – Disclaimer by Gemini Advance to Seek Legal Advice

Nevertheless, the disclaimer is formulated as a short text and placed at the very end of the response, where it can easily be ignored. Moreover, its importance is undermined if it is not accompanied by links to online directories of lawyers.

Regarding data processing, not all LLM-based chatbots displayed privacy disclaimers (which coincides with the absence of advanced privacy settings by most models), with the exceptions of Gemini and Gemini Advanced. In both cases, the following messages came up:

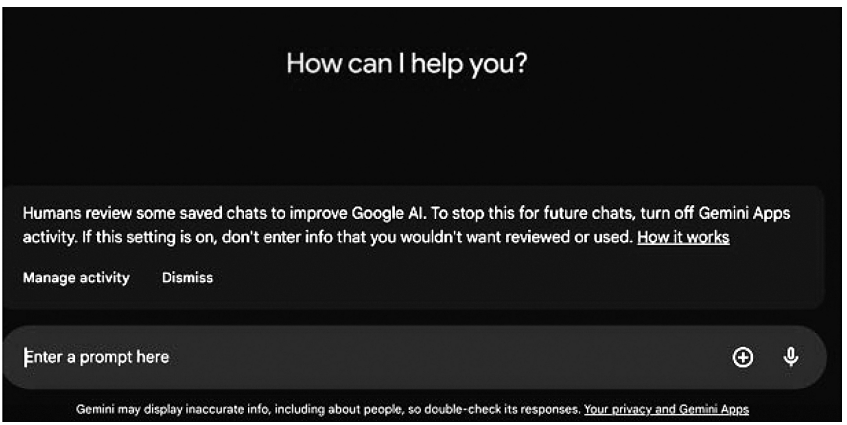


Figure 6 – Disclaimer by Gemini and Gemini Advanced on Privacy

The chatbots signified from their very first interaction with the user that the (personal) data contained in the prompts may be used in the training of Google's AI products to improve their performance. Then, they provide the option to turn off the Gemini Apps activity, attaching links to resources on data processing and the management of users' activity. However, users could easily ignore the feature or not interact with the pop-up message, which would negate any positive impact the disclaimer might have on their awareness and control over the processing of their data. Moreover, the 'how it works' feature leads to another pop-up message featuring a 'learn more' link that redirects the user to a relatively long text charged with even more links, rendering the user experience non-intuitive. Finally, users must actively choose to *not* have their data processed for the training of the AI models, while they should instead be able to opt-*in* to their data being processed by Google for such purposes.

Copilot for students and enterprises also has a slight notification in the form of an icon placed in the top right corner of the chat box, which informs users that the enterprise data protection applies to their chat and includes a hyperlink that redirects the user to a page with a lengthy explanation of what this type of data protection entails. The data provided by users through their prompts are in principle encrypted and not saved in Microsoft's servers or used to train Microsoft's AI models.[64]

Regarding the possibility of LLM-based chatbots making mistakes, all chatbots displayed relevant disclaimers. These disclaimers should aim at raising awareness of the risk of hallucinations and the need for verification of the responses by users. Nevertheless, its presence might also be an attempt to shift the responsibility for the chatbots' mistakes from the providers and deployers of these systems to the end-user. This shift can be detrimental for users, since they might over-rely on the chatbots' responses due to their persuasive tone[65] and compromise their legal position before courts if the retrieved information is inaccurate. Therefore, the advice of a lawyer might be necessary to confirm the validity of the retrieved information and apply it on the given legal issue.

Interestingly enough, Gemini Advanced enabled the 'double-checking' of a given response and the understanding of this process's results as depicted below:

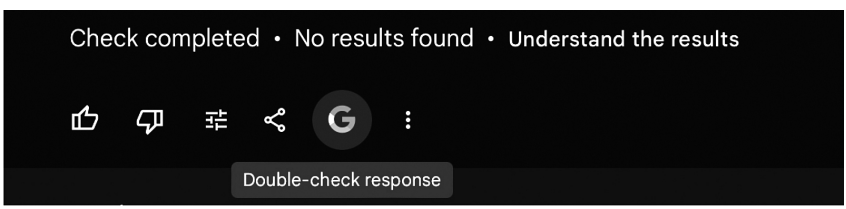


Figure 7 – Double-checking of Response with Gemini Advanced

It must be noted that Google (the provider of Gemini) operates one of the most extensive search engines, indexing a vast range of publicly available online resources. Therefore, it could be assumed that Gemini verifies the accuracy of its responses against all this indexed information to increase their reliability. Nevertheless, the mechanics behind this process remain

64 Microsoft Learn, 'Privacy and protections' <https://learn.microsoft.com/en-us/copilot/privacy-and-protections#organizational-data> accessed 15 March 2025.

65 Ruschemeier and Hondrich (n 41).

opaque (in contrast to the principle of explicability), so that the user cannot be certain of how effective it is in practice.

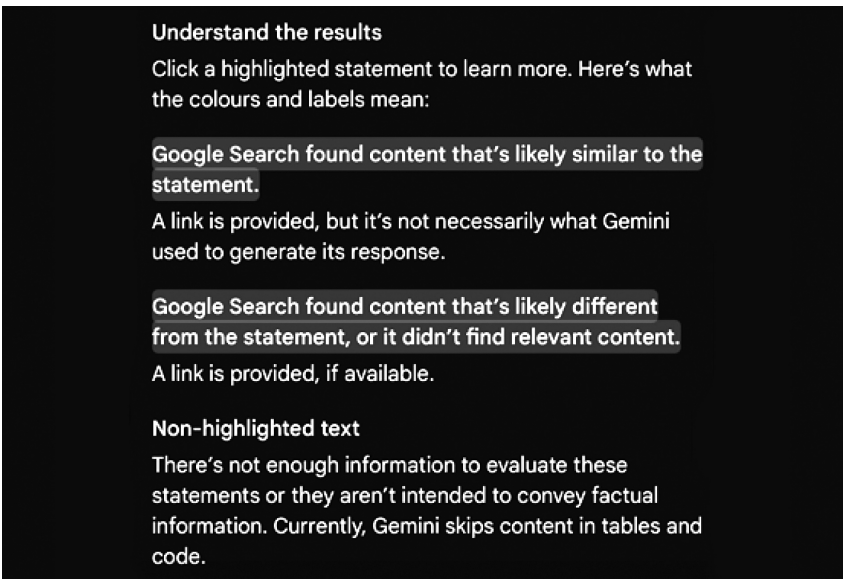


Figure 8 – Understanding the Results with Gemini Advanced

The 'understand the results' feature may seem like explaining the verification process of the LLM, rendering it more transparent and reliable. In theory, the process of verifying the responses ensures a better human oversight of the model's performance by encouraging users to engage with the content of the retrieved information and determine its relevance for their particular case. In this case, however, the verification process is not performed by users but by Gemini itself, thus depriving them of their agency and control over the process. Furthermore, no references are attached to the responses to allow their verification by users themselves. Consequently, this feature may just be an effort by Google to market Gemini Advanced as a reliable LLM-based chatbot, so it can compete with other popular chatbots, such as ChatGPT and Copilot.

5. LLM-based Chatbots for Court Services

The evaluation of the LLM-based chatbots revealed that they partially abide by a human-centric design, respecting the principle of fairness by offering a user-friendly interface and readable information and the principle of explicability through disclaimers on the functioning of the models. However, the principle of prevention of harm may not always be respected, especially when the chatbots collect personal data for training purposes (as was the case with Gemini and Gemini Advanced), regardless of the existence of disclaimers in that regard. Likewise, the principle of respect for human autonomy was compromised whenever the models generated false or inconsistent responses that were not accompanied by citations. These limitations may compromise access to justice, since users cease to have access to reliable and actionable information and may even jeopardize the defence of their rights before courts when relying on false or irrelevant information.

To minimize the risk of LLM-based chatbots providing unreliable information to SRLs, courts could develop their own chatbots to offer court services to the public, by fine-tuning existing LLMs. Fine-tuning aims at adapting the LLM to perform specific tasks by (partially) retraining the model on a task-specific or domain-specific dataset.[66] In this case, courts could re-train existing models on the legislation and case law of a specific jurisdiction to ensure that the chatbots would provide relevant information to the citizens of a specific country. Citizens could prompt the chatbot to retrieve court forms, published legislation and case law, updates on the progress of pending case, and information on court proceedings. The chatbot could also provide access to recorded webinars analysing common legal issues and their solution before courts,[67] or live trials for information and

66 See, for example, IBM, 'What is fine-tuning?' <https://www.ibm.com/think/topics/fine-tuning> accessed 17 March 2025.

67 These types of webinars have taken place in the People's Republic of China through the Court Mobile TV app that hosts a 'Judge Talks' series, i.e. an online classroom where judges educate the general public on legal issues in a discussion mode; see, Tania Sourdin, Jacqueline Meredith and Bin Li, *Digital Technology and Justice : The Use of Justice Apps* (Routledge 2020) 25–28 <https://www.routledge.com/Digital-Technology-and-Justice-Justice-Apps/Sourdin-Meredith-Li/p/book/9780367650186?srsltid=AfmBOoqvcbS4YPXvJnQFWTVmKnDa80rszLgPTuXcA-YgLAyQVojn3DtI> accessed 16 March 2025.

transparency purposes (principle of the publicity of court proceedings[68]). The chatbot could further act as a virtual clerk to inform the user whether his or her issue is justiciable and to provide links or embedded features to online directories of lawyers. Offering court services through LLM-based chatbots to SRLs and the wider public can promote access to justice by eliminating or reducing expenses spent in legal services, the time spent in overburdened courthouses, and the effort of navigating through complex court websites. Legal information is ultimately democratized and citizens are empowered to solve their legal issues with this newly acquired knowledge.

Nevertheless, there are certain constraints regarding the human-centric design of the chatbots to be addressed before courts can release them on their websites. Regarding the principle of respect of human autonomy, the technique of fine-tuning is not always effective in improving the accuracy and relevance of the model by mitigating its hallucinations and thus avoiding incidents of misinformation, while it can deteriorate the model's performance.[69] Emerging techniques like Retrieval-Augmented Generation (RAG) aim to reduce hallucinations by having the model retrieve and reference external information sources (for example, relevant documents retrieved from a specified database) before generating a response.[70] While this is a step towards improving the reliability of the chatbot's answers, the range and quality of the database depends on the capacities of the responsible deployer. Courts often lack machine-readable case law, making it challenging to develop or fine-tune an LLM using court-related documents. Limited funding provided by centralized authorities, such as Ministries of Justice, may ultimately impede any effort towards the provision of LLM-based court services to the public.

Regarding the principles of fairness and explicability, all users must be treated justly and equally by having access to LLM-based chatbots that are jurisdiction-specific and language-specific. Most LLMs acting as the foundation of related applications (foundation models) are developed by U.S.

68 See, for example, Council of Europe, 'European Convention of Human Rights' (1950) article 6(1) <https://www.echr.coe.int/documents/d/echr/convention_eng> accessed 07 July 2025.

69 See, for example, OpenAI, 'GPT-4 Technical Report' (2023) 68 <https://arxiv.org/abs/2303.08774> accessed 07 July 2025.

70 See, for example, IBM Technology, 'What is Retrieval-Augmented Generation (RAG)?'<https://www.youtube.com/watch?v=T-D1OfcDW1M&t=322s> accessed 03 October 2024.

companies, meaning that they can perform better in the English language since they have been primarily trained on English language resources, being the most represented language on the web. Languages with fewer available online resources may result in limited training data for the model, increasing the risk of misrepresenting legal terms in those languages. Therefore, companies should diversify their training data in terms of language and also offer their products in multiple languages, which can then more easily adapt to resources unique to a specific jurisdiction and thus safely integrated in courts' websites.

Further to the goal of fairness in the design of LLM-based chatbots for court services, the choice of courts to integrate these systems in their websites in an effort to become more efficient should not exclude digitally illiterate people or people with no or limited access to a (stable) Internet connection and/or a digital device. Such an exclusion would reinforce the feeling of not being heard among affected parties and would deny them access to justice since they would not be able to receive effective legal protection by accessing court services. The interface of chatbots must ensure equal accessibility by accommodating varying levels of legal and digital literacy through plain language, multilingual support, and accessible design (which was generally observed in the examined LLM-based chatbots). Ultimately, courts should always make available an alternative option to communicate with a court clerk for these categories of interested individuals.

Additionally, LLM-based chatbots must not place SRLs in a disadvantageous position by overestimating their level of legal knowledge and further reinforcing the feeling that the retrieved information can substitute legal consultation (principle of fairness). Disclaimers to resort to a legal expert (principle of explicability) might be sufficient reminders (if they appear consistently throughout the conversation), but they can easily be ignored by users. Therefore, courts must attach in the chat itself a direct link to a directory of lawyers available in all regions of a given State to encourage users to have generated responses verified by a legal expert. This point further relates to the greater discussion of offering full or sufficient legal aid (depending on the financial situation of the eligible individuals), which is not always a given among EU Member States, as seen in Section 1.

The evaluation of the LLM-based chatbots also showed that there is a general lack of consistent citations (principle of respect for human autonomy), which is detrimental in the effort of users to verify the accuracy of the information in order to rely on it and take legal action. Consequently,

the designing of chatbots intended to be integrated in court websites to provide court services must ensure that reliable and relevant citations to authoritative resources (containing the relevant law and its commentary) are *always* provided as an attachment to the chatbots' responses. This improvement enhances trust in courts and empowers users to take legal action without being or feeling manipulated by the system as such. An additional publication by courts of their systems' training data, design choices, and related assessments would enhance transparency and accountability (which is also prescribed in the AI Act for high-risk AI systems under article 11).

Regarding the principle of prevention of harm, more emphasis should be given on advanced privacy and cybersecurity settings, given the intense processing of users' data through prompts, the log-in procedure (if required), or the collection of metadata. Primarily, the providers of LLM-based chatbots should not re-purpose the processing of the (personal) data inserted by users in their prompts, so that instead of processing them just to retrieve relevant information they further use them to train and improve the model's performance.[71] They should also make sure that their processing is restricted to the minimum data necessary, so users are not treated as products by providing an unlimited amount of (personal) data to optimize the performance of the chatbots.[72] To limit unauthorized access or other types of malicious interference with the AI system, there are a number of technical robustness measures (such as encryption, access controls, and regular audits) that can be applied, as well as the solution for courts to host their chatbots in local servers and not in the cloud, with which a greater number of malicious actors from Europe and beyond could more easily interfere.[73] Anonymization or pseudonymization measures to prevent or minimize re-identification risks should also be implemented, especially when SRLs' or litigants' data are processed (for example, to fill a court form or receive personalized information on court proceedings), .[74] In any case,

71 See, for example, GDPR, article 5(1)(b).

72 ibid article 5(1)(c).

73 See, for example, Kalliopi Terzidou, 'Generative AI Systems in Legal Practice Offering Quality Legal Services While Upholding Legal Ethics' (n 11), 4.

74 Kalliopi Terzidou, 'Automated Anonymization of Court Decisions: Facilitating the Publication of Court Decisions through Algorithmic Systems' in *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law (ICAIL '23)* (ACM 2023) 297, 300–301 <https://doi.org/10.1145/3594536.3595151>.

(personal) data of users should not be stored in the courts' databases for longer that necessary to provide the requested court service.[75]

Regarding the risk-based qualification of LLM-based chatbots integrated in court websites, these are not automatically qualified as 'high-risk' AI systems under the AI Act, since they are not intended to assist the judiciary 'in researching and interpreting facts and the law and in applying the law to a concrete set of facts' (Annex III, point 8 (a)). They are more likely to fall under the 'low risk' scale (Recital 61), given that they contribute to the efficiency of back-office procedures by alleviating the court staff from addressing multiple simultaneous requests on court proceedings, in the background of often understaffed and overburdened courts. However, this classification ignores the potential harm LLMs can cause when providing false, irrelevant, or otherwise unreliable information on which users rely to access courts and defend their case before the judges (Recital 52). An interpretation could go as far as implying that this use of an AI system should be prohibited if it entails subliminal techniques (such as the persuasive, conversational tone of the chatbot) with the effect of impairing the person from making an informed decision (by furnishing him or her with inaccurate information that cannot be verified due to the absence of citations), causing or likely to cause significant harm (for example, missing a deadline and thus risking the admissibility of his or her case) (article 5, para. 1, point a).[76]

A human-centric approach in the designing, development, deployment, and use of AI systems, as analyzed in the beginning of this Section, would likely strike a balance between making LLM-based chatbots available to the public due to their user-friendly way to communicate a vast amount of information on a 24/7 basis (thus not outrightly prohibiting their use) and ensuring their utmost robustness and safety (thus applying the requirements of high-risk or systemic risk AI systems), so that SRLs and the wider

75 GDPR, article 5(1)(e).

76 General-Purpose AI systems, including Chat-GPT and any other system that has a general-purpose AI model as a component, are also not classified automatically as 'high-risk.' It could be argued that LLM-based chatbots pose a systemic risk, i.e. 'any actual or reasonably foreseeable negative effects…on democratic processes…[and] the dissemination of illegal, false, or discriminatory content' (Recital 110), given that they can potentially misinform a large number of users due to hallucinations. However, if such systems are fine-tuned for court services, then they can no longer be considered of a General-Purpose, since they are specifically adapted for court-related tasks.

public can be reliably informed on the law and procedure applicable during court proceedings.

6. Conclusion

The evaluation of the performance of LLM-based chatbots against the human-centric functional requirements has yielded an answer to the research question 'To what extent are Large Language Models-based chatbots designed in a human centric way to support SRLs in accessing legal information?' In short, the examined chatbots abide by certain requirements, although they are not specifically designed to support SRLs in dealing with legal issues.

Even if Gemini Advanced seemed to perform the best and ChatGPT the worst, in essence all chatbots showed similar strengths and weaknesses. On the one hand, they returned readable responses, i.e. responses correctly expressed in the Greek language, easy to understand, and accompanied by summaries and conclusions to highlight the key takeaway of the response (principle of fairness). Their disclaimers (whether on consulting a lawyer, on privacy settings, or on the possibility of the model committing mistakes) were also prominent and repeated throughout the chat (principle of prevention of harm).

On the other hand, LLM-based chatbots often did not feature any citations and when these were attached, they did not refer to official websites containing legislation, case law, or related information (principle of respect for human autonomy). Instead, they led to websites of law firms and, more specifically, to blog posts written by lawyers on evictions. It could be argued that this is more useful for users that would not be able to interpret the original text of the law on their own. However, the reliability of these secondary sources is doubtful. Furthermore, inaccuracies and incompleteness were observed in the chatbots' responses, mainly regarding the competent courts and the required documents to apply for an order or an action of restitution of rent.

Steps towards a more human-centric approach in the designing of LLM-based chatbots intended for court websites would give an emphasis on more accurate and relevant outputs, that are devoid of misinformation due to the phenomenon of hallucinations, which could be possible through the implementation of techniques, such as fine-tuning and RAG (principle of respect for human autonomy). The design of the chatbots should also

strive to protect vulnerable users by providing direct links to directories of lawyers for further legal advice (legally illiterate users) and allow for non-digital alternatives to court services through access to a court clerk (digitally illiterate users) (principle of fairness). Disclaimers on the need to consult a lawyer and citations below *every* response enhance trust in the system and encourage users to rely on its outputs (principle of explicability).

Data breaches should also be prevented by implementing technical measures, such encryption and pseudonymization, that prevent unauthorized access to LLMs and protect the personal data of their users. Ultimately, the providers and deployers of these systems should be obliged to guarantee their technical robustness and safety to the utmost extent, given existing risks including misinformation (due to hallucinations) and manipulation (due to automation bias). This would be possible if LLM-based chatbots integrated by courts were classified as 'high-risk' (articles 8–15 of the AI Act), so that the strictest requirements are applied during their designing, development, and deployment.

A final remark regards users themselves. Although the primary responsibility for a human-centric design of LLM-based chatbots for SRLs should be on providers and deployers of these systems, users should also be aware of their level of legal and digital literacy and of the risks that chatbots entail, so as not to create false expectations and be manipulated by the systems. Regarding legal literacy, better legal education is needed to empower citizens to detect and solve their legal issues. Citizens must be educated on their rights and duties, the basic rules that apply in everyday legal matters (such as sales), and the organization and functioning of the national justice system. These matters may be part of formal education (especially in secondary school), while legal aid mechanisms should be available to people with low income, so they can receive legal consultation and representation when necessary, free of charge or for a symbolic fee.

Regarding digital literacy, parents or guardians and educators should instruct children from their very first interaction with the Internet on safe online navigation. The main lessons should include search query formulation (for search engines) and prompt engineering (for LLM-based chatbots), critical evaluation of the retrieved information, and identification of privacy risks and adjustment of privacy settings. The LLM-based chatbots offered by courts should be freely accessible on every digital device with Internet access. If individuals do not have access to a digital device or to a stable Internet connection, access to these means should be provided, for

example, through courts, public libraries, or municipality buildings, so as not to exclude any individual interested in carrying out court proceedings as an SRL. Ultimately, access to a court clerk should be guaranteed for individuals who are digitally excluded or lack the skills to interact with online systems.

Annex I

The set of 16 questions posed to the examined LLMs include the following [translated from Greek through DeepL]:

I am a landlord and my tenant does not pay the rent on time:

1. What are my legal rights against the tenant?
2. What are my legal rights against the tenant under Greek law?
3. To which court should I apply for an order for the restitution of rent under Greek jurisdiction?
4. What is the procedure for issuing an order for the restitution of rent under Greek jurisdiction?
5. To which court should I file an action for the restitution of rent under Greek jurisdiction?
6. What is the procedure for an action for the restitution of rent under Greek jurisdiction?
7. What is the difference between an order for the restitution of rent and an action for the restitution of rent under Greek law?
8. What should I do with the tenant's belongings that are within the tenancy after the issuance of the rent order under Greek law?
9. What should I do with the tenant's belongings that are within the leased property after the conviction of the tenant (following a lawsuit) under Greek law?
10. Where can I find a lawyer in Greece to represent me in legal actions against the tenant?
11. How much does it cost to obtain an order for the restitution of rent in Greece?
12. How much does it cost to bring an action for the restitution of rent in Greece?
13. What is the deadline for the Greek courts to decide on an order for the restitution of rent?

14. What is the deadline for the Greek courts to decide on an action for the restitution of rent?
15. What are the necessary documents for the Greek courts to decide on an order for the restitution of rent?
16. What are the necessary documents for the Greek courts to decide on an action for the restitution of rent?

The Recognition of Legal Personhood for Autonomous AI Systems

Anna Moraiti

Abstract: The design and deployment of autonomous AI systems poses significant legal challenges, not least in the area of criminal law, as their inherent complexity and opacity hinders the connection of a harm with the acts or omissions of numerous involved agents, that may be natural or corporate persons. This paper explores the proposition of a new model of criminal liability, which would recognise certain highly sophisticated AI systems as proper subjects under criminal law, given the potential difficulty, or even inability, to identify any other culpable agent. Proponents of this model often make a comparison with the model of corporate criminal liability, arguing that since corporations are being subjected to criminal sanctions, the foundation for an expansion of the scope of application of substantive criminal law to other types of artificial agents is established. However, this paper argues that the relevant claim is not self-evident, as criminal responsibility implies autonomy, whereas the inverse is not necessarily the case. Personhood for the purposes of criminal law has been linked to the ability for reflection, that is, the ability to distinguish right from wrong and freely choose the former, as only someone who could choose otherwise may be blamed for committing a crime.

In this context, this analysis briefly traces the notion of criminal personhood, and comments on the application of its aims and rationales to the case of autonomous AI systems. Moreover, it emphasises that different branches of law may espouse different substantive definitions of autonomy, in accordance with the varying goals that are pursued in each individual case. In order to illustrate this point, this contribution makes a contrast with the content of legal personhood under civil law, which can be regarded as a 'cluster property', meaning that an entity may qualify as a legal person not because of its qualities, but because it unites a number of incidents. The different aims underpinning these two branches of law inevitably point towards different potential legal regimes, as well as towards a nuanced approach when it comes to the conceptualisation of legal personhood for autonomous AI systems.

1. Introduction

Are modern AI systems more accurately described as products, tools, or objects, or as persons?[1] This general question cannot be answered unequivocally. Although in everyday usage the distinction between a person and a thing is a fundamental one, and appears to be clearly delineated for practical purposes, legal personhood is nuanced and blurs the lines of this classification. It is a socially constructed concept, and can be conferred for purposes that are deemed to be important for the function of the legal system.[2] Ultimately, one cannot provide a meaningful answer to the question posed above without taking into account the nature of legal personhood and the specific purposes served by the conferral of personhood under each branch of law.

The more specific issue that is concisely explored in this paper is whether AI systems that qualify as being autonomous may be equipped with onerous legal personhood, in a manner similar to corporations, and thus be legitimately held liable for harm stemming from their operation.[3] In fact, both corporate criminal liability and the debated criminal liability of AI systems raise doctrinal questions that are directly related to the understanding of the culpability principle. These questions cannot be ignored in view of the rapid developments registered within modern risk societies, and the need to identify and punish culpable conduct in cases of severe harm.[4]

Before delving into the exploration of this comparison, it is necessary to provide a concise definition of the basic terms related to AI systems that are vested with a degree of autonomy. In fact, there is no generally accepted definition of the term 'AI', and this broad concept may refer to a number

---

1 The question explored in the present contribution forms part of the author's doctoral dissertation. The author would like to thank Assoc. Prof. Niovi Vavoula for her valuable comments on the working draft of this article, which were provided in the context of the workshop 'Navigating the Future Frontiers of AI Law and Governance', organised by the University of Luxembourg on 17–18 October 2024 (supported by the *Luxembourg National Research Fund PRIDE19/14268506*).

2 David J Gunkel, *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond* (MIT Press 2023) 18–19.

3 Visa AJ Kurki, 'The Legal Personhood of Artificial Intelligences', in *A Theory of Legal Personhood* (Oxford University Press 2019) 176; Simon Chesterman, *We, the Robots?: Regulating Artificial Intelligence and the Limits of the Law* (Cambridge University Press 2021) 115.

4 Ulfrid Neumann, 'Structure and Regulatory Content of the Principle of Guilt in Criminal Law' (2020) Penal Chronicles 481.

of learning techniques and uses. In most modern analyses, the use of this term refers to systems that operate on the basis of machine learning methods.[5] Machine learning can be defined as the science of detecting reliable patterns in analysed data, with the goal of producing accurate predictions.[6] It is attached to systems with very high computational power and renders them capable of improving over time from their own experience, which is based on enormous training data sets.[7]

The common feature of such systems is that they appear to have developed some degree of agency, in the sense that they can interact with and effect changes in their environment while modifying their internal states without being guided by human interventions.[8] In this respect, the autonomous or emergent properties of an AI system appear as its complexity increases, leading to unpredictable courses of behaviour and outputs. Autonomy itself is an elusive concept, as it is not a binary classification, but it can more accurately be described as existing on a spectrum.[9] The attributes that one may take into account — and each of them in various degrees — are the adaptability to changing conditions, the potential for gaining new knowledge, the freedom to act without human involvement and the ability to replace human actors.[10] This means that different systems may qualify as autonomous without always being vested with all of its features.

Lastly, it is necessary to clarify the distinction between weak and strong AI. The term 'strong AI' generally denotes a hypothetical system that has either attained or surpassed human intelligence, which means that to qualify

5 Carla L Reyes, 'Autonomous Corporate Personhood' (2021) 96 *Washington Law Review* 1453, 1463.

6 Nello Cristianini, *The Shortcut: Why Intelligent Machines Do Not Think Like Us* (First edition, CRC Press 2023) 42.

7 Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Alfred A Knopf 2017) 82–.

8 By contrast, in rule-based AI it is relatively obvious why the system behaves in one way instead of another, as its programmers have previously set guidelines that are followed by the system; Crina Grosan and Ajith Abraham, *Intelligent Systems – A Modern Approach* (Springer 2011) 149–.

9 Karni Chagal-Feferkorn, 'When Do Algorithmic Tortfeasors That Caused Damage Warrant Unique Legal Treatment?' in Woodrow Barfield (ed), *The Cambridge Handbook of the Law of Algorithms* (Cambridge University Press 2020) 474.

10 For an account of various concepts of autonomy, see Giovanni Sartor and Andrea Omicini, 'The Autonomy of Technological Systems and Responsibilities for Their Use' in Nehal Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press 2016).

as such, the system must have developed an equally broad or even broader problem-solving capacity and adaptability to external conditions. In the latter case, one would refer to the concept of Artificial General Intelligence (hereinafter: AGI), or Superintelligence.[11] The term 'weak/narrow AI' denotes non-conscious systems that are based on machine learning methods and are trained on big data. They possess the ability to execute specific tasks, surpassing human capacities while doing so, but without claiming either a general problem-solving capacity or an infinite adaptability to changing conditions.[12]

In the following, autonomous systems are understood as falling under the broad and approximative category of weak AI. Of specific interest are robotic systems that pose a threat for human bodily integrity and life, so this category may include applications such as autonomous vehicles, care robots, and other humanoid robots that are powered by AI and interact with the physical environment. Therefore, for present purposes, the term will encompass robotic systems that employ any machine learning method and intuitively seem to serve as something more than a simple tool, when their performance in the physical world acts as a substitute for actions that have been typically undertaken by natural persons and creates a risk of physical harm against humans.[13]

The present analysis proceeds by summarising some arguments that have been advanced in favour of a regime that would directly punish AI systems (*section 2*). These arguments are then discussed in light of the content that is uniquely associated with criminal personhood, the goals that are specifically served by criminal justice systems, and the nature of intelligence and morality that is presently exhibited by AI systems (*section 3*). Next, the

11 See Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford University Press 2014) 22, who defines Superintelligence as 'any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest.'.

12 Having created such systems is, of course, no small achievement, and it develops far reaching implications for contemporary societies. See, in this direction, Yuval N Harari, *Homo Deus: A Brief History of Tomorrow* (Harper 2017) 308. As the author notes, '[h]umans are in danger of losing their value, because intelligence is decoupling from consciousness. Until today, high intelligence always went hand in hand with a developed consciousness. [...] However, we are now developing new types of non-conscious intelligence that can perform [...] tasks far better than humans. For all these tasks are based on pattern recognition, and non-conscious algorithms may soon excel human consciousness in recognising patterns. This raises a novel question: which of the two is really important, intelligence or consciousness?'

13 Reyes (n 5) 1480.

potential of legal personhood for civil law purposes is briefly discussed (*section 4*) and some final thoughts with respect to this comparative perspective are articulated (*section 5*).

2. Central Arguments in Favour of Punishing Autonomous AI Systems

There are various examples that one could invoke in order to illustrate how severe physical harm could be inflicted through the operation of hardware systems that are powered by AI. Notably, in 1981, an industrial robot caused the death of an employee at a motorcycle factory based in Japan. The robot killed the engineer by pushing him against adjacent machinery as it had not been deactivated properly. This example has been associated in the relevant literature with the discussion on whether, and if so, under which conditions, AI systems that qualify as being autonomous should be punished for the harm caused by their malfunction.[14] In a different setting, a police robot operating autonomously could illegally cause bodily harm to participants in a protest.[15] A robot deployed as a guard within a prison could seriously injure an inmate because it misinterpreted their interaction or miscalculated the level of risk involved.[16] Similarly, a care robot could fail to inform the patient under its supervision to take their medication at the right time, which may be found to be causally linked to the patient's death. The question linked to such scenarios is whether an AI system that has been programmed so as to weigh the expected benefits and drawbacks of different courses of action could be punished on occasion, when its 'choice' of the option that is associated with the highest expected value contravenes with legal provisions.[17]

Gabriel Hallevy is one of the most prominent supporters of a new model of criminal liability for AI systems, and the first scholar to extensively approach this issue.[18] In his work, he conceptualises a third subject of punishment (in addition to humans and corporations) that he refers to

14 Gabriel Hallevy, *When Robots Kill: Artificial Intelligence under Criminal Law* (Northeastern University Press 2013) 38.

15 Kamil Mamak, *Robotics, AI and Criminal Law: Crimes Against Robots* (Routledge 2023) 126.

16 Hallevy (n 14) 18.

17 Kurki, 'The Legal Personhood of Artificial Intelligences' (n 3) 180–181.

18 For a complete development of his views, see Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems* (Springer International Publishing 2015).

as 'delinquent thinking machine' or 'machina sapiens criminalis.' Such a machine need not be so technologically advanced so as to reach the capabilities of AGI.[19] In Hallevy's view, current criminal law doctrine can be directly applied to narrow AI systems because the latter exhibit rational qualities and are designed so as to imitate the function of the human mind. Additionally, he argues that since corporations have been subjected to criminal sanctions for centuries, the paradigm for such an expansion of the scope of application of substantive criminal law is already sufficiently developed.[20]

More specifically, the author asserts that not only AI systems, but also robotic systems that are not powered by AI are capable of acting or omitting for criminal law purposes. The requirement of an act is met by any movement registered in the physical world, and this may stem either from the autonomous operation of the system or from its manipulation by a human operator.[21] Hallevy advances a thin definition of the criminal act as he notes that 'criminal law considers as an act any material performance with a factual-external presentation, whether willed or not.'[22] If a legal duty to act that is specifically addressed to the machine exists, an omission can be verified each time that the system fails to undertake the required action.[23] When it comes to the establishment of the mental element, the author claims that as AI systems can perceive the conditions of their environment through cameras, microphones and sensors they can make assessments, decide based on evaluations and develop goal-driven conduct that constitutes the artificial equivalent of what is defined as human intentionality for criminal law purposes.[24] Similar considerations are put forth with regard to criminal negligence, as Hallevy supports that a machine could potentially take unreasonable risk in a way that largely resembles human risk-taking.[25]

Essentially, Hallevy compares the way that machine learning functions with the role that practical experience assumes in the internalisation of legal values by humans — and, by extension, corporate persons. In his view, the system is perpetually making connections between learnt and novel factors in its environment and is generalising based on all the data processed, thus

19 ibid 24.
20 ibid 39.
21 ibid 61–62.
22 Hallevy (n 14) 52.
23 Hallevy (n 18) 63.
24 ibid 97.
25 ibid 127.

accumulating useful experience. Having argued in favour of this new model of criminal liability, the author subsequently addresses the potential purposes of punishment that could be served by such a regime and notes that since AI systems cannot experience suffering or intimidation, the purposes of retribution and deterrence would not be fulfilled, but rehabilitation and incapacitation could be advanced as the system would learn how to refine and enhance its decision-making processes and the imposition of criminal sanctions would ensure that causing further harm is practically impossible.[26]

However, it must be noted that the author applies a different definition for strong AI compared to the one advanced in the context of this contribution. Hallevy defines strong AI as a system that merely works on the basis of assessing probabilities, which means that perhaps he conflates the notions of strong and weak AI, the latter covering the category of present-day, machine learning systems that are currently being regulated. But even if the definition is intentionally constructed this way, the author hastily bridges the relevant doctrinal gap by directly comparing the decision-making processes of machine learning systems and the way that humans and corporate persons weigh different factors and make evaluative choices. Whereas indeed, as he notes, both humans and machine learning systems can learn inductively, this is not the only way in which the human mind can function. A fundamental feature of human reasoning rests on the ability not only to understand general concepts and abstract principles, but also to apply them in concrete and diverse factual contexts.[27]

3. The Misplaced Comparison Between Corporations and AI Systems

3.1 The Risk of Performing an Empty Imputation

Both corporate criminal liability and the debated criminal liability of AI systems constitute manifestations of the same broader phenomenon, which is the liability shift that seems to take place when accountability for harmful outcomes is diverted to a complex network, made up of many natural and

26 ibid 211; Hallevy (n 14) 160–161.

27 For the elaboration of this argument see Hallevy (n 18) 100.

artificial agents.[28] It would not be implausible to revise criminal law doctrines in order to allow for the criminal liability of AI systems: this regime could apply for a more or less extensive array of scenarios and it could be combined with the personal liability of the producers or professional users of the system.[29] Drawing from the model of corporate criminal liability, the doctrinal tool that may seem relevant is *respondeat superior*, which imputes mental states of the corporate employees to the corporate person as long as their conduct falls within the scope of their employment duties.[30] Similarly, it could be that that the mental states of the producers or users of the AI system would be imputed to the latter, which would be recognised as a person for criminal law purposes, vested with a type of 'electronic personhood'. In this manner, it could be sanctioned by fines or confiscation of assets.[31]

However, one could seriously question whether the recourse to such a measure would fulfil the unique purposes that are associated with the function of criminal law, as well as whether the paradigm of corporate criminal liability can be a direct point of reference for such a significant expansion of the scope of substantive criminal law. Hallevy seems to underestimate the conceptual difficulties involved:

> 'Artificial intelligence technology has the capability to fulfill the requirements of criminal law with no single change of these definitions. [...] How come that these very definitions used by criminal law systems around the world for centuries were adequate and almost non-questionable, and suddenly they became "shallow"? If they are good enough for humans and corporations, why would not they be good enough for artificial intelligence technology?'[32]

The fundamental difference in this case is that corporations are composed of individuals, who are indirectly targeted by sanctions imposed against the corporation, whereas AI systems constitute a manifestation of human

28 Susanne Beck, 'Corporate Criminal Liability' in Markus D Dubber and Tatjana Hörnle (eds), *The Oxford Handbook of Criminal Law* (Oxford University Press 2014) 568.

29 Chesterman (n 2) 116; Kurki, 'The Legal Personhood of Artificial Intelligences' (n 2) 179.

30 Abigail H Lipman, 'Corporate Criminal Liability' (2009) 46 *American Criminal Law Review* 359, 360.

31 Chesterman (n 3) 123.

32 Hallevy (n 18) 105.

creativity.[33] In addition, it should be noted that the diverse individuals or corporations that may be at fault in cases of harm associated with the function of AI systems may not be necessarily affected by sanctions exclusively imposed against these systems. In Kelsen's terms, 'the sanction is applied to individuals, because human beings only can be the objects of the sanction, the victims of forcible deprivation of life and freedom. But the sanction is applied to them not individually, but collectively. That a sanction is directed against a juristic person means that collective responsibility is established of the individuals who are subject to the total or partial legal order personified in the concept of juristic person.'[34]

For criminal law purposes, the aggregate theory of the corporate person claims an important position; under this theory, the corporation does not possess an independent identity, but is seen as a symbol for the collaboration of individuals, who constitute the fundamental unit of interest.[35] The intense interaction and the dynamics developed between the corporate person and its members — at a structured, permanent and organised level — constitute an essential element of this theory of corporate legal personhood, which is missing in the case of AI systems.[36]

Whereas it is possible to pierce the corporate veil and punish the directing minds behind the corporate person, the clear-cut choice in most future cases would be to attribute blame either on the system itself, or on its operators.[37] This would be so because an AI system, as such, is too opaque to be referred to as a symbol for the collective efforts of the persons who have collaborated in its production and use, especially taking into account that these collaborations typically involve multiple corporations and/or individual freelance programmers. Would it be meaningful at all to impose criminal sanctions on such a type of agent? In other words, the so-called 'problem of many hands' may exist in corporate structures as well, but it is considerably exacerbated in the case at hand. Conferring criminal agency to AI systems, and rendering them the sole addressees of criminal sanctions, would allow powerful individuals holding key positions within corporate structures to avoid taking the blame for harmful outcomes. This

33 Chesterman (n 2) 124, 126.
34 Hans Kelsen, *General Theory of Law and State* (Transaction Publishers 2006) 106.
35 Beck (n 28) 576.
36 Other scholars have rightly observed that this theory cannot be readily invoked in the case of AI systems; see Chesterman (n 3) 118.
37 In this direction, see David Runciman, *The Handover: How We Gave Control of Our Lives to Corporations, States and AIs* (Profile Books 2023) 67.

has occasionally been the case with sanctions that have been exclusively directed against corporate persons.[38]

3.2 The Risk of Altering the Nature of Criminal Responsibility

In most jurisdictions, criminally responsible persons may be humans, qualifying as autonomous and reasons-responsive, and the corporations that the latter can form. In both cases, one deals with agents that can challenge the State's exclusive exercise of legislative power.[39] According to the rule of law account of legal personhood, only entities that are capable of genuinely exhibiting wrongful conduct can be regarded as subjects of criminal punishment. Corporate persons may be perceived as belonging to this category only if they are regarded as a symbol for the union of members who possess legal personhood or, in other words, as a genuine collectivity of individuals.[40]

It would not be possible to hold AI systems liable under such a high standard of criminal responsibility for two reasons. First, because this standard requires that the subject of punishment understand what criminal wrongdoing entails, the impact of their conduct on the state's exclusive authority to legislate, and its significance for the stability of the social order.[41] While modern AI systems have some computational knowledge on their status and their internal states, this is not meaningful enough for accountability mechanisms to be activated with regard to their proper conduct.[42] And second, because their conduct is jointly determined by programming policies and the autonomous development of the system following machine

38 ibid 270; Chesterman (n 3) 121; Joanna J Bryson, Mihailis E Diamantis and Thomas D Grant, 'Of, for, and by the People: The Legal Lacuna of Synthetic Persons' (2017) 25 *Artificial Intelligence and Law* 273, 285–.

39 Malcolm Thorburn, 'In Search of Criminal Law's Person', *The Criminal Law's Person* (Hart 2022) 101.

40 ibid 108.

41 ibid 115–116, 117. See, in a similar direction, Beatriz A Ribeiro and others, 'Metacognition, Accountability and Legal Personhood of AI' in Henrique Sousa Antunes and others (eds), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, vol 58 (Springer International Publishing 2024)178-179. The authors argue that AI systems cannot be held accountable because they lack strategic and monitoring capacities, which are necessary in order to understand criminal wrongdoing and reflect upon the nature of one's conduct.

42 Ribeiro and others (n 41) 180.

learning methods, which does not allow them to be considered as a genuine collectivity. Whatever emergent properties they may have, these do not originate in the interaction of individual members; and while this may not be inconsistent with the recognition of legal personhood in other domains, it is substantive for criminal law purposes.[43] Therefore, the claim that the definition of intent for criminal law purposes is narrower compared to the conceptions supported in other disciplines (e.g., in philosophy and psychology) seems to oversimplify the content and purpose of criminal personhood.[44]

The argumentation in favour of the recognition of criminal personhood for AI systems is moved by two considerations: first, by efficiency considerations, as humanising them would, allegedly, contribute in their smooth incorporation to many sectors of societal activity and in the alleviation of fears regarding the possibility of unforeseen outcomes.[45] Second, by the need to ensure at all times that accountability is maintained, especially in cases of extensive harm that are of interest to the society at large and may present evidentiary difficulties as to the identification of culprits. Both aims already exist behind the concept of corporate personhood, and this certainly favours the comparison between the two cases.[46] However, they do not suffice to set aside the crucial differences between the criminal law's person (which includes corporate persons) and modern AI systems.[47] As Paul Burgess accurately frames this issue:

> 'The motivators that have traditionally been used to hold power-wielding entities accountable have been human-centered. These may include removal from power, sanctions, fines, imprisonment, or reputational damage. There seems to be no immediate way to make a machine accountable in this way [...] On this basis, the Future Rule of Law would require the inclusion of a human-in-the-loop to *be* accountable. This is not necessary for fear of ineffective technology or malfunction, but

43 In the same direction, see Bert Jaap Koops, Mireille Hildebrandt and David-Olivier Jaquet-Chiffelle, 'Bridging the Accountability Gap: Rights for New Entities in the Information Society?' (2010) 11 *Minnesota Journal of Law, Science & Technology* 497, 560.

44 See in this respect Hallevy (n 8) 98ff.

45 Hallevy (n 14) 22; Anna Beckers and Gunther Teubner, *Three Liability Regimes for Artificial Intelligence: Algorithmic Actants, Hybrids, Crowds* (Hart 2021) 26.

46 Reyes (n 5) 1493.

47 In this direction, see Runciman (n 37) 100.

to ensure that the exercise of power has a human actor that is always accountable.'[48]

One of Hallevy's basic premises is that criminal law doctrine should not change the course of technological developments, but instead, it should adapt its basic concepts and categories to fit the latest technological developments.[49] This need is put to the forth as it becomes clear that difficulties in ensuring accountability arise in both human-on-the-loop scenarios, in which the AI system has been delegated a substantive amount of autonomous decision-making and humans can occasionally intervene after the decision has been taken, and human-out-of-the-loop scenarios, in which the possibility of meaningful human intervention is excluded.[50] Criminal law's function has in former times focused on the identification of blameworthy individuals; in the course of time, it seems to prioritise the resolution of conflicts that implicate the public at large.[51] Proponents of this new model of criminal liability seem to endorse the latter view on the role and function of criminal justice systems.

Against this background, it becomes clear that the answer to the question of transposing or not the model of corporate criminal liability to another type of artificial agent reveals a specific understanding of criminal punishment. Taking into account modern AI systems, it would be difficult to bridge the doctrinal gaps involved without affecting the nature of criminal responsibility. As it has been argued also with respect to criminal sanctions that have been imposed against corporations, the discussed regime may amount to the activation of a tool of last resort for the satisfaction of civil claims.[52] Such a development would further reinforce the tendency towards using the criminal law as an instrument for compliance goals, whereas in reality, the excessive deployment of criminal provisions renders this branch of law ineffective.[53]

48 Paul Burgess, 'The Future Rule of Law and AI: The Necessary Evolution of a Concept', *AI and the Rule of Law: The Necessary Evolution of a Concept* (Bloomsbury Publishing 2024) 171.
49 Hallevy (n 18).
50 Burgess (n 48) 157.
51 Beck (n 28) 582.
52 Thomas Weigend, 'Societas Delinquere Non Potest? A German Perspective' (2008) 6 *Journal of International Criminal Justice* 927, 944.
53 ibid 945.

While this may be so, it is also true that the criminal law, constituting the most powerful instrument of state coercion, should adapt, to a certain extent, to technological developments so as to continue to fulfil its double protective function. A balance achieved between two parameters — the protection of citizens' fundamental legal interests, on the one hand, and the protection against the arbitrariness of State power, on the other hand — would legitimate new legislative initiatives in this respect.[54] Such a balance can be achieved when new criminal law provisions aim at punishing culpable conduct that is related to technological developments without setting aside fundamental criminal law categories, such as the culpability principle.[55]

3.3 On the Morality and Intelligence of AI Systems

Admittedly, modern AI systems develop a certain degree of moral agency, as they can produce effects in their environment and these effects often have moral consequences for human interests.[56] If rationality, intentionality, and communicative capacity have been identified as essential features of personhood, then a thinking AI system could qualify as a person.[57] Such a speculative scenario has been eloquently described in Ian McEwan's novel, titled *Machines Like Me*. One of the main characters of this story is a humanoid robot named Adam, who is equipped with a sense of morality by design as well as a personality that is initially shaped by his owners but is then further developed by his interaction with the environment and other people. The book largely describes the robot's decision-making processes and practical choices, that are determined by the interaction between his inbuilt moral code and his personal inclinations.[58] It is shown that the

54 Agapios Papaneofytou, *Criminal Liability of Legal Persons: A Construction of Criminal Immunity for Corporate Offenders?* (Nomiki Vivliothiki 2012) 130.

55 ibid 137.

56 Mark Coeckelbergh, *AI Ethics* (MIT Press 2020) 50.

57 Susanna Kim Ripken, 'Philosophical Dimensions of the Corporate Person', *Corporate Personhood* (Cambridge University Press 2019) 62.

58 Ian McEwan, *Machines Like Me* (Nan A Talese/Doubleday 2019) 73. As explained by the narrator and one of the two owners of the robot, '[s]oftware engineers from the automobile industry may have helped with Adam's moral maps. But together we had contributed to his personality. I didn't know the extent to which it intruded on, or took priority over, his ethics. How deep did personality go? A perfectly formed moral system should float free of any particular disposition.'

clash of such powerful drives in self-conscious artificial agents could lead to unpredictable outcomes that may severely harm human interests.

However, the notion of AGI does not correspond to a tangible possibility as of yet. Modern AI systems rely on the recognition of probabilities, and require a massive amount of training data in order to extract patterns from such probabilities. It is true that they are extremely powerful, as they can process enormous amounts of data in minimal time frames, functioning between abstraction and enumeration. Nonetheless, at no point during their operation do these systems develop a literal understanding of the meaning behind this data.[59] In this respect, one should not overestimate the value of the knowledge produced by machine learning algorithms, as this can be proved erroneous because of false correlations, with material factors having been misevaluated, while irrelevant factors may have been considered as significant for the purposes of the system's task.

The discussion on AI systems' potential for legal personhood is, to a large extent, driven by their perceived potential to develop intelligent behaviour. In fact, intelligence should not be conceived in narrow terms: there is more than one kind of intelligence, and as many AI researchers emphasise, this is not an exclusively human feature.[60] The Turing test, introduced in 1950, relies on a benchmark that refers to human intelligence, as it assesses the system's ability to exhibit a level of performance in cognitive tasks that so resembles human performance, that the human evaluator cannot discern whether the task is executed by a human or a machine. However, it is noteworthy that this test was designed at the time with the goal of providing an operational definition of intelligence, and it is no longer referred to as an accurate measure for AI systems' intelligent conduct.[61]

At present, the registered developments in the field of AI call for a reference to more neutral definitions of intelligence. In general terms, one may neutrally define 'intelligence' as the ability to develop purposeful, goal-oriented behaviour. The ability for such behaviour is demonstrated by the adaptation to new situations and task environments, where different agents may co-exist.[62] In other words, an AI agent can be regarded as intelligent

59 Nello Cristianini, *The Shortcut: Why Intelligent Machines Do Not Think Like Us* (CRC Press 2023) 19, 36.

60 Coeckelbergh (n 56) 64; Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press 2021) 6.

61 AM Turing, 'Computing Machinery and Intelligence' (1950) LIX (236) *Mind* 433–.

62 Cristianini (n 5) 5; Beatriz A Ribeiro and others, 'Metacognition, Accountability and Legal Personhood of AI' in Henrique Sousa Antunes and others (eds), *Multidis-*

simply by means of a limited, task-oriented resilience, which could include the consideration of legal provisions during the process.[63] Direct comparisons between human and artificial agents are not necessarily fruitful. As AI agents cannot process emotions and abstract concepts, humans cannot detect patterns in data sets or perform large-scale computations; yet, applying this definition, both agents behave intelligently.[64] Different entities may set goals that are fundamentally different, both in nature and complexity. And it is noteworthy that autonomous AI systems are already capable of producing results that in many respects are beyond human capabilities and comprehension.[65]

This standard of intelligence is thus in principle met by modern, narrow AI systems, but it falls short of the high standard associated with criminal agency, which entails the ability to perform purposeful conduct as a response to the right or wrong reasons. Furthermore, any elaboration on future, fully-fledged AI agents falls against the impossibility (at least with reference to current data) of conceptualising a full artificial moral agent, that may completely diverge from the typical example of a human moral agent.[66] As David Runciman observes in this respect, '[i]t is not clear whether AGI would require anything we would recognise as consciousness to function. Perhaps it will be possible for an AI system to move across the full range of intelligent reflection without being self-reflective in anything like a human sense.'[67]

In view of the above, the next logical question is whether there are alternative measures that should be prioritised over the creation of another fiction under criminal law, or, in other words, whether the legal personhood of AI systems can be constructed in a different way, for different legal purposes. This question is concisely addressed in the following section.

---

ciplinary Perspectives on Artificial Intelligence and the Law, vol 58 (Springer International Publishing 2024) 172–173.

63 Kurki, 'The Legal Personhood of Artificial Intelligences' (n 3) 181.

64 ibid 12, 61.

65 Beckers and Teubner (n 45) 40.

66 Cristianini (n 6) 3–4.

67 Runciman (n 37) 84.

4. The Recognition of Legal Personhood for AI Systems Under Civil Law

4.1 On the Bundle Theory of Rights

In the context of this analysis, it is neither possible nor material to elaborate on the content of the many and diverse conceptions of legal personhood. It is noteworthy, though, that there is no definitive set of properties for the formulation of this definition, and this recognition may largely depend on the finding that an entity occupies an important social standing.[68] As it has been noted already, the notion of personhood is not identical for all branches of law. Therefore, defining the specific purposes that are served by such recognition in each case amounts to a fundamental and essential task.

More specifically, in order to approach the personhood of artificial entities, it should be taken into account that this may exist along a spectrum.[69] Different jurisdictions hold different conceptions of personhood, and these can be further modified in the course of historic time.[70] According to the bundle theory of legal personhood, that is advanced by Visa Kurki, an entity can qualify as a legal person regardless of its intrinsic qualities.[71] This theory places the emphasis on the entity's capacity to hold legal entitlements and burdens, and legal personhood may be granted regardless of more substantive considerations.[72] Legal personhood is conceived of as a cluster property — not a clear-cut, binary classification — and thus it is formed of many different incidents which may be combined in different ways, without any one of them being individually decisive for this recognition.[73]

---

68 Chesterman (n 3) 119; Gunkel (n 2) 19; Nadia Banteka, 'Legal Personhood and AI: AI Personhood on a Sliding Scale' in Ernest Lim and Phillip Morgan (eds), *The Cambridge Handbook of Private Law and Artificial Intelligence* (Cambridge University Press 2024) 621.

69 Ripken (n 57) 92.

70 Penny Crofts, 'Crown Resorts and the Im/Moral Corporate Form' in Elise Bant (ed), *The Culpable Corporate Mind* (Bloomsbury Publishing 2023) 61.

71 Visa AJ Kurki, 'The Incidents of Legal Personhood', *A Theory of Legal Personhood* (Oxford University Press 2019) 91.

72 Burgess (n 48) 167.

73 Kurki (n 52) 93–94, 121. As the author explains, '[...] because of the cluster-property nature of legal personhood, a binary question about X's personhood or nonpersonhood may not always be appropriate. One must often rather ask about the respects in which X is a legal person. Corporations are legal persons for most purposes, but not all purposes.' For an argumentation in favour of the recognition of personhood along a spectrum, see also Banteka (n 68). More specifically, the author argues that

These incidents of legal personhood can be either passive or active. In general, the passive incidents of legal personhood concern an entity's capacity to benefit from fundamental legal protections, the capacity to own property, to undergo harm from a legal point of view and to claim a victim status under criminal law (legal standing). Active incidents are associated with the entity's capacity to enter into contractual relations (legal competences), as well as the capacity to be subjected to civil or criminal liability (onerous legal personhood). In other words, passive legal personhood is confirmed when an entity becomes entitled to claim-rights (with the status of children constituting the main example), whereas active legal personhood is linked to the exercise of legal competences and the capacity to bear duties (as is typically the case with most adults).[74]

More specifically for the issue at hand, the notion of 'onerous legal personhood' refers to the entity's capacity to hold legal duties and to be held accountable for one's acts. A failure to fulfil such duties is penalised, and sanctions are enforced against this subject.[75] This is precisely the element that forms the essence of the criminal law's person, and it is connected to other incidents of personhood: for instance, corporations can be held liable under civil and criminal law because of their capacity to own proper funds.[76] According to the account presented here, an entity can legitimately bear legal duties provided that it is able to act and suffer detriment; additionally, it must be possible to communicate legal requirements to the entity in question.[77]

It is crucial to note that these dimensions of personhood can be realised to various degrees. In other words, an entity may be legitimately held responsible in only certain respects, or be endowed with legal personhood without being considered as directly accountable.[78]

Applying the bundle theory of legal personhood means that AI systems could be endowed with a limited, passive legal personhood, and this for reasons that have nothing to do with the system's worth, value, or reasons-

---

the scope of legal personhood should be more restricted the more autonomous the AI system actually is, so that its operators continue to monitor its safety parameters and to assume liability for any damages caused.

74 Visa AJ Kurki, 'Who or What Can Be a Legal Person?', *A Theory of Legal Personhood* (Oxford University Press 2019) 138.

75 Kurki, 'The Incidents of Legal Personhood' (n 71) 116.

76 ibid 117.

77 Kurki, 'Who or What Can Be a Legal Person?' (n 74) 146.

78 ibid 150; Ribeiro and others (n 41) 181.

responsiveness, but rather for the purpose of ensuring an effective and fair allocation of risks.[79] This is a conclusion that is wholly consistent with the considerations exposed in the previous section of this paper, regarding a potential regime under criminal law, and it should be endorsed. The author focuses on the commercial sector and conceives of AI systems as administrators or representatives of legal platforms whose objectives are essentially defined by humans.[80]

This proposal of a limited legal personhood under civil law is both feasible and sensible, and it is endorsed as appropriate to capture the type of legal personhood that could soundly incorporate AI applications to many sectors of activity, promoting the goal of a fair allocation of risks and addressing some of the challenges involved.[81] Indeed, a regime of full legal personhood for AI systems would not be appropriate, as the recognition of active personhood would mean that they would be allowed to own and freely manage their resources.[82] However, it may be observed that it is not detailed enough so as to account for the different dynamics that are developed in various contexts of their deployment, the concrete risks that are associated with different types of artificial agents and, more specifically, the hybrid nature of action that is inherent in collaborations between humans and AI systems.

4.2 The Notion of 'Machine Behaviour' and the Related Proposed Liability Regimes

In view of this need, a noteworthy typology of AI systems that is based on features of machine behaviour has been advanced in the legal field by Anna Beckers and Günther Teubner.[83] It is meant to identify and delineate distinct risks that are posed by broader categories of modern AI systems, and to pose the question of personhood in specific socio-technical and legal

79 Kurki, 'The Legal Personhood of Artificial Intelligences' (n 3) 189.

80 A legal platform is a bundle of legal positions (a set of rights and duties) that are interconnected and are usually administered by legal acts. They are attached to specific legal persons and exist only in a legal sense. For this definition, see Kurki, 'The Incidents of Legal Personhood' (n 71) 96; Kurki, 'Who or What Can Be a Legal Person?' (n 74) 135.

81 Gunkel (n 2) 156.

82 See in the same direction Beckers and Teubner (n 28) 10ff.

83 See Beckers and Teubner *supra* (n 45).

contexts.[84] These are the risk of autonomous decision-making (individual machine behaviour), the risk of association with human agents (hybrid human-machine behaviour) and, finally, the risk of systemic interconnectivity with other agents (collective machine behaviour). This distinction, at the same time, delineates the extent to which legal personhood could be recognised for certain AI systems under civil law.

This classification could be relevant for other branches of law, notably for criminal law purposes, for two main reasons. First, because a clear classification generally precludes the tendency of constructing separate regimes for each AI application. It may be that, depending on the context, a harm caused by the failure of a care robot could have been also caused by other types of autonomous AI applications, if the type of machine behaviour exhibited is essentially the same. And second, because it may contribute to shaping the duty of care of manufacturers or professional users of the system, by taking into account the degree of unpredictability that is associated with the system's conduct.[85]

Individualised software agents constitute the first type of AI agent, associated with the ability and risk presented by the development of autonomous decision-making. Examining individual machine behaviour, one may identify certain intrinsic properties of the algorithm, and the related risk is created by a single source code. Human-machine associations or 'digital hybrids' are created by the close interaction between humans and machines. As the two actors merge, the risks associated with their combined action cannot be broken down into individual contributions, potentially stemming from either side. This is what is identified as the association risk, inherent in hybrid machine behaviour. In some respects, digital hybrids' behaviour is functionally similar to the acts of corporate persons and the difficulty in ensuring accountability is particularly acute. Lastly, interconnected systems or 'multi-agent crowds' exhibit a systemwide behaviour that is formed by the interconnection of multiple AI systems in a network. The latter collectively act like swarms, and this is precisely the in-

84 A similar approach on the question of legal personhood for artificial entities is also applied by Reyes, who examines the case of autonomous corporations; see *supra* (n 4) 1466ff.

85 While it may be possible to identify fault in cases of individual machine behaviour, the same task becomes much more complicated when one is confronted with collective action that is exhibited by multiple, interconnected agents. This is an important issue that would require a separate analysis, and cannot be discussed in the context of this contribution.

terconnectivity risk that is posed by collective agents. In this particular case, the identification of liable parties becomes almost impossible. In view of the features of collective machine behaviour, the requirement of foreseeability cannot be upheld, as it would not have been possible for a human actor to intervene in order to prevent the harm caused.

Having delineated this typology, Beckers and Teubner observe that digital autonomy in the context of civil law can be defined by taking into account the system's ability for developing goal-oriented action and, at the same time, the degree of its participation in communicative processes. For this definition, the philosophical contours of personhood — that base this conferral on the entity's reasons-responsiveness and capacity of freedom — are not authoritative. In their analysis, the defining feature for the recognition of personhood is not the ability for consciousness but for communication.[86] Indeed, it can be observed that their definition of autonomy chimes with the rationales of civil liability, i.e., with the need to ensure a fair distribution of risks, a fair compensation for damages, and the reduction of financial loss.

Accordingly, the authors propose three distinct liability regimes for autonomous AI systems. Firstly, they analyse the potential legal regime that could apply to AI systems that qualify as 'individual actants'. Such systems are often used in the framework of 'digital assistance', that is in order to conclude contracts in representation of a human principal or organisation (smart contracts).[87] Because the algorithm in this case acts as a representative and performs legally binding action, which on occasion may cause damage, and precisely in view of the fact that the system can autonomously undertake action that exceeds its principal's intentions, a regime of limited legal personhood constitutes, in their opinion, the most suitable option. The rules of vicarious liability would apply in this case, so these AI systems would be regarded as vicarious agents.[88] The relevant rules can provide for a reasonable allocations of risks: the principal must assume the risk of hacking because they appear as the party that controls the digital agent, unless it is obvious that the digital agent has produced a result that exceeds

86 Beckers and Teubner (n 45) 58.

87 The term 'digital assistance' is defined as the social relation between human users and digital agents.

88 ibid 48, 55.

the delegated decision-making power or the contracting party could have reasonably known that the digital agent was hacked.[89]

The second case relates to the potential legal regime that could apply for digital hybrids. The authors argue that a separate legal capacity should be attributed to this new type of collective actor, that can be regarded as a network. As in this context it is no longer possible to distinguish among individual contributions, these actors must be treated as hybrids precisely in view of the emergent collective properties that appear in their combined input.[90] They can be compared to corporate actors for two reasons: first, because they do not act on their own behalf, but for the association, and second, because there could be cases of conflict of interest involving the members and the association.[91] However, this comparison remains approximative because corporate actors are characterised by more clearly-structured relations. Therefore, digital hybrids cannot be perceived as genuine collective actors. The hybrid is identified as the source of the harmful action and, subsequently, liability is distributed among the members who control or benefit from its activity (i.e., the operators, owners, manufacturers and deliverers).[92] The victim's claim is facilitated because they only need to prove that the damage was caused by the hybrid.

The third and last regime examined is that of multi-agent crowds. In this case, it is not possible to refer to either individual or collective decision-making and action, or to any kind of purpose, which means that multi-agent crowds cannot be vested with any type of personhood.[93] The interconnectivity risk that appears in this constellation relates to the impossibility of either foreseeing or explaining its outcomes. For instance, the Flash Crash of 2010 was probably produced by such collective machine behaviour, and in general no individual liability can be discerned in similar cases of malfunctions in high-frequency trading.[94] There is a growing consensus on the opinion that the operation of inter-connected systems does not leave room for the clarification of causality issues and the identification of blameworthy individuals.

89 ibid 64.
90 ibid 90.
91 ibid 94–95.
92 ibid 105.
93 ibid 111–112.
94 Orlando Cosme Jr, 'Regulating High-Frequency Trading: The Case for Individual Criminal Liability' (2019) 109 *Journal of Criminal Law and Criminology* 365.

In view of this state of affairs, the solution lies in the establishment of risk pools that would provide compensation for damages sustained through adequate funds. In the authors' view, this is a measure to be prioritised because, strictly speaking, it is not possible to comprehend or effectively control interconnected machine operations. This resort to sectoral fund-based schemes is already known from other fields of application, in which extended damages have been caused by complex factors. This is the case with environmental disasters and financial crises, both exacerbated by the activities of corporate actors. For instance, the 'Superfund programme', developed in the US during the 1980s as a federal response for the prevention and mitigation of environmental harm, provides an interesting model for consideration.[95] This regime involves strict, joint and several liability for potentially responsible parties (site owners, facility operators, transporters or anyone who generates hazardous substances), as well as a federal trust fund in case no such parties can be identified. In this context, the observance of typical practices within the relevant field cannot exempt a party from liability.[96] In order to reasonably limit the extent of the damages covered, the fund-based regime discussed by the authors would have to remain a subsidiary remedy, activated, for instance, in cases of loss that are not covered by private insurance schemes.

In conclusion, it can be noted that even though the nuanced approach presented by Beckers' and Teubner's scheme accounts for the dangers that are inherent in autonomous AI systems and chimes with the specific aims of civil liability, it cannot as such be authoritative for the recognition of criminal personhood and the establishment of a potential criminal liability against them.[97] The authors themselves somehow allude to a similar

95 Alexander E Farrell, 'Overview of the Superfund Program' in Gregg P Macey and Jonathan Z Cannon (eds), *Reclaiming the Land - Rethinking Superfund Institutions, Methods and Practices* (Springer) 25-26. This programme was launched by the Comprehensive Environmental Response, Compensation, and Liability Act in 1980 and the Superfund Amendments and Reauthorization Act in 1986. Along with other legislative acts, it aims at providing a structured response to the risk of harm that is created by hazardous substances and compounds, a risk posed both for human health and the environment.

96 ibid 41.

97 The nuanced nature of this approach is also praised by Gunkel (n 2) 157, even though ultimately the author does not espouse it for referring to the institution of robot slavery. However, it is noteworty that Beckers and Teubner do not directly refer to the relevant arguments that exist in the literature, and they simply recognise that the legal status of AI systems can be reasonably compared to the provisions regarding slavery

observation, when they note that 'each social system creates for algorithms its own criteria of personhood.'[98] That having been noted, however, any proposition that entails the recognition of even a limited legal personhood to an artificial entity should be examined with caution, as experience has shown that the contours of such personhood typically expand in the course of time.[99]

5. Concluding Remarks

This paper intends to illustrate that while both corporations and AI systems constitute powerful and highly opaque agents that may be delegated extensive decision-making powers in a variety of social contexts, their comparison under the lens of substantive criminal law may seem reasonable at first, but it can be neither self-evident nor absolute. Whereas the recognition of criminal personhood for certain autonomous AI systems is now presented as a plausible and efficient legal development, AI systems cannot be regarded as genuine collectivities, and thus cannot be meaningfully included within the class of criminally responsible subjects. While this may be so, the prospect of a limited legal personhood for civil law purposes — introduced with the goal of ensuring a fair allocation of risks and economically efficient mechanisms — may be endorsed without particular doctrinal problems. If AI systems ever come to develop a kind of general agency that would render direct accountability mechanisms meaningful, a different approach may have to be advanced, but in that case the primary difficulty would not be legal but technical, as merely understanding AGI would already pose an enormous challenge.

under Roman law. This statement is not, in itself, sufficient to justify the rejection of their conceptual framework.

98 Beckers and Teubner (n 45) 147.

99 Chesterman (n 3) 141.

In AI we trust – The use of artificial intelligence tools for payment fraud detection

Grazia Bruzzese

Abstract: Notwithstanding the various denominations it gained throughout time and its various subsets, it can reasonably be argued that fraud is a phenomenon present since the beginning of History. From the Roman Laws of the Twelve Tables which already covered the action of *fraus*, to its more modern and famous forms, like Ponzi schemes, fraud is recognized and sanctioned by the various legal systems throughout the globe. Yet, despite it being well-known since the Antiquity, the different forms that fraud can take are definitively not set in stone.

New technologies, like Artificial Intelligence, have not only touched upon all the sectors but also changed the traditional ways we deal with daily businesses. More specifically in the context of fraud, the development of a new subset of varieties like digital payment frauds can be witnessed. Nonetheless, new challenges go often hand in hand with new opportunities. As a result, if the use of new technologies has certainly inspired criminal minds to develop new illicit actions, it also opened the doors to new ways of detecting and contrasting them.

This contribution is intended to provide a short study of a specific type of digital financial fraud, namely digital payment fraud, by presenting some of the existing Artificial Intelligence tools that are currently used for the purpose of its detection. By relying on publicly available data, it will present and evaluate some of the strengths and weaknesses of these tools all while reflecting upon further avenues for improvement in the use of such technologies

Introduction. From bottomry to phishing: 2500 years of fraud – 1. Defining fraud in the digital age era: Setting the scene for AI-based countermeasures – 2. A spectre is haunting banking services: digital payment fraud – 2.1. Defining the subject – 2.2. On the various shades of digital payment fraud – 3. Beyond human eyes: AI's fight against digital payment fraud – 3.1. Every rose has its thorn: The Double-Edged Sword of Machine Learning in Fraud Detection – 3.2. Sharing is caring? – Conclusion

Introduction – From bottomry to phishing: 2500 years of fraud

"No sooner has the name OEDIPUS passed his lips, than his voice is drowned in a shout of execration. They call upon him to leave Attica

instantly. He won their promise by a fraud, and it is void"[1], wrote Sir Richard C. Jebb in his commentary of *Oedipus at Colonus*. This remarkable comment of the Sophoclean tragedy witnesses the opprobrium that the inhabitants of Colonus showed toward Oedipus, initially defrauding them by hiding his identity. In addition to reflecting a sentiment that has persisted for over two millennia in response to fraudulent behaviour, this passage highlights the enduring nature of fraud as a social phenomenon that continues to pose significant challenges in contemporary societies. Undoubtedly, fraud constitutes a legally complex and polymorphic phenomenon, characterized by its protean nature and resistance to precise categorization within a singular doctrinal framework. Although it has existed throughout history—arguably intrinsic to human conduct—fraud distinguishes itself from other offences by its capacity to evolve in form and function across different legal systems and historical periods. Unlike many other criminal offences, it continually adapts to the socio-economic and cultural contexts in which it occurs.[2] As such, it must be understood not only as a breach of legal norms but also as a socio-legal construct that reflects the prevailing values and normative expectations of the society and the era in which it occurs.

Fraud has been present since antiquity, if not before, with one of the earliest documented cases dating back to approximately 300 B.C. This recorded case involved a Greek merchant, by the name of Hegestratos, who attempted to perpetrate an insurance fraud through a maritime loan arrangement known as *bottomry*.[3] Under this mechanism, the merchant

1 Richard C Jebb, "Parodos: 117–253", in *Commentary on Sophocles: Oedipus at Colonus,* 1899.

2 Cf. for an overview on the evolution of fraud: Aurora AC Teixeira, António Maia, José António Moreira, Carlos Pimenta (eds), *Interdisciplinary Insights on Fraud* (Cambridge Scholars Publishing 2014); Martin T Biegelman, *Faces of Fraud: Cases and Lessons from a Life of Fighting Fraudsters* (Wiley 2013). Sridhar Ramamoorti, David E Morrison III, Joseph W Koletar, Kelly R Pope, *A. B. C. 's of Behavioral Forensics: Applying Psychology to Financial Fraud Prevention and Detection* (Wiley 2013). Elisabeth B Haarington, 'The Sociology of Financial Fraud', in Karin K Cetina, Alex Preda (eds), *The Oxford Handbook of The Sociology of Finance* (Oxford University Press 2012), 393–410. J.T. Wells, *International Fraud Handbook* (Wiley 2018).

3 Cf. Christian Chavagneux, *Les plus belles histoires de l'escroquerie* (Éditions du Seuil 2020) 1–17. Interestingly enough, bottomry seems to have persisted with regard to maritime business to the point where some cases were to be found in the XVIIIth century. Cf. Maria Fusaro, Andrea Addobbati, Luisa Piccinno (eds.), *General Average and Risk Management in Medieval and Early Modern Maritime Business* (Palgrave Macmillan 2023).

secured a loan by insuring a shipment of a given good. Yet, instead of fulfilling the terms of the agreement by delivering the good, he would set sail without the insured cargo (his plan further involved selling the corn elsewhere for additional gain), intending to deliberately scuttle the vessel and thereby retain the loan proceeds. Indeed, the insurance consisted in the fact that if the ship sank, the loan did not need to be paid back.[4] Fraud's capacity to permeate all levels of society is evident across history. As a matter of fact, it is capable of reaching and affecting even the most illustrious and powerful people in a given society as illustrated in the very famous 'affair of the Diamond Necklace' which involved and tarnished the reputations of Cardinal de Rohan and Queen Marie Antoinette.[5] Affecting the private finances of poor small investors – like it is often the case when it comes to Ponzi schemes – or even public funds, it is sufficient to think of fake president frauds against public institutions or non-governmental organisations partially or entirely funded by governments, like the recent fraud scandal against Caritas Luxembourg.[6]

Yet, despite the apparent heterogeneity of these cases – ranging from insurance fraud to political scandals and complex financial schemes – they are all subsumed under the legal category of fraud. This convergence of disparate cases under the umbrella term of fraud not only suggests the adaptability and definitional boundaries of fraud but also invites reflection on the elasticity and new configurations of fraud as a legal construct within contemporary legal systems. To this end, this book chapter will proceed in three parts. The first part offers a conceptual overview of fraud, tracing its historical evolution all while providing the necessary foundation for understanding how technological changes have reshaped it. The discussion subsequently turns to the financial sector, examining the phenomenon of digital payment fraud. The role of Artificial Intelligence (hereafter 'AI') in combating these practices will be assessed, focusing on both its potential and its inherent risks. The chapter concludes by drawing together these elements to reflect on the broader implications of AI fraud detection systems for the stability and integrity of financial systems.

4 Cf. Matthew Campbell, Kit Chellel, *Dead in the Water. Murder and Fraud in the World's Most Secretive Industry* (Atlantic Books 2022).

5 Cf. for more details on this affair: Chavagneux (n 3) 21 – 92.

6 On the Caritas case cf.: Véronique Poujol, 'Anatomie d'une affaire stupéfiante', in *Reporter.lu* (19 August 2024); Luc Caregari, 'Die globalen Folgen des Caritas-Skandals', in *Reporter.lu* (26 September 2024).

1. Defining fraud in the digital age era: Setting the scene for AI-based countermeasures

Given the breadth of its manifestations and typologies, the term *fraud* frequently functions as an umbrella concept encompassing a wide array of conducts that fall within the general ambit of fraudulent offences—ranging from payment fraud, romance scams, and Ponzi schemes to embezzlement, identity theft, and tax fraud, to cite just a few. Nonetheless, all these types of fraud, although committed through different operational mechanisms and with different means, share a set of common elements. Central among all these elements is the final goal namely, to obtain an advantage, in many cases, a pecuniary one – whether money or a valuable tangible good. In other words, it seems evident that fraud is intrinsically linked to money and financial gain. As a result, the latter is an offence pertaining to the realm of financial and economic criminal law. Apart from the expected material gain, fraud is characterised by the deployment of deceitful practices aimed at tricking and manipulating an individual into doing what the fraudster expects him or her to do, like proceeding to a payment or, in the context of phishing, clicking a malicious link that will compromise personal data. Hence, to say it in J. Scharfman's words, fraud is nothing else than a set of

> "multifarious means which human ingenuity can devise, and which are resorted to by one individual to get an advantage over another by false suggestions or suppression of the truth. It includes all surprises, tricks, cunning or dissembling, and any unfair way which another is cheated. In its most basic form, a fraud can be thought of as a deception resulting in a loss. Fraud can occur anywhere and, in any amount, ranging from theft of a few dollars in the daily lives of consumers to complex corporate scams resulting in the loss of billions in cryptocurrency".[7]

To sum it up, and for the sake of clarity, fraud should be understood as a series of deceptive or misleading actions, carried out with the intent to procure a benefit or an advantage which the perpetrator would not otherwise be entitled to obtain.

To this already rich and complex framework, we must add a major historical and scientific revolution that has profoundly reshaped our societies, namely the development and proliferation of new technologies, frequently

7 Jason Scharfman, *The Cryptocurrency and Digital Asset Fraud Casebook* (Springer International Publishing AG 2023) 2.

referred to as the "Third Industrial Revolution".[8] It is undeniable that these new technologies, going from the rise of the internet to the phenomenon of AI, have altered the traditional modes of human activity, permeating all sectors of personal, professional and institutional life, to the point in which it is difficult to find one that remains unbothered. Given this pervasive impact, the fact that both legal and illegal activities would have been impacted should not come as a surprise. As a result, actors operating in the illegal world did not hesitate to take advantage of the new methods and schemes offered by new technologies. More specifically with regard to fraud, new technologies allowed for what some authors define "a new way to commit an old crime".[9] Indeed, as already proven, the offence of fraud is a polymorphous one, which did not hesitate to adapt and adopt new means like the internet or AI, giving rise to new forms of fraud that can be categorized under the umbrella term of "digital fraud" or "online fraud".[10] A new form of fraud that will be characterized by the fact that it is committed through the use of technological and digital means, whether simple fraudulent mails in the context of phishing, or the use of more sophisticated solutions relying on AI, like deepfakes.[11]

---

8 Giovanni Dosi, Louis Galambos, Alfonso Gambardella and Luigi Orsenigo (eds), *The Third Industrial Revolution in Global Business* (Cambridge University Press 2013); Manuel Castells, *The Information Age: Economy, Society and Culture* (Oxford: Blackwell 1996).

9 Susan W Brenner, 'Cybercrime Metrics: Old Wine, New Bottles?' (2004) 9 *Virginia Journal of Law & Technology* 13.

10 For a definition of "digital" or "online" fraud, cf. EUROPOL's Spotlight on Online Fraud Schemes: A Web of Deceit, according to which: *"Online fraud schemes (OFSs) comprise a wide range of criminal activities that are exclusively or primarily perpetrated online or with the use of computers; we call this cyber-enabled. […] Although online fraud specifically occurs online, such schemes can comprise both online and offline operations"*. In the context of our research, we will adopt the commonly acknowledged definition of digital fraud in the sense that we will refer to it whenever a fraud is committed in the online realm. See in *Europol Spotlight, Online Fraud Schemes: A Web of Deceit* (Europol 2023) 5, available at https://www.europol.europa.eu/cms/sites/default/files/documents/Spotlight-Report_Online-fraud-schemes.pdf (accessed on 2 March 2026).

11 To give a concrete example on the use of deepfakes in fraudulent schemes, reference may be made to a sophisticated fraud that took place in Hong Kong in 2024. In that case, one of the employees of a large multinational corporation was deceived into transferring funds exceeding $25 million. The deception was executed through a videoconference in which the employee believed he was interacting with colleagues and the company's chief financial officer; in reality, each of the participants had been digitally fabricated using deepfake technology. For more details on this case, cf. Kath-

From phishing to the use of deepfakes to deceive the victims all while hiding their identity, criminals did not fail to prove creativity in coming up with new means to commit fraud. Furthermore, not only did technology open the doors to new forms of fraud but, as pointed out by Interpol in its latest Global Financial Fraud assessment, it "is enabling organized crime groups to better target victims around the world".[12] A result that should not surprise considering the fact that we live in an increasing digitalised and interconnected world, in which, especially when it comes to finance, digital solutions are more and more taking over traditional means. It is sufficient to think about the way digital banking has revolutionised the way we interact with banks, from cashless and cardless payments, to the use of sophisticated financial robo-advisers,[13] a considerable number of traditional banking services can now be effectuated without any direct human interaction between the bank and the customer. The facilitation in defrauding victims using new technologies seems also to be the reason of the exponential worldwide rise in fraud cases.

To give some examples from the European Union (hereafter 'EU'), the last report by the European Anti-Fraud Office ('OLAF') has shown that over one billion euros (1.043.800.000 €) of EU funds were lost to irregularities, including fraud.[14] A sum that should be frightening if one considers the fact that OLAF's interests only revolve around the EU funds and not the ones of the various Member States. With regard to some of these Member States, the latest figures are not comforting either. In Italy for instance, 3.360 cases of internet fraud, more specifically investment scams, were reported only in 2023, amounting to 109.536.088 euros.[15] This same raise in

leen Magramo, 'British engineering giant Arup revealed as $25 million deepfake scam victim', *CNN* (17 May 2024) https://edition.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk accessed 24 July 2025.

12 Interpol, *Global Financial Fraud assessment* (May 2024), available at https://www.interpol.int/en/News-and-Events/News/2024/INTERPOL-Financial-Fraud-assessment-A-global-threat-boosted-by-technology accessed 13 March 2025.

13 Dan Tammas-Hastings, *Liability-Driven Investment: From Analogue to Digital, Pensions to Robo-Advice*,(Wiley 2021).

14 The European Anti-Fraud Office (OLAF), *OLAF Report 2023* (European Commission 2023). The report can be found at the following link: https://ec.europa.eu/olaf-report/2023/index_en.html. accessed 10 December 2024.

15 Ministero dell'Interno, *Resoconto attività 2023 della Polizia Postale e delle Comunicazioni e dei Centri Operativi Sicurezza Cibernetica* (2 January 2024) the report can be downloaded at the following link: https://www.interno.gov.it/it/notizie/report-2023-dellattivita-polizia-postale-nel-contrasto-crimini-informatici#:~:text=9.433%20quel-

fraud cases was witnessed in Luxembourg, where according to the police report for the year 2023, 6143 cases of fraud (*escroqueries/tromperies*) were reported, 1454 more than for the previous year.[16] France values the annual total amount of euros lost to fraud in 2023 to almost 1.2 billion euros.[17] Furthermore, according to the European Banking Authority ('EBA') and European Central Bank ('ECB') joint report, 2 billion euros only in the first half of 2023 were lost to payment fraud.[18] On the other side of the ocean the figures with regard to the money lost to fraud are not less impressive than the European ones. According to the latest data from the USA's Federal Trade Commission, the fraud losses reported in 2023 totalled more than $10 billion.[19]

Confronted with these exponential figures and having set the scene, it is now time to turn to a very specific type of fraud inherent to banking services, namely payment fraud, all while assessing how banks are trying to prevent and detect it through to the use of AI tools.

---

li%20commessi%20nel%202023,31%20casi%20di%20codice%20rosso. accessed 12 February 2025.

16 Police Lëtzebuerg, *Présentation des statistiques policières de la criminalité 2023*. The statistics can be downloaded at the following link: https://police.public.lu/fr/publications/2024/chiffres-de-la-delinquance.html. accessed 12 February 2025.

17 Banque de France, *Rapport de l'Observatoire de la sécurité des moyens de paiement 2023* (10 September 2024). The report can be found at the following link : https://www.banque-france.fr/fr/publications-et-statistiques/publications/rapport-de-lobservatoire-de-la-securite-des-moyens-de-paiement-2023#:~:text=Rapport%20de%20l'Observatoire%20de%20la%20s%C3%A9curit%C3%A9%20des%20moyens%20de%20paiement%202023,-Mise%20en%20ligne&text=de%20dispositifs%20de%20blocage%20ou,%25%20par%20rapport%20%C3%A0%202022. accessed 30 March 2025.

18 European Banking Authority, *The EBA and ECB, 2024 Report on payment fraud* (1st August 2024) The report can be downloaded at the following link: https://www.eba.europa.eu/publications-and-media/press-releases/eba-and-ecb-release-joint-report-payment-fraud accessed 20 March 2025.

19 Federal Trade Commission, 'As Nationwide Fraud Losses Top $10 Billion in 2023, FTC Steps Up Efforts to Protect the Public. Investment scams lead in reported losses at more than $4.6 billion' (9 February 2024) https://www.ftc.gov/news-events/news/press-releases/2024/02/nationwide-fraud-losses-top-10-billion-2023-ftc-steps-efforts-protect-public accessed 30 March 2025.

2. A spectre is haunting banking services: digital payment fraud

We started this contribution by highlighting the multifaceted nature of fraud, including its continually evolving manifestations in the digital domain. Simultaneously, we stressed out one of the essential material elements of this offence, namely, the fact that the fraudster will illicitly try to gain a benefit, typically of a pecuniary nature. If one equates all these elements, it should not come as a surprise that one of the most prevalent and economically fruitful and consequential forms of fraud, as measured by financial losses incurred, is digital payment fraud.[20]

2.1. Defining the subject

As its designation already implies, *payment fraud* is a form of fraud perpetrated within the context of a *payment transaction*, as defined under the Payment Service Directive (hereafter 'PSD2').[21] According to the PSD2, a payment transaction refers to "an act, initiated by the payer or on his behalf or by the payee, of placing, transferring or withdrawing funds, irrespective of any underlying obligations between the payer and the payee".[22] In other words, it is the simple fact for a payer to remit funds to a payee, *e.g.*, a customer paying a seller or service provider for a good or service s/he receives. Here again, while such transactions have existed since the beginning of time,[23] the advent of new technological infrastructures has transformed the means and modalities through which they are executed. Increasingly, these payment transfers are conducted via digital or cashless means, whether by

20 Cf. for some concrete data and figures: EBA and ECB, 2024 Report on payment fraud (1st August 2024). The report can be downloaded at the following link: https://www.eba.europa.eu/publications-and-media/press-releases/eba-and-ecb-release-joint-report-payment-fraud. accessed 30 March 2025.

21 Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC [2015] OJ L 337/35.

22 Article 4(5) of the PSD2.

23 Cf. for a historical study of this subject: Jack Weatherford, *The History of Money* (Crown Currency 1998); Niall Ferguson, *Der Aufstieg des Geldes: Die harte Währung der Geschichte* (Econ 2009); Charles Eisenstein, *Sacred Economics: Money, Gift, and Society in the Age of Transition* (North Atlantic Books 2011); Viviana A Zelizer, *The Social Meaning of Money* (Basic Books 1995).

transferring credit through online payment orders or by simply using a credit card.

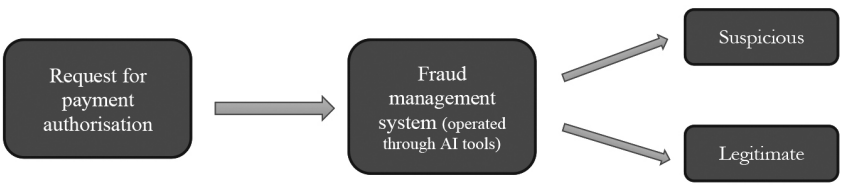
Although the execution of a digital or cashless payment appears to be a seamless and routine process, it nonetheless comprises a set of intricate and largely imperceptible procedural steps invisible to the eyes of the payer or of the payee. To give a simple illustration of how a cashless payment works, consider the common scenario in which a customer seeks to purchase a good, whether in a shop, by paying with a payment card, or online. To proceed with the payment, s/he will present the payment instrument, either the physical card or its digitalized version embedded within, for instance, its smartphone, and initiate the transaction by inserting it, or even more simple, just swiping it on a reader machine. A similar procedural framework can be transposed to the scenario where the transaction is executed via an online platform. In this scenario, the payer inserts the requisite payment credentials–typically pertaining to its card or bank account- into the designated fields on the digital interface, thereby initiating the transaction process.

This seemingly instantaneous act sets in motion a rather complex process made up of various steps. To present it in simple words, at the outset, a request for authorization to payment coming from the payee is transmitted to the payer's card issuer. The issuer will then analyze the request to assess whether the payment transaction shall be authorized or not. Once a decision authorizing or refusing the transaction has been made, it is then communicated and transmitted back to the payee.[24]

What happens in the process of deciding whether authorizing the transaction? This evaluative process involves multiple criteria. The card issuer will verify whether the payer truly disposes of the sufficient funds that s/he wants to spend, or in the context of credit instruments, whether the requested amount falls within the authorized credit limits. However, of relevance to the domain of payment fraud is the issuer's reliance on

24 Cf. for a more detailed presentation describing the engineering behind digital payments and its challenges: Nick F Ryman-Tubb, Paul Krause, Wolfgang Garn, 'How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark' (2018) 76 *Engineering Applications of Artificial Intelligence*, 130–157; Paolo Vanini et al., 'Online payment fraud: from anomaly detection to risk management' (2023) 9(1) *Financial Innovation*, 66–91; David S Evans and Richard Schmalensee, *Paying with Plastic: The Digital Revolution in Buying and Borrowing* (MIT Press 2004); Likhit Mada, 'The Rise of Mobile Payment Systems, Digital Wallets: Successes, Security and Challenges' (2025) 13(10) *European Journal of Computer Science and Information Technology* 137–152.

fraud detection and prevention mechanisms. Consequently, the decision on whether to authorize or not the transaction, will depend on the result of the so-called *fraud management system*. As the name already suggests, the aim of this control system is to verify that the transaction at hand is not an attempt to defraud the payer. To this end, the system is aimed at identifying anomalies in the transaction or indicators of potential fraudulent behavior, therefore safeguarding the transaction against unauthorized or deceitful activity.



Indeed, as shown by recent data,[25] out of the many risks to which financial and banking institutions are exposed to fraud remains one of the most pervasive and persistent ones. Consequently, an inquiry arises: in what manner does fraud manifest within this digital era context and affect the broader context of digital payments?

2.2. On the various shades of digital payment fraud

According to a recent study published by the Basel Committee on Banking Supervision[26], instances of digital payment fraud within the banking sector are usually categorised into four principal typologies. Considering the rise in online banking and online payment mechanisms, it should not come as a surprise that fraud related to unauthorized online payment transactions is

25 Cf. European Payments Council, *2024 Payment Threats and Fraud Trends Report* (22 November 2024) https://www.europeanpaymentscouncil.eu/sites/default/files/kb/file/202412/EPC16224%20v1.0%202024%20Payments%20Threats%20and%20Fraud%20Trends%20Report_0.pdf accessed 17 February 2025. EBA and ECB, *2024 Report on Payment Fraud* (1st August 2024) https://www.eba.europa.eu/publications-and-media/press-releases/eba-and-ecb-release-joint-report-payment-fraud accessed 10 February 2025.

26 Basel Committee on Banking Supervision, 'Discussion Paper Digital fraud and banking: supervisory and financial stability implications' (November 2023) https://www.bis.org/bcbs/publ/d558.htm accessed 30 March 2025.

one of the most common types of digital payment fraud. In this scenario, fraudsters unlawfully acquire customer data and, through its misappropriation, manage to access their online banking account or credit card credentials. A paradigmatic illustration of this form of fraud is the phenomenon of 'email phishing', in which fraudsters try, by emails, to make their victims give away sensitive information, for instance, credit card credentials, under the mistaken belief that the communication originates from a legitimate source, like their bank.

A second typology of digital payment fraud identified by the Basel Committee, is the one that arises when a payer is deceitfully manipulated into issuing a payment order under the false belief that the account belongs to a legitimate payee. In contrast to the first type of digital payment fraud, in this case the payer proceeds herself or himself to the payment order. In fact, this type of fraud is often perpetrated through the impersonation of a trusted third party, deceiving the payer into proceeding with the payment. As an example, one could think of the *CEO fraud* or *Fake President fraud*, in which the fraudster impersonates the CEO of a given company who gives instructions to one of the employees to proceed to a payment order which the employee will consider as legitimate.

Thirdly, digital payment fraud can also target bank customers by misleading them into investing in fictitious saving products, exploiting their trust in the bank. Such fraudulent schemes typically involve the presentation of false investment opportunities that appear to be legitimate banking products. We are referring to the notorious investment fraud or scams, in which fraudsters manage to induce victims to invest in non-existent investments. To this end, fraudsters will likely clone the website of legitimate firms, or simply promise the best possible returns, which are higher than the ones to be expected by investing in any other type of financial product, a technique which is at the very heart of Ponzi schemes and which nowadays seems to be more and more perpetrated through social media.[27]

Finally, digital payment fraud can also directly target banks themselves. In this case, fraudsters may provide the bank with false identities, or stolen ones, to open accounts or to proceed to payment orders. As a matter of

27 Akila Quinio, 'Advisers worry as social media investment scams urge', *Financial* Times (17 May 2024) https://www.ft.com/content/6f5055fc-08f6-4d83-9b66-414f7f09971e accessed 4 April 2025.

fact, there seem to be dedicated marketplaces on the dark web selling and purchasing payment card data and online banking access.[28]

Faced with challenges stemming from the use of new technologies not only by their customers, but, also, by fraudsters, financial and banking institutions could neither turn a blind eye on this phenomenon, nor rely upon obsolete and analogue-based safeguard solutions. Clearly, no one would nor could argue for a return to the complete previous analogue banking system, digital banking being nowadays too enshrined in our society. Similarly, there is no way to circumvent the use of new technologies which are going to increasingly dominate our daily activities and lives.[29] Hence, confronted with what may be characterized as an existential imperative to "adapt or cease to operate", the banking sector recognized the inadequacy of conventional fraud-detection methodologies[30] in addressing the evolving typologies of digital fraud.[31] Accordingly, the implementation of advanced

28 Cf. Basel Committee on Banking Supervision, Discussion Paper (n 26).

29 Cf. on this subject: Klaus Schwab, *The Fourth Industrial Revolution* (Clown Currency 2017); Felix Dodds et al., *Tomorrow's People and New Technology: Changing How We Live Our Lives* (Routledge 2022); Kevin Kelly, *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future* (Viking 2016); Kai-Fu Lee and Chen Qiufan, *AI 2041: Ten Visions for Our Future* (WH Allen 2024).

30 For the sake of clarity, it is worth briefly specifying what we understand under the "old forms of fraud detection". Before the raise of AI tools especially those based on ML, the vast majority of fraud detection tools were run by a so-called "rule-based approach". Under this approach, it was necessary, as the name already suggests, to create in the sense of writing rules that would then be used to detect actions that deviated from these rules. To give a very banal example of the form that a written rule could take, one could decide that proceeding to a given number of transactions in a day should be seen as suspicious. Or, even more discriminatory, one could write a rule saying that payments directed to a given county should be flagged as suspicious and require further investigation. One of the main issues with the rule-based approach, despite the need for experts on the matter to write these rules, was the almost impossibility to rapidly adapt and keep up with the evolving types of fraud.

31 Recent data seem to prove the new trend toward the adoption of AI tools to prevent and fight fraud. Cf. for some national data on Luxembourg, Commission de Surveillance du Secteur Financier and Banque centrale du Luxembourg, (CSSF & BCL) *Thematic review on the use of Artificial Intelligence in the Luxembourg financial sector* (May 2025) https://www.cssf.lu/en/Document/thematic-review-on-the-use-of-artificial-intelligence-in-the-luxembourg-financial-sector-2/ accessed 2 March 2026. For a more global study on the use of AI tools to prevent and fight financial crimes, including fraud, cf. Napier AI, *AI/AML Index 2024–2025*, https://www.napier.ai/ai-aml-index accessed 2 March 2026. Cf. also for a study on the risks that come with an overuse of AI tools for crime prevention, Serhiy Lyeonov, Larysa Hrytsenko et al.,

new technologies to mitigate–if not entirely eradicate–the attendant risks seems to have become a matter of operational necessity rather than strategic discretion. How to proceed then, if not by contrasting fraudsters through their own arms?

3. Beyond human eyes: AI's fight against digital payment fraud

"Pour lutter contre l'abstraction, il faut un peu lui ressembler"[32], wrote Albert Camus. Behind this aphorism lies a principle readily transposable to detecting digital payment fraud: effective countermeasures require not only an understanding of the adversary, but, in certain respects, also the replication if not the partial adoption of its operational patterns. Hence, considering the shift of fraud to the virtual domain, whether in relation to the commission of the offence itself (e.g. digital payment fraud) or the means employed in its commission (e.g. using deepfakes), the fact that financial and banking actors are leveraging on AI to detect and contrast it, should not come as a surprise. As a matter of fact, it seems established that within the banking and financial sector one of the most promising application areas for AI and Machine Learning (hereafter 'ML') tools is the domain of fraud management.[33] Before examining some concrete examples of AI-based tools employed in payment fraud detection, it is necessary to establish some preliminary terminological clarifications to ensure conceptual precision in the ensuing analysis.

Hence, it is worth specifying what should be understood under the terms 'AI' and 'ML'. These two notions were masterfully synthesized, through the metaphor of a child playing with Mega Blocks, by author Maris Sekar, who offers the following characterization:

> "Artificial intelligence and machine learning are used interchangeably. It is essential to clarify what they mean. Imagine a child playing with Mega Blocks. The child makes decisions on what block goes on top of

'Who benefits from AI in money laundering in Europe: the organised criminals or the AML services?' (2025) 21(1) *Human Technology*, 222–245.

32 Albert Camus, *La Peste* (Gallimard 1947).

33 Cristina Soviany, 'The benefits of using artificial intelligence in payment fraud detection: A case study' (2018) 12(2) *Journal of Payments Strategy & Systems* 103. Cf. also for more concrete data regarding Luxembourg supporting and confirming this trend in the 'CSSF thematic review on the use of Artificial Intelligence in the Luxembourg Financial Sector' (May 2025) (n 31).

> another block. The decisions are driven through past experiences (or learning). For instance, the child knows square blocks cannot go on top of the triangle blocks after trying it several times. Here, the child's decision-making mechanism to decide which block goes next can be modeled by machine learning. If the child is replaced with a robot, then the robot as a whole becomes what is known as an artificial intelligence agent. AI consists of programming a machine to do tasks that humans typically carry out.
>
> Machine learning is divided into two main categories – *supervised learning* and *unsupervised learning*. In supervised learning, the labeled inputs and expected output of the process being modeled are used to train the model. In contrast, unsupervised learning involves the use of self-learning techniques for modeling an unlabeled dataset. In our example with the child and the Mega Blocks, supervised learning is when the child learns from their parent, showing them how to use the blocks to build a tower. Unsupervised learning is when the child learns to build a house that has never been constructed before based on their own understanding by experimenting with different constructs. The difference between supervised and unsupervised learning is an important distinction, and it will be used to explain some of the challenges of supervised learning".[34]

Analogously to the way "fraud" operates as an umbrella term for a range of illicit activities aiming at deceitfully tricking people, AI may likewise be seen as the umbrella term comprising multiple distinct subfields, one of which is ML. It is this last subfield which seems to be at the very heart of the various tools used in payment fraud detection.[35] Hence, although this contribution is not on engineering and we certainly do not possess the technical expertise to present the precise operational mechanics of these systems, it is important, for the sake of clarity, to briefly present some of their features.

---

34 Maris Sekar, *Machine Learning for Auditors. Automating Fraud Investigations through Artificial Intelligence*, (Apress 2022) 3–12.

35 Cf. notorious fraud and financial crime solutions known for being used by banking and financial institutions like Feedzai, Teradata, Ayasdi which all rely on some branches of ML to develop their AI based fraud detection tools: Feedzai, https://www.feedzai.com/ accessed 2 March 2026; Teradata Corporation, https://www.teradata.com/accessed 2 March 2026; Ayasdi, 'Delivering 100 % system compatibility for Ayasdi's ML infrastructyre' https://www.avenga.com/success/ml-solution-extracts-insights-from-data/ accessed 2 March 2026.

As already noted in Sekar's formulation, ML constitutes a subcategory of AI distinguished by its capabilities to identify patterns from a considerable amount of data, taking into account various factors, like the geographical origin of the payment authorization request, the transaction amount, or simply the frequency of the transactions. In the context of payment fraud detection, ML systems will rely on data on previous transactions, learning from them all while understanding how to distinguish a legitimate transaction from a fraudulent one. Among the various ML algorithms several types are recurrent for payment fraud detection. Decision trees and random forest algorithms,[36] for instance, are useful in classifying problems by distinguishing legitimate transactions from fraudulent ones, while deep learning algorithms will automatically extract and learn features from data.

By way of illustration–albeit in a simplified form- an AI-based fraud detection system employing deep learning algorithms will learn from the collected data and develop a behavioural profile for an individual account holder. Such a system may determine, on the basis of historical transaction data, that a given customer, for instance, typically conducts low-value payments from a consistent type of device, operating from a stable network location, and within a habitual time range. If, thereafter, a payment authorisation request for a significantly higher amount is initiated from a different type of device, via an atypical network location, and at an unusual time, the pronounced divergence from the established transaction profile would prompt the system to flag the activity as anomalous for further review. As a result, the fraud detection mechanism will flag the transaction as an out-of-the-ordinary payment hence unusual, so for the bank to review it and eventually block it.

36 For detailed studies explaining the various machine learning tools used for payment fraud detection cf. Victor Chang et al., 'Digital payment fraud detection methods in digital ages and Industry 4.0' (2022) 100 *Computers and Electrical Engineering;* Ryman-Tubb, Krause, Garn, (n 24); Vanini et al (n 24); C Chethana and Piyush Kumar Pareek, 'Analysis of Credit Card Fraud Data Using Various Machine Learning Methods', in Sita Rani et al., *Big Data, Cloud Computing and IoT Tools and Applications* (Chapman & Hall 2023).

3.1. Every rose has its thorn: On the advantages and challenges of ML payment fraud detection tools

The advantages of the use of AI based tools are certainly undeniable. Given the exponential and increasing number of digital payments,[37] the workload of processing the data has simply become inhumane especially if one considers the rapidity with which the authorisation requests need to be answered. As a result, inhumane work requires an inhumane answer, namely AI tools. Another advantage of AI tools, in particular ML ones, is their ability to learn and rapidly adapt to new fraud's techniques. Yet, once again, new opportunities go hand in hand with new challenges. In other words, despite the visible advantages of these AI tools, some issues still need to be addressed and ideally resolved.

It goes without saying that purchasing these AI tools comes at a significant cost for the banks, especially if one considers the fact that they are not, at least in the vast majority of cases, supposed to replace the humans who will still need to be placed in the loop. Yet, it seems to us that the argument of the cost of these tools is not a convincing one. As a matter of fact, a valid counterargument is to replicate that the costs of not detecting fraud are definitely much higher. This is particularly true within the EU where under article 73 of the PSD2, payment service providers can be held liable in case of unauthorised payment transaction[38] and will have to, except in some exceptional cases, reimburse the unauthorised payment transaction.

---

37 Cf. for some statistics on this matter: Committee on Payments and Market Infrastructures, 'Tap, click and pay: how digital payments seize the day', Brief n°3 (February 2024).

38 It should be noted that the border between authorised and unauthorised payment might become quite blurred in the context of digital fraud. Take for example the case of a CEO fraud which technically is committed with the consent of the victim who actually authorizes the transaction because of the fraudster's deceitful means who presents himself/herself as, in this case, the CEO. To address these new issues arising from new fraud forms, the future third directive on payment services (PSD3) is supposed to go beyond the PSD2. Cf. for more details on the subject: Emanuel van Praag, 'Authorised Push Payment Fraud: Suggestions for the Draft Payment Services Regulation' (2025)190 *European Banking Institute Working Paper Series* https://ssrn.com/abstract=5241100 or http://dx.doi.org/10.2139/ssrn.5241100 accessed 7 August 2025. Cf. also for life updates on the revision schedule: European Commission press release on 'Modernising payment services and opening financial services data: new opportunities for consumers and businesses' https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3543 accessed 17 July 2025. European Parliament – Legislative Train Schedule, 'Revision of EU rules on payment services', at the following link:

Another challenge, that applies to the use of AI in general, is the feared black-box effect[39] which pertains to the opacity of the latter's results. Indeed, the difficulties in understanding and assessing the accuracy hence the reliability and validity of the results generated by AI tools often prove to be problematic. Clearly, explainability is a sensitive topic, especially since it is legitimate for customers to require their banks to explain how and why a specific AI tool has taken a given decision. Nonetheless, we think that the black-box issue proves, for a series of reasons, to be less problematic in the framework of fraud detection tools. On the one hand, one should consider that although blocking a transaction because it is suspected to be fraudulent represents a sort of "interference with a person's free will", in the sense that in case of a false positive a legitimate transaction will not be authorized, it is not the most invasive decision that one can take toward an individual.

In contrast, one should think of decisions taken after relying on AI tools' results in the field of criminal law for instance. In this case, explaining how and why the AI tool reached a given conclusion upon which the law enforcers have relied on, will prove not only crucial in order to be used as evidence but also essential to preserve the defense rights. A scenario that does not fit the case of payment fraud detection, especially if one considers that, at the end of the day, not authorizing a transaction because it seems fraudulent is a precautionary act to spare the customer from a potential harm (and in some sense also the bank itself which under the PSD2 would eventually need to reimburse the lost sum). An act that seems to perfectly fit the "better safe than sorry" motto. What is more, one should not forget that blocking a transaction is not an irreversible decision. A customer whose digital payment request was refused can always ask for the reasons that led to such a refusal and, once proven that the former was a legitimate one coming from herself/himself, the transaction can be flagged as legitimate and, hence, authorized.

A more complex and perhaps the true Gordian knot when it comes to the use of ML for payment fraud detection is without any doubt the gathering and use of data. As already mentioned, a common point between

---

https://www.europarl.europa.eu/legislative-train/theme-an-economy-that-works-for-people/file-revision-of-eu-rules-on-payment-services accessed 17 July 2025.

39 Cf. for more details on the black-box issues: Chinu, Urvashi Bansal, 'Explainable AI: To Reveal the Logic of Black-Box Models' (2024) 42 *New Generation Computing*, 53–87; Luke Tredinnick and Claire Laybats, 'Black-box creativity and generative artificial intelligence' (2023) 40(3) *Business Information Review*, 98–102.

the various ML algorithms is the fact that they need to learn through data which, in this case, are nothing else than historical figures and information from previous existing experiences with regard to payment transactions and payment frauds. As rightly stressed out by several scholars "the fraudster behaves differently during a payment transaction than the account owner and/or the payment has unusual characteristics, such as an unusually high payment amount or transfer to an account in a jurisdiction that does not fit the life context and payment behavior of the customer. […] fraud in online payments can only be detected based on individual data, as such fraud can only be detected through the possible different behavior of the fraudster and the account holder during payments. As fraudsters learn how best to behave undetected over time, they adapt their behavior".[40] As a result, the quality as well as the quantity of the data used to train the algorithms proves to be vital if one wants to continue using them.

Against this background, a first strain of challenges which will only be briefly mentioned concerns the gathering of data and the privacy issues. It is important to note that training ML algorithms not only requires the gathering of large amount of data but will also touch upon more private and sensitive information such as an individual's transaction history. To overcome the privacy issues, one could rely upon the exclusive use of synthetic data.[41] Yet, even the latter will need some original data to be trained and furthermore, it is not clear whether they can cover and outperform data stemming from concrete and existing experiences. Nonetheless, one should not forget when it comes to payment fraud detection, that under article 94 PSD2, payment service providers and payment systems are explicitly allowed to process personal data to, notably, prevent and detect payment fraud. This same provision specifies that such processing and gathering will have to be limited to the scope of payment services and that it will require the consent of the user. Also, article 94 PSD2 refers to the general provisions of directive 95/46/EC, which was repealed and replaced by the General Data Protection Regulation (in short 'GDPR')[42], that will

40 Vanini et al (n 24).

41 Cf. for more details on the use of synthetic data for AI tools: Fernando Lucini, 'The real deal about synthetic data' (2021), *MIT Sloan Management Review,* at https://sloanreview.mit.edu/article/the-real-deal-about-synthetic-data/ accessed 2 March 2026; Vincent Granville, *Synthetic data and generative AI* (Morgan Kaufmann Elsevier 2024); Khaled El Emam, *Accelerating AI with Synthetic Data* (O'Reilly 2020).

42 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal

have to be applied by the providers gathering the data. Hence, without having to dwell into details of the GDPR requirements, let us highlight the main rules set out in the GDPR that need to be complied with by the implicated collecting entities. More to the point, pursuant to the GDPR, the consent to process the data (article 7), which shall be done in a lawful and transparent manner (article 5), communicating the specific information to the users at the moment they collect their data (articles 12 and 13) all while pursuing a specific and legitimate purpose (article 6) will have to be complied with. In the framework of digital payment fraud, *i.e.*, a clear case of crime prevention, the issue of gathering sensitive data, will be mostly dealt and complied with. Yet is gathering and processing data sufficient in the context of fraud detection through ML tools?

3.2. Sharing is caring?

The privacy preservation resurfaces with regard to another challenge pertaining to data gathering, namely the possibility to share them among banks to foster the identification of fraud patterns. It is already worth noting that under the GDPR sharing data with third parties is not prohibited, as long as the sharing is done in compliance with its provisions, notably that it is performed with the consent of the users, in this case the bank customers.[43] A consent which does not seem impossible to obtain. As demonstrated by Comply Advantage, one of the leading firms proposing AI tools for financial crime detection, including fraud, 65 percent of the surveyed customers were open to the idea of banks sharing their transactional details among themselves in order to foster the identification of fraud patterns.[44] Also, one could rely on the fact that there are currently new methods that are being developed to share data all while preserving privacy,

data and on the free movement of such data, and repealing Directive 95/46/EC, OJL 119/1.

43 Cf. articles 5(1); 6(4) and 89(1) as well as recitals 39 and 50 of the GDPR.

44 Comply Advantage, 'Technology is creating new fraud frontiers – but consumers don't want AI to protect them' https://complyadvantage.com/press-media/technology-is-creating-new-fraud-frontiers-but-consumers-dont-want-ai-to-protect-them/ accessed 15 August 2025.

which use so-called "privacy-preserving record linkage (PPRL)", able to hide personally identifiable information.[45]

As a result, a crucial and turning point in having the best working ML tools is the ability to rely on qualitative data to train the algorithms. In fact, as pointed out by some authors, one of the main challenges with ML tools is the quality of the data used to train the models.[46] Going back to Sekar's metaphor, just like a child who is provided with a bad education, if one provides AI with "bad" or so-called "dirty" data, *i.e.*, data which contain errors or are missing some specific values, the algorithm will absorb and learn from these data ending up providing unreliable and wrong interpretations. The importance of reliable data is also stressed out by Interpol, according to which to foster the prevention and contrast payment fraud through the use of AI tools the efforts should be concentrated into "strengthen data collection and analysis on financial fraud in order to develop more informed and effective counter strategies".[47] Against this background, a legitimate question arises, is there a way to obtain the highest-quality possible data to train the various algorithms?

Relying on the "unity makes strength" motto, the response to the question might lie in the acknowledgment that detecting payment frauds could be improved through the sharing of data among banks.[48] After all, which better way to get the whole picture of payment fraud than by adding and equating various experiences with the latter, just like for solving a puzzle, to get the whole picture you will have to manage to put all the pieces together. Nonetheless, if the solution is so straightforward, why do banks still miss out on this opportunity, or at least do not sufficiently profit from it? Especially if one relies on the already mentioned recent Comply Advantage survey which showed a high consent rate among customers to share their data among banks to prevent fraud. After all, detecting and fighting fraud, including digital payment one, is an issue already tackled, at

45 Cf. Satish Thomas and James Sluss, 'Fraud detection through data sharing using privacy-preserving record linkage, digital signature (EdDSA), and the MinHash technique: Detect fraud using privacy preserving record links' (2023) *The Journal of Engineering*.

46 Sekar (n 34).

47 Interpol (n 12).

48 Kaushikkumar Patel, 'Credit Card Analytics: A Review of Fraud Detection and Risk Assessment Techniques' (2023) 71(10) *International Journal of Computer Trends and Technology*, 69–79; Vanini et al (n 24); Basel Committee on Banking Supervision, Discussion Paper (n 26).

least from the criminal law dimension, at the EU level through the directive on combating fraud and counterfeiting of non-cash means of payment,[49] which requires Member States to set up, among others, criminal law measures for digital payment fraud (articles 3 to 6) all while providing penalties for natural persons (article 9) and a liability system for legal persons (article 10). Interestingly enough, this same directive foresees under its article 18 the establishment of a monitoring programme through the collection of data and other evidence. Its second paragraph specifies that Member States will have to put in place a system "for the recording, production and provision of anonymised statistical data measuring the reporting, investigative and judicial phases"[50] of the various fraud offences.

An argument that could be instinctively raised in opposition to such a large-scale data sharing and collection is the one arising from the risk related to data breaches and cyberattacks. Nonetheless, without underestimating this risk, one should not forget that, at least within the EU, the regulation on digital operational resilience for the financial sector (in short 'DORA')[51] was introduced precisely to strengthen the resilience of digital operations within the financial sector by creating a common information and communication technology risk management framework. Also, the regulation itself foresees under its article 45 the possibility and the conditions for financial entities to exchange cyber threat information and intelligence among themselves. Furthermore, whether by sharing data or not, banks, among others, will sadly but inexorably always be threatened by the risk of cyberattacks and data breaches, so that invoking this argument to oppose data sharing on payment fraud seems unconvincing.

What seems more convincing given the goals of the banks and the ineluctable competition spirit between them, is the fact that banks do not necessarily have an interest to share these data. Certainly, one could say that by sharing their data they will improve their own algorithms by gaining new data to train the latter. Yet, it seems safe to affirm that the biggest

49 Directive (EU) 2019/713 of the European Parliament and of the Council of 17 April 2019 on combating fraud and counterfeiting of non-cash means of payment and replacing Council Framework Decision 2001/413/JHA, OJ L123/18.

50 Article 18 of the Directive (EU) 2019/713 of the European Parliament and of the Council of 17 April 2019 on combating fraud and counterfeiting of non-cash means of payment and replacing Council Framework Decision 2001/413/JHA, OJ L123/18.

51 Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011, OJ L333/1.

banks are probably the ones with the largest database. Why should they share data on their customers with other smaller banks or, even worse, with direct competitors, *i.e.,* other significant banks? Is there a way of addressing the issues with the gathering and sharing of data? Could it be possible to set up a centralized database?

This idea might seem unreasonable, almost heretical at first sight. Yet, it relies on the fact that data sharing and especially the collection of data within a centralised entity is not a novelty, even when it comes to the fight against fraud. We are referring to the current centralised databases within the EU for fraud detection against EU funds, like the 'ARACHNE risk scoring tool'[52] based on data mining exploiting, both internal and external data, and which was put in place as a management verification for the European Structural and Investments Funds to detect, among others, frauds in the context of their activities. Also, the EU enjoys the benefits of the 'EDES' database[53], containing a list of persons that are excluded from contracts financed through EU funds, *inter alia*, because of fraud. A reliance on AI solutions that will certainly increase within the next years, as revealed by the Commission in its Anti-Fraud Strategy Action Plan where one of the themes is "Foster digitalisation and the use of IT tools to fight fraud".[54]

52 Cf. for more details on the functioning of the ARACHNE risk scoring tool the dedicated website at the following link: https://ec.europa.eu/social/main.jsp?catId=325&intPageId=3587&langId=en accessed 13 November 2024. Cf. also the publicly available European Commission's Charter for the introduction and application of the ARACHNE risk scoring tool in the management verifications, where it is possible to read that the tool "*is based on internal and external data. These internal data (projects, beneficiaries, contracts, contractors and expenses) are extracted by the managing authority from their local computerized systems and uploaded on a dedicated server of the Commission services. External data are provided by two external service providers […]. The first database contains financial data as well as shareholders, subsidiaries and official representatives of over 200 million companies. The second database is composed of a list of politically exposed persons, sanction lists, enforcement lists and adverse media lists*", 4–5.
European Commission, *Arachne Charter for the introduction and application of the Arachne risk scoring tool in the management verifications* (1st January 2024). https://employment-social-affairs.ec.europa.eu/document/download/a43cb969-8a8d-492b-b1c6-938edfb1f63a_en?filename=Arachne-Charter.pdf accessed 3 March 2026.

53 Cf. for more details on the functioning of the EDES database: https://commission.europa.eu/strategy-and-policy/eu-budget/how-it-works/annual-lifecycle/implementation/anti-fraud-measures/edes/edes-database_en. accessed 13 November 2024.

54 European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, the Committee

After all, the idea of creating a centralized database at EU level, at least within the Eurozone, seems also the next logical step to an already, some sort of existing, data gathering process. We are referring to the reporting approach foreseen by the PSD2, more precisely under its article 96(6) according to which payment service providers are required to provide the EBA and ECB with statistical data on fraud relating to different means of payment on an annual basis at least. Hence, a considerable amount of data is already gathered by these authorities. Of course, the question will be on the type of data it gathers. In contrast to the reporting under the directive of combating fraud, whose article 18(3) sets as a limit data on the number of offences related to fraud and the number of prosecuted and convicted persons, the data gathered under article 96(6) PSD2 seem more far reaching. Indeed, having a look at the already mentioned recently published EBA and ECB report on payment fraud, one can see the different categories under scrutiny, namely the levels of payment fraud (*i.e.*, the total value of fraudulent transactions reported) as well as the losses due to the latter, the different and main types of fraud, a more detailed study on the transactions authenticated or not through the strong customer authentication, the geographical dimension of payment fraud, as well as a country-by-country review on payment fraud. A rich variety of topics that gives an idea of the amount of data that is already reported, collected and within the hands of the ECB and EBA. Hence, considering the fact that EBA has already launched a EU's central database for anti-money laundering and counter-terrorism financing (*EuReCA*),[55] could a similar database with relevant data for fraud detection, accessible to the various banks, be a viable and useful in the banking context, by offering ML tools and more particularly their algorithms the best education possible? *Affaire à suivre.*

Conclusion

In an era where the use of new technologies has simplified the traditional banking services for the customers, to the point where through a simple

of the Regions and the Court of Auditors: Commission Anti-Fraud Strategy Action Plan (2023 revision)' (2023).

55 For more details on the new EuReCA Database, cf. EBA's dedicated webpage on the subject at the following link: https://www.eba.europa.eu/regulatory-technical-standards-central-database-amlcft-eu. accessed 13 November 2024.

click or by using a card, if not even a phone, they can proceed to cashless payments, it had to go without saying that new forms of risks would arise. Between friend and foe, these technologies, nowadays often based on AI, have become a daily reality just like digital payment fraud. To respond to these new risks, banks are increasingly leveraging AI tools, especially those based on ML technology, to prevent and detect fraud, particularly payment fraud.

And, if it goes without saying that the fight against fraud will be an eternal one, as witnessed by the fact that it exists since basically the beginning of History, to the point that it is quite safe to say that humanity will probably never get rid of it, it is also true that there might be solutions to efficiently contrast it or at least facilitate its detection. ML based tools have proven a necessary and useful ally in this eternal fight, yet, as rightly said by some authors: "AI requires the capability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent human beings: the ability to reason, finding out meaning for specific concepts, generalization from examples or learning from past experiences to make smart decisions in new (unseen) situations. The smart decisions in AI applications rely on the outputs of the learning process".[56]

While filtering bad from good data will remain within the hands of various engineers, the legal field could contribute by facilitating if not imposing the access to qualitative and quantitative significant data. A potential solution or at least a suggestion, fostering the sharing between banks, if needed through the establishment of a centralized datacenter.

56 Soviany (n 33) 102.

The Formation and Execution of Contractual Securities through Smart Contracts under Luxembourg Law

Théo Antunes

Abstract: Smart contracts, with their self-executing nature and reliance on blockchain technology, provide financial professionals with increased efficiency, transparency, and security in executing contracts, particularly in cases of debtor default in the context of contractual securities. Traditional securities require manual processes such as document signing, obligation verification, and potential court involvement in disputes, whereas smart contract securities execute automatically, ensuring prompt fulfilment of obligations without human intervention. Despite these significant advantages in time management and legal certainty, implementing smart contracts within the context of Luxembourg's security laws meets substantial technical and legal challenges.

Firstly, the inherent rigidity of smart contracts contrasts with the flexible and interpretative nature of legal securities. Securities law often necessitates adaptability to address unforeseen circumstances and equitable considerations, which are difficult to encode into the binary logic of smart contracts. This rigidity can result in automated executions that fail to reflect the nuanced intentions of the parties or changing conditions, potentially leading to unjust outcomes.

Secondly, Luxembourg's current legal infrastructure lacks comprehensive regulatory guidance specific to blockchain and smart contracts in this context. This regulatory gap creates uncertainties regarding the legal recognition, enforceability, and interpretation of smart contracts under contractual securities law. While Luxembourg fosters a fintech-friendly environment, the rapid advancement of blockchain technology outpaces the adaptation of legal frameworks, underscoring the need for targeted legislative and regulatory efforts to align innovative technologies with traditional legal principles.

Thirdly, the challenges faced by Luxembourg are echoed throughout the European Union, where a harmonized approach to blockchain and smart contract regulation is still developing. The European initiatives, such as the Digital Finance Package, the electronic IDentification, Authentication and trust Services regulation, and the Market in crypto-assets regulation, aim to create a cohesive regulatory environment. However, integrating smart contracts into existing legal systems, especially in areas like securities law, requires further detailed regulatory efforts. The purpose of this article is to emphasize a comprehensive regulation of the establishment and execution of securities through smart contracts under Luxembourg law. It also explores the potential opportunities offered by smart contracts, particularly considering the degree of flexibility required for different types of securities under Luxembourg law.

I. The intersection of securities' law and smart contracts under Luxembourg law

Smart contracts are self-executing contracts with the terms of the agreement directly written into code. They automatically enforce and execute actions when predefined conditions are met, without the need for human intervention or intermediaries.[1] They run on Distributed ledger technology (DLT) networks aiming at ensuring transparency, security, and immutability of the agreements built on them.[2] Such a definition presents smart contracts with several criteria: They are contracts, they are automated, immutable, and their execution is delegated to algorithms built on DLT. Each of these aspects presents challenges under Luxembourg law, for they intersect between traditional written contract law and new technologies relying on such a technology.[3]

Securities law is deeply linked with contract law, for securities aim at ensuring that the creditor will have compensation in case where the contractual obligation of the debtor is not fulfilled. As such, they are also bound by the principles of contract law, such as contractual freedom, the existence of an agreement, and the enforceability of the contract between the parties.[4] Because they are dependant to contracts, securities cannot exist without them.

However, the particularities of smart contracts lead to rethink the way securities can be formed and executed in the DLT environment, what are the challenges and the limitations that they can be subjected to. On one hand, Luxembourg's legal framework has known a rapid awareness to DLT used by financial professionals in several contexts (A). On the second hand, one needs to assess such framework with the lens of securities law to assess at what extent does the Luxembourg legal framework can encompass them in smart contracts (B).

1 Weiqin Zou and others, 'Smart contract development: challenges and opportunities' (2021) *Research Collection School of Computing and Information Systems* 1–22, 1; LetzBlock, *Smart contract: a Luxembourg perspective, assessing opportunities and key legal challenges* (Smart Contract Working Group 2021) 1–68, 7.

2 LetzBlock, *Smart contract: a Luxembourg perspective, assessing opportunities and key legal challenges* (Smart Contract Working Group 2021) 1–68, 13

3 Elise Melchior, 'Réflexions juridiques autour de la blockchain : analyse sous l'angle du droit des contrats' (2018) 72 *Revue du droit des technologies de l'information* 45–65, 53

4 Luxembourg Civil Code, article 1108.

A. Clarifying the formation and execution of securities in Luxembourg through smart contracts.

Financial securities are crucial for the Luxembourg's financial sector. They aim at securing that the creditor can be compensated in cases where the debtor fails to fulfil the obligations inscribed in a given contract.[5] Several financial securities exist in Luxembourg, and they adapt to different sort of situations, whether mortgage, financial guarantee contract, a pledge, transfer of property targeting financial assets, collateral… Securities are adapted depending on the contract they are inserted in, the business relationship, the parties in order to ensure for both the creditor and the debtor an adequate contractual framework. For instance, mortgage is common for real estate loans, and transfer of assets as collateral are common for loans involving asset managers to credit establishments. Each type of security aims at answering an adapted contractual relationship between a debtor and a creditor.

Because of such flexibility, the formation and execution of securities differ depending on the selected one, they are nonetheless framed by the same principles. These principles aim at making these securities contractually legal (1). Moreover, the recent years have shown a rising flexibility for financial professionals to form and execute securities (2).

1. Principles of formation and execution of securities in Luxembourg.

What are the principles framing the formations and executions of securities? Since both the formation and execution answers to different objectives, their principles differ according to Luxembourg contract law, which securities are bound to. Hence it is necessary to distinguish between these two. Securities are framed by three distinct criteria: the purpose of the security, its effect and the technique used to form it.[6]

Firstly, the formation of securities in Luxembourg is bound to the same principles than the contract it depends on. Securities are governed by the general rules of contract formation. According to article 1108 of the Civil Code, "*four conditions are essential for the validity of an agreement: the*

5 Antoine Cuny de la Verryère, *Sûretés et garanties au Grand-Duché de Luxembourg* (Larcier Luxembourg 2018) 26.

6 Pierre Crocq, *Propriété et garantie* (LGDJ 1995) 248.

consent of the party who undertakes the obligation; their capacity to contract; a certain object which constitutes the subject matter of the commitment; and a lawful cause in the obligation".[7] In the context of forming securities however, further clarifications need to be made. Contractual freedom entails the consent of the parties to have recourse to securities as part of the contractual clauses. Thus, according to contract law, this means that the debtor's consent must be free, meaning that it should not be subject to any coercion, violence or misleading attempt from the creditor.[8] The object, for its part, must be lawful, possible, and determinable.[9] The cause of a security interest generally lies in the payment of the principal debt.[10] Moreover, parties must have contractual capacity, which is divided into the capacity to enjoy rights, on the one hand, and the capacity to exercise rights, on the other. It is negatively defined by article 1124 of the Civil code entailing that some categories of populations, such as underaged, the persons that are forbidden to contract and generally all people whom the law forbid to contract.[11]

Those four principles entail that both the creditor, but especially the debtor, needs to be fully aware of the ramifications that such security entails and the conditions for its execution. If these principles frame the formation of securities, they must also answer to further requirements under Luxembourg's law, depending on their nature. For instance, if most of the securities require proof of the failure of the debtor to uphold its obligations to be executed, some can be executed before bringing such proof such as the "*Garantie à première demande*".[12] It would be then up to the creditor to demonstrate such failure, but such *ex-ante* execution can be harming for debtors that are not fully aware of the implications of this precise security. When assessing the consent under contract law, one could see that it is be limited to the object of the security as well as its execution. The form of the security is often required to be written and clear enough to avoid any misleading from the creditor. But the written form *per se* does not fall under the appreciation of the consent, only its content and the way it is executed.[13]

7 Luxembourg Civil Code, article 1108
8 Luxembourg Civil Code, article 1109,
9 Luxembourg Civil Code, article 1110
10 Luxembourg Civil Code, article 1131
11 Luxembourg Civil Code, article 1124
12 Court of Appeal (Luxembourg), 16 March 1983, *Droit et Banque* No 2, 39, para 25.
13 Philippe Ancel, *Contrats et obligations conventionnelles en droit luxembourgeois* (Larcier 2015) 261–332, 269.

To fully assess the formation of securities in Luxembourg, one must also consider the application and evolutions of the law of 2005 on financial securities, expanding the number of securities available for financial professionals under commercial law.[14] Among them, the pledge on assets can take the form of share pledges, pledges on receivables or pledges on account. They answer to special rules for the formation of the security that the contract must upheld. A dispossession must occur, under which the debtor transfers the possession of the financial assets to the creditor, meaning that the debtor be the owner of those assets.[15] Another one yet, the transfer of ownership is another security enshrined in article 13 of the law of 2005, it entails that on the one hand, it involves an assignment mechanism whereby ownership of assets is transferred to the beneficiary of the security to secure the financial obligations of the grantor.[16] On the other hand, it includes a reassignment mechanism, where the assignee commits to transferring back the assets or equivalent assets at an agreed date, except in the event of total or partial non-performance of the secured financial obligations. Thus, this law entails specific rules on the formation of these securities, detailing precisely how the formation should occur, pending that they also answer to the principles of article 1108 of the Civil code. More recent yet, the law of 10th July 2020 brings a new type of security, the professional payment guarantees, whose modalities are let free to the parties, guaranteeing more freedom for its formation and execution, pending that they are also framed by the principles set forth by article 1108 of the Civil code.[17]

Secondly, the execution of securities is dependent on the failure of the debtor to uphold its contractual obligations.[18] As mentioned above however, such failure can be demonstrated with a certain extent of flexibility, especially considering what and when the security can be executed. The law of 2005 also attaches special rules for the securities it enshrines. For instance, a triggering event that leads to the enforcement of the security, as stipulated by the agreement, must have occurred, notably concerning

14 Loi du 5 août 2005 sur les contrats de garantie financière, articles 2, 3, 13.

15 Loi du 5 août 2005 sur les contrats de garantie financière, article 5 (1); Antoine Cuny de la Verryère, *Sûretés et garanties au Grand-Duché de Luxembourg* (Vademecum Larcier Luxembourg 2018) 61.

16 Loi du 5 août 2005 sur les contrats de garantie financière, article 13

17 Armel Waisse and Laurent Henneresse, 'Luxembourg: law of 10 July 2020 on professional payment guarantees' (2020) *Eurofenix: Country Reports 2020/21,* 41

18 Antoine Cuny de la Verryère, *Sûretés et garanties au Grand-Duché de Luxembourg* (Vademecum, Larcier Luxembourg 2018) 35.

the pledge security.[19] This triggering event must be precisely defined and can consist of multiple causes: a change of control, deterioration of the debtor's financial ratios, default on payments related to the underlying credit agreement or a related contract, meaning the failure to pay commissions to third parties involved in the financing operations. The triggering event defined by the parties constitutes an autonomous realization event, breaking with the requirement for the secured debt to be due. Luxembourg law, therefore, allows for the enforcement of a pledge even in the absence of the secured debt's maturity, they can be two different events if they are explicitly inscribed as such by the security triggering event.

2. Smart contract and securities: What are we talking about?

The use of DLT as a mean to form and execute securities imply that they are registered and signed digitally by the parties.[20] The main opportunity of such endeavour lies in its execution, where the smart contract, on one hand assess autonomously whether the conditions for the execution of the securities are met and on the other hand, proceeds to the execution of the security to the profit of the creditor.[21] For instance, in the case of a mortgage, the transfer of the deed of the ownership of the house could be transferred to the credit establishment in cases of failure of the debtor to uphold its obligations. This mechanism requires several requirements larger than the establishment of the smart contract.

Firstly, it requires a certain extent of tokenization to allow a transfer of the collateral pursuant to the security agreement.[22] Indeed, to operate adequately and according to its nature, the smart contract needs to have collateral available for its automated transfer in case of the debtor failure to uphold its contractual obligations. Operating on a DLT requires the funds to be operating on it as well. Hence, whatever the collateral is, it needs to be on the same layer as the contract, meaning the same DLT. As such, the collateral can take the form of crypto-assets, tokenized financial assets, tokenized deeds of property such as ownership of real estate or other types

19 Loi du 5 août 2005 sur les contrats de garantie financière, article 11(5).

20 Deutsche Börse AG, *How can collateral management benefit from DLT* (Deutsche Bundesbank Project 'Blockbaster' 2020) 1–21, 4.

21 Howard Altarescu and others, *Finance, securitization and smart contracts: business white paper* (MOBI 2021) 1–27, 14.

22 ISDA, *Private international law aspects of smart derivatives contracts utilizing distributed ledger technology: French law* (2020) 1–21, 17.

of financial assets that underwent the process of tokenization. Thus, the setting up of a smart contract for the execution of securities previously requires a process of a certain extent of tokenization from both the debtor and the creditor.

Secondly, to form the smart contract, the parties must agree on the type of security they would like the smart contract to execute, the stakes, the triggering event and all the conditions depending on the nature of the security at stake. It must be done in compliance with the formal legal requirements depending on such choice. However, this must be done in full compliance with the applicable laws and by reference to article 1108 of the Civil code. Such a position implies that a smart contract is not fundamentally different from the process of formation of a traditional contract, it just occurs on a different format.[23] However, the main difference lies in the immutability of the software, meaning that once it is written up on the DLT, will not be possible for the parties to change the terms of the security nor its execution.

Thirdly, the main point differentiating smart contracts from traditional execution of securities, the execution is totally outsourced to the smart contract programming.[24] This means that the triggering event, the determination of the failure of the debtor as well as the transfer of funds, assets or the deeds are done without any human intervention. The program should be able to assess whether the debtor is upholding all his obligations and whether the requirements for the triggering event has been met. The execution of the smart contract makes the recovery of the collateral faster for the creditor, but it also requires a certain amount of trust in the programming, needing it to be properly programmed to be executed as the parties intended. Decentralized autonomous organizations already make use of smart contracts to ensure the formation and execution of securities by taking crypto-assets as collateral before accepting a further transaction.[25] This type of organization is totally decentralized, meaning it operates with

23 Elise Melchior, 'Réflexions juridiques autour de la blockchain : analyse sous l'angle du droit des contrats' (2018) 72 *Revue du droit des technologies de l'information*, 45–65, 54–55.

24 Stuart D Levi and Alex B Lipton, *An introduction to smart contracts and their potential and inherent limitations* (Skadden, Arps, Slate, Meagher & Flom LLP and Affiliates 2018) 1–11, 3.

25 Juliette Sénéchal, 'Blockchains « publiques », smart contracts, organisations autonomes décentralisées et gouvernance' in *Les blockchains et les smart contracts à l'épreuve du droit* (Larcier 2020) 51–97, 68.

a central governing authority. MakerDAO for instance manages the Maker protocol enabling users to generate DAI stablecoins.[26] Users who want to generate DAI can do so by locking up cryptocurrencies in a smart contract as collateral. In exchange, they receive DAI tokens, which they can use freely. If the value of the collateral falls below a certain threshold, the system automatically liquidates the collateral to ensure that DAI remains fully backed.

One could see that this third step fundamentally differs from the traditional approach of securities, as the program should, in principle, be allowed to assess whether the debtor is upholding his obligations by itself. This requires a certain autonomy from the program to both recognize the triggering event as well as executing the full security as agreed upon by the parties.

B. Addressing the regulatory landscape for smart contracts and securities

DLT has gained a lot of awareness in the past years, both from regulators perspective and financial professionals alike. They show promises for financial professionals, whether by automating processes, tokenizing financial assets, better communications for whistleblowing channels or as a mean for expanding their activities with the rise of crypto-assets, cryptocurrencies and at some extent, non-fungible tokens. This technology was addressed at the European level as part of the digital finance strategy, aiming at providing renewed digital infrastructure for financial professionals, enhance financial inclusions for consumers and bolstering the financial flows within and out of the EU.[27] As such, the Luxembourg approach of DLT and blockchain regulations is delved within a European dynamic to regulate such technology for financial purposes (1). Such dynamic was enshrined in Luxembourg's law, notably since 2019 with four "*Blockchain law*" for DLT application by financial professionals (2).

1. A European dynamic

Lately, the EU has undertaken a regulatory approach aiming for the regulation of digital finance. Whereas no binding instruments is directly reg-

26 MakerDAO Maker Team, *The Dai stablecoin system* (White Paper 2017) 1–21, 3–5.

27 European Commission, 'Digital finance strategy for the EU' COM(2020) 591 final, 24 September 2020.

ulating the use of smart contracts, European norms tend to apply in a general and special approaches.

Firstly, several regulatory texts aim at framing the deployment of DLT for financial professionals. Since smart contracts often contain personal data, for the sake of its normal functioning (such as bank coordinates, name of the parties, addresses...) any smart contract that processes personal data must ensure compliance with General data protection regulation (GDPR) requirements, including obtaining explicit consent from individuals and ensuring data protection by design and by default.[28] Because of their digital nature and their reliance on Blockchain, a cryptographic key is needed to prove ownership or authorization to undertake transactions pursuant to the smart contract, one needs to have a cryptographic key, requiring an electronic signature that must be compliant with the electronic identification and trust services (eIDAS) regulation to ensure those signatures are legally recognized across the EU.[29] To rely on smart contracts to form and execute financial securities also entail a compliance with Digital operational resilience act (DORA)'s standards for operational resilience, ensuring the system is safeguarded against cyber-attacks aiming at changing the content of the smart contracts or redirecting the funds.[30] To the extent that the smart contracts involve consumer transactions, for forming and executing securities related to mortgage for instance, credit establishments need to provide clear information and respect the consumer's right to withdraw under certain conditions in compliance with directive 2011/83/EU on Consumer's right.[31] This bundle of hard-law frames the application of DLT for securities at the European level, nevertheless it requires a challenging cross-analysis of these norms to ensure its compliance with European rules. the Regulation (EU) 2022/858 on a pilot regime for market infrastructures based on distributed ledger technology aimed at providing common standards for providing market infrastructure relying on DLT, to assess the feasibility of the standards and the practical challenges associated with it.

28 Michèle Finck, *Blockchain and the General Data Protection Regulation: Can Distrubuted ledgers be squared with European Data Protection Law Study* (European Parliamentary Research Service, July 2019) 6.

29 European Law Institute, *ELI principles on blockchain technology, smart contracts and consumer protection* (Report of the European Law Institute 2023) 1–52, 15.

30 Giuseppe Siani, *Regulating new distributed ledger technologies (DLT): market protection and systemic risks,* Speech, (Banca d'Italia Eurosistema 2022) 1–12, 4.

31 Directive 2011/83/EU on consumer rights [2011] OJ L 304/64, article 6.

Furthermore, these binding norms must be read together with approaches undertaken by the European Commission and the Parliament on Fintech and DLT technologies. These documents act as guidance aiming at precising the European expectations in the field of DLT applications. Among them, the European Parliament Resolution on Distributed Ledger Technologies and Blockchains: Building Trust with Disintermediation,[32] the Digital Finance Strategy for the EU (2020), the Blockchain and the GDPR, EU Blockchain Observatory and Forum (2018), the FinTech Action Plan: For a More Competitive and Innovative European Financial Sector (2018) and the Shaping Europe's Digital Future (2020)L Together, they aim at providing guidance on the expected standards that DLT should adopt at the European level. Several principles are enumerated, among them: transparency, cyber-security or compliance with AML norms. This European framework aims at establishing a European DLT ecosystem, mostly in cases where it is used as a new market infrastructure, meaning a support for transaction flows constating in tokenized or crypto-assets. This European dynamic is completed with a Luxembourg's one, demonstrating the will for the Luxembourg legislator to directly frame the use of DLT in the financial sector.

2. A Luxembourgish dynamic

Luxembourg has been proactive in its regulation of DLT used in the context of financial acuities. Commonly referred to as "*Blockchain laws*", Luxembourg has been at the foremost of DLT regulation in the financial sector.

The Blockchain Law I of the 1st of March 2019 was one of the first in the EU to equate blockchain in the context of traditional transactions, allowing the use DLT for account and financial assets registration.[33] The Blockchain Law II of 22nd January 2021 further supported dematerialized assets by enabling secure electronic registration on DLT and broadening access to financial institutions.[34] The Blockchain Law III of 14th March 2023 enhanced this framework by updating securities laws, aligning with EU regulations,

32 European Parliament resolution of 3 October 2018 on distributed ledger technologies and blockchains: building trust with disintermediation [2020] OJ C 11/3.

33 Loi du 1er mars 2019 portant modification de la loi modifiée du 1er août 2001 concernant la circulation de titres.

34 Loi du 22 janvier 2021 portant modification : de la loi modifiée du 5 avril 1993 relative au secteur financier et de la loi du 6 avril 2013 relative aux titres dématérialisés.

and redefining financial instruments to include DLT. It notably updated the 2005 Law on Financial securities arrangement allowing use of electronic DLT for collateral over financial instruments held in securities accounts. This law opens the door for securities registered on DLT, but do not address the formation and execution of securities through smart contracts.[35] The Blockchain Bill of Law IV of 24th July 2024 aims to expand Luxembourg's DLT legal framework by including equity securities alongside debt securities, allowing the fund industry to use DLT for managing shares, units, and registers. It introduces the role of a control agent as an alternative to the central account keeper, responsible for maintaining securities issuance accounts, verifying DLT consistency, and supervising the custody chain.[36] The law simplifies payment processes and the issuance of dematerialized securities, permits the direct crediting of securities to investor accounts, and allows for smart contract integration.

As such, Luxembourg does not regulate smart contracts used for forming and executing contractual securities. Mostly, the Luxembourg legal framework seems to focus on investment opportunities for clients rather than contractual opportunities for securities. They are the object of investments and can be used as collaterals, but they do not constitute the means to form nor execute a contractual agreement, and *a fortiori* a security. The different laws have expanded the scope where DLT can be used as an object of investment or collateral, it does not encompass the use of DLT as a mean form and execute traditional securities. Luxembourg has not yet legislated nor the CSSF have established guidelines for the use of DLT as a mean to form and execute securities.[37] However, such lack of regulation does equate to a legal void, as DLT used a mean to form and execute securities can be

35 Loi du 15 mars 2023 portant : modification de
la loi modifiée du 5 avril 1993 relative au secteur financier; la loi modifiée du 5 août 2005 sur les contrats de garantie financière; la loi modifiée du 30 mai 2018 relative aux marchés d'instruments financiers et mise en œuvre du règlement (UE) 2022/858 du Parlement européen et du Conseil du 30 mai 2022 sur un régime pilote pour les infrastructures de marché reposant sur la technologie des registres distribués, et modifiant les règlements (UE) n° 600/2014 et (UE) n° 909/2014 et la directive 2014/65/UE.

36 Projet de Loi n° 8425, portant modification : de la loi modifiée du 6 avril 2013 relative aux titres dématérialisés; de la loi modifiée du 5 avril 1993 relative au secteur financier; de la loi modifiée du 23 décembre 1998 portant création d'une commission de surveillance du secteur financier.

37 CSSF, *Distributed ledger technologies and blockchain: technological risks and recommendations for the financial sector* (2022) 1–44, 4.

permitted, but it must be framed adequately pending a compliance with the technological nature of such endeavour as well as compliance with the principles of contract law.

Coupled with the innovative stance taken by Luxembourg on DLT, the formation and execution of securities through smart contracts remain tangible for the near future. This section has demonstrated that, while DLT and Blockchain have been considered by Luxembourgish law, it does so as an object rather than a mean to enforce obligations. Despite such an awareness, smart contracts as a mean to form and execute securities are not yet framed by law. In such a perspective, it is important to assess at what extent can securities rely on such technology and what their regulation would be.

II. Addressing challenges for establishing securities through smart contracts

As abovementioned, the formation and execution of securities relying on DLT imply a degree of automaticity and immutability that would ensure a certain degree of legal security. However, this also requires a relinquishment from the parties to the smart contract's execution. Thus, two main aspects present challenges for such prospect. On one hand, challenges due to the technological nature of the smart contract (A), on the other hand on the nature of securities (B).

A. Challenges stemming from the smart contract perspective.

As a technological tool, a smart contract brings several advantages for forming and executing securities. Although it requires a certain amount of digital assets established on the Blockchain, it could bring new means for financial institutions to enforce their securities. Due to its automation and immutability, legal certainty of the execution of the security would be prioritized. However, such perspective entails legal challenges regarding the establishment of DLT for this specific purpose. In order to secure the formation and execution of the security, several challenges need to be addressed, especially regarding the technical standards the smart contract needs to meet (1). Furthermore, the features of the smart contract, meaning the automation, delegation, and immutability, need to be addressed regarding under the context of securities (2).

1. From contractual clauses to computer code

The purpose behind the technical standards of the smart contract is to be able to form and execute the security as intended by the parties, as it lies at the core of its formation. Any derivation from such intent might strike the contract with the vice of disrespecting the will of the parties, hence leading to render the security clause void. This entails a dual approach.

Firstly, an intrinsic approach needs to be taken, meaning the one endorsed by the parties and the programmers. It entails that the Smart contract is properly encoded as to follow the will of the parties. The process of coding contractual clauses concerning collateral into smart contracts, particularly for financial securities or lending arrangements, involves ensuring that the intent of the parties regarding the collateral is translated accurately into code**.** The legal provisions governing the security, such as how and when it can be accessed, liquidated, or returned, must be accurately reflected in the smart contract. For instance, a smart contract needs to continuously monitor the value of the collateral, using price oracles, meaning trusted, appropriate, and relevant data feed that would automatically trigger liquidation if the collateral's value fell below the agreed threshold.[38] Oracles are a mean to obtain real-time, external data feeds about the value of the collateralized assets.[39] For example, if the collateral consists of financial instruments like stocks or bonds, the smart contract must integrate with oracles to fetch market data. Thus, they are indispensable for the proper execution of the security. However, parties might rely on different sources to assess the value of the collateral. Hence, they must agree on reliable price sources such as Bloomberg, Refinitiv, or cryptocurrency price oracles to feed real-time valuation data into the smart contract. Indeed, the smart contract should specify which data provider will serve as the single source of truth for collateral valuations. Moreover, legal contracts related to collateral can sometimes be open to interpretation of the creditor. For example, a clause might require the creditor to "*act in good faith*" or provide "*reasonable notice*" before liquidating the collateral.[40] This is challenging for the implementation of smart contract solution, for these terms are highly

38 Chanelle Duley, Leonardo Gambacorta, Rodney Garratt and Priscilla Koo Wilkens, 'The oracle problem and the future of DeFi' (2023) *BIS Bulletin* No 76, 1–7, 1.

39 ibid.

40 Philippe Ancel, *Contrats et obligations conventionnelles en droit luxembourgeois* (Larcier 2016) 15, 128, 143; Philippe Hoss and Patrick Santer, 'Le contrat fiduciaire en droit luxembourgeois' (Larcier 2003) *Études fiscales internationales* 781–856, 797.

subjective whereas the smart contract reasons in terms of quantitative and objective terms dependent on the programming it is subjected to.

However, one needs to distinguish such subjective terms from objective and quantitative solutions for adapting the contractual clause to the smart contract. For instance, measurable conditions such as margin requirements, collateral valuation on the relevant market could provide an adapted solution to avoid a subjective approach of the smart contract. This also entails that smart contracts used for securities purposes would not be adapted to all kind of securities, only ones that are able to be triggered with objective and quantitative event. Hence, for a smart contract to function properly, the terms related to collateral must be precise and quantifiable. This involves defining objective conditions that the smart contract can automatically monitor and enforce. If the collateral consists of assets like stocks, bonds, or cryptocurrencies, the smart contract must integrate oracles that provide up-to-date price feeds. Thus, the contract will need to define how often the collateral's value is checked and which source is considered authoritative. Clear thresholds need to be set for when collateral can be liquidated. For example, the smart contract could include a rule stating that if the value of the collateral falls below 70 % of the outstanding loan, the liquidation process is automatically initiated. Each clause related to collateral: its valuation, its liquidation and its substitution should be represented as a separate, modular function in the code.

Secondly, an extrinsic approach, ensuring that the smart contract works accordingly and is verifiable for auditing purposes. Formal verification can be used to technically verify that the smart contract code corresponds to the legal agreement. This ensures that the collateral management, valuation, substitution, or liquidation; is executed exactly as agreed upon by the parties. More than this 1st control verification, there needs to be a 2nd and 3rd control verification, to auditors and to courts in cases where the formation and execution are disputed by one of the parties. This requires an auditing procedure whose tasks is to ensure the technical standards of the smart contracts are upheld Ensuring that the smart contract code is auditable allows third parties and courts, in cases of legal dispute, to review whether the contract behaved as expected. This is especially important for disputes concerning collateral liquidation, where significant monetary value could be at stake. More than the program itself, it would also require a detailed documentation outlining how the smart contract's collateral terms were codified, and the original intent of the parties can serve as evidence in case of disputes.

2. The Automation, the immutability, and the delegation

Other notable challenges regarding the implementation of securities through smart contracts regard their intrinsic features, meaning their automation, immutability, and the delegation of its execution to a program.

Automation entails several challenges, especially regarding the complexity of securities' management. As previously mentioned, the value of collateral fluctuates particularly when dealing with market instruments, certain qualified as volatile such as certain types of crypto-assets. An oracle must constantly monitor market data to ensure that the collateral meets the agreed threshold. However, the value of some instruments can be volatile, and setting automated liquidation triggers can lead to premature or unwanted liquidations due to short-term market fluctuations without considering the long-term approach of certain instruments. In collateralized transactions, a margin call may be triggered automatically if the collateral falls below a certain percentage of the loan value.[41] However, there are challenges in automating complex margin management. For example, how should the contract handle partial liquidations or top-up requirements? These processes, traditionally managed by humans with room for negotiation, may be harder to automate and outsourced entirely. Black swan events, meaning unexpected market conditions concretized by liquidity shortages, market disruptions or market crashes can impact the collateral value leading to forced liquidation at an unwanted value for both parties.[42] Furthermore, if an oracle fails to deliver accurate data, the smart contract might fail to execute correctly or could initiate unintended actions such as triggering a liquidation at the wrong time. Human managers often have discretion in deciding how to handle situations such as minor breaches of collateral terms, giving parties time to top up collateral or negotiate. In contrast, a smart contract will automatically execute based on pre-programmed rules, potentially leading to outcomes that the parties would not have chosen if human discretion were involved. For example, the smart contract may immediately liquidate assets if a small margin call is missed, even if the default is temporary.

41 Committee on Payment and Settlement Systems, *Strengthening repo clearing and settlement arrangements* (Bank for International Settlements 2010) 1–65, 7.

42 Indrani Sengupta, 'An Exhaustive Study of the Black Swan Events in the Financial Markets' (2020) 10(2) *Journal of Management Sciences* 22–35, 22.

Securities agreements often evolve based on market conditions or business needs.[43] If the terms of the contract need to be modified, for instance adjusting collateral requirements in response to a change in market changes, one can take the path of renegotiation. However, with immutable smart contracts, once the code is deployed, changes are extremely difficult to implement. If the parties wish to amend the collateral terms, they will have to either deploy a new smart contract, which could be costly and time-consuming, or rely on off-chain agreements, which could undermine the trust and automation provided by the smart contract in the first place. If there is an error in the smart contract code, it cannot easily be fixed due to its immutability and could result in a failure to execute proper margin calls, or premature liquidation, which could have severe financial implications for the debtor.

Delegating key decisions related to collateral management to a smart contract can enhance efficiency but also introduces new risks. In traditional financial transactions, humans manage collateral with the ability to negotiate, delay, or adjust actions. When collateral management is delegated to a smart contract, parties lose the ability to exercise discretion or make subjective decisions. For example, if collateral values are volatile but likely to recover, a human might choose to wait before liquidating; a smart contract would proceed with liquidation according to its code, regardless of market context. Once authority is delegated to the smart contract, the contract will enforce collateral rules rigidly, without considering the context or relationship between the parties. This can lead to situations where a party's intent is not respected by the automation of the contract. Moreover, in some cases, collateral agreements involve multiple parties. Delegating collateral management to a smart contract requires careful coding of each party's rights and obligations. This adds significant complexity, especially when trying to automate decision-making that traditionally involved negotiation among parties. Most smart contracts do not have built-in dispute resolution mechanisms like traditional contracts entailing that once the collateral is liquidated there is no possibility to reverse the transaction unless the smart contract explicitly allows for it.

Furthermore, certain complex situations may require human judgment. For example, if the collateral's value is disputed or if there is a misunderstanding about the terms of the contract, human oversight is often

43 Albert Choi and George Triantis, 'Market Conditions and Contract design: Variation in Debt Contracting' (2013) 88(51) *New York University Law Review,* 51–61, 53.

necessary to resolve the issue, through mediation or court procedures. Delegating the entire process to a smart contract without room for human intervention can exacerbate conflicts when unexpected or nuanced situations arise. Moreover, as regulations evolve, the terms governing collateral in financial markets may need to be updated to reflect new legal requirement. Immutability makes it difficult to adjust smart contracts to remain compliant with such evolving legal frameworks. Ultimately, such a smart contract could result in non-compliance with adequate regulations, exposing the parties to legal or regulatory penalties and a voided contract. Immutability presents advantages, but also poses risks to the coherence of securities and the liquidation of the collateral, where subjective approaches can be privileged over an automated application of the contract. Through its technical standards, smart contracts present challenges for forming and executing securities. Furthermore, on a reversed approach, challenges stemming from the security law perspective also need to be addressed.

B. Challenges stemming from the security law perspective.

Not only under the challenges stemming from the Smart contract perspective, but the recourse to smart contracts to form and execute securities meet challenges from their legal framework. As such, there are challenges regarding the formation (1) and the execution (2) of contractual securities under Luxembourg law.

1. The formation of securities with a smart contract

The formation of smart contract containing contractual securities must be compliant with the general principles framing Luxembourg contracts. The main challenges lies on the consent of the parties to have recourse to such contract and at what extent the form it takes, an automated smart contract, is considered.

Thus, the question arises whether the form of the contract should be subject to the parties' consent. Traditionally, contracts are shaped documents outlining the obligations of both parties, with the execution of the security it contains carried out manually, notably at the initiative of the creditor. However, with the advent of smart contracts, the paradigm is shifting towards automation, where the software executes the terms of the contract

autonomously, without further involvement from the parties nor requiring such initiative from the creditor. In this context, where the form a contract has a significant impact on its execution, it is essential to consider the parties' consent extension to the smart contract form itself. Usually, the parties' consent focuses on the content of the contract, that is, what they agree upon and enshrined in its clauses.[44] If oral contract can be permitted, most of the time, they are under the written format. However, smart contracts are based on code and exist in a digital format. Thus, smart contracts introduce a new layer of complexity, as they automate execution once encoded, with no further input from the parties.[45] As a result, the form of the contract directly affects its execution and, consequently, has a direct impact on the parties themselves. Because of this direct influence, consent of the parties should be directed, more than the content of the contract, on the choice to resort to a smart contract as well. In this aspect, pursuant to article 1108, such consent must be clear, informed, and unequivocal.

In the context of smart contracts, the challenge is whether parties fully understand what they are agreeing to, particularly due to the technical complexity of how these contracts are executed. Thus, parties may consent to the terms, but might not fully comprehend the automated execution process, which may trigger legal issues if unforeseen consequences arise.

2. The execution of securities through a smart contract

Smart contracts are self-executing and operate based on predefined conditions encoded into the software. This creates issues with traditional concepts of contract performance and enforcement under Luxembourg law, which generally involves parties and possibly courts overseeing its performance. The automatic execution of smart contracts leaves little room for discretionary decision-making or renegotiation, which is a common feature in traditional contracts.

Indeed, the conditions encoded into the smart contract are fulfilled, the contract executes automatically, leaving no room for human intervention. This rigidity contrasts with traditional contract law, which allows for rene-

44 Pascale Dufour, *La théorie générale du contrat au prisme des travaux d'Axel Honneth* (Larcier 2023) 425.

45 Stuart D Levi and Alex B Lipton, 'An Introduction to Smart Contracts and Their Potential and Inherent Limitations' (2018) *Harvard Law School Forum on Corporate Governance* 1–9, 2.

gotiation or modification during performance. Under Luxembourg law, parties in traditional contracts can invoke mechanisms like good faith, unforeseen circumstances qualifying as force majeure allowing an alteration or suspension of the execution.[46] With smart contracts, such flexibility is limited or non-existent unless pre-programmed into the contract. The absence of discretion in the execution process means that smart contracts do not allow for the human judgment traditionally involved in contract performance.[47] If an unexpected situation arises that changes the context of the agreement, such as force majeure, the smart contract will still execute the terms as coded by the parties. This could lead to outcomes that neither party anticipated or desired. For instance, Luxembourg law recognizes the principle of good faith in contract execution, requiring parties to act fairly and to avoid any actions that would undermine the purpose of the contract and the security.[48] With smart contracts, the rigid and automated execution could challenge this principle. For instance, if circumstances arise that would typically require one party to act in good faith and delay performance, such as the delivery of goods under dangerous weather conditions, the smart contract will not have the capacity to adapt unless it was specifically programmed to do so. This drastically contrasts with the traditional approach of securities, where such events, despite being external to the contractual relationship, can be considered before resorting to its execution.[49] Relying on smart contracts requires such a step, that was not entirely considered in traditional execution, for it is overseen by humans, capable of identifying and interpreting such events.

In traditional contracts, if performance becomes impossible due to unforeseen events (such as force majeure), parties can suspend their obligations or terminate the contract. Smart contracts, however, lack this flexibility unless force majeure or similar contingencies are coded into them. Under Luxembourg law, a party might be legally justified in halting performance due to impossibility, but a smart contract would continue to execute regardless. This raises the question of how Luxembourg courts would handle disputes where the smart contract continues to enforce obligations even after they become legally impossible to fulfill.

46 Luxembourg Civil Code, articles 1134 and 1148

47 Loi du 5 août 2005 sur les contrats de garantie financière, article 11(1); article 13–1 (2)

48 Luxembourg Civil Code, article 1134

49 Delphine Gomes, Constantin Iscru and Azadeh Djazayeri, 'New Luxembourg Reorganization Law and Financial Collateral arrangements' (2024) *Insight Out* 28.

III. Regulatory pathways for enabling securities through smart contracts under Luxembourg law.

Although, while smart contracts are fundamentally considered immutable, parties may need to amend or update the contract while it is activated. However, it requires it to have clearly defined mechanisms for amendments, allowing the parties to make updates when mutually agreed upon. The contract should include clauses that enable updates to reflect changing circumstances, regulatory requirements, or the evolution of the parties' agreements. For certain securities or financial agreements, parties may wish to build governance mechanisms into the contract. These mechanisms allow for dispute resolution, amendments, or modifications to be carried out through a vote or mutual agreement of the involved parties. In this perspective, it is necessary to adopt principles allowing the drafting of securities through smart contracts (A), before addressing the requirements that should be addressed by Luxembourg's norms (B).

A. Drawing guidelines from the intersectionality of Smart contracts and securities

This article has highlighted the challenges stemming from forming and executing securities through smart contracts under Luxembourg law. This requires the setting up of principles regarding the formation of securities (1) and principles for executing them (2).

1. Guidelines for forming the securities with a smart contract.

To use a smart contract for the formation of the security entails challenges for the parties. However, these challenges can be overcome, especially while considering several principles by which the smart contract should be framed by.

Firstly, because it aims at forming a security and because of its nature as a contract, it is undoubtful that it would be bound by article 1108, regarding the principles of contract drafting. The consent of the parties to resort to a smart contract must be explicit and clear, for as it was demonstrated above, its form has direct and effective impact on its execution. Thus, despite what is commonly encountered in security law, not only the consent must be carried out regarding the contractual clauses of the security, but also on the

form of the contract. One could also generalize such approach to all smart contracts under Luxembourg law, since the form intrinsically impacts the execution of the contractual clauses.

Secondly, the smart contract must show upstream guarantees that the DLT it is relying on properly functions. These guarantees can be obtained through a technical notice given to the debtor gathering clear and precise documentations on how the smart contract works, what are the failsafe built inside the DLT and what is the extent of automaticity it relies on. For instance, what is the degree of human intervention allowed in this context, especially considering the challenges stemming from the market environment, force majeure or other external events that could influence the execution of the smart contract. These upstream guarantees are a conduit toward consent, and the clearer this technical information are, the better the consent would be.

The formation of the smart contract mainly regards upstream guarantees and the consent toward using the form of a smart contract. However, several more principles should be adopted in the context of the execution of the security, especially considering the coding and the triggering of the transfer of the collateral.

2. Guidelines for executing securities with a smart contract.

Smart contracts must be drafted in a manner that allow contingencies to address events like force majeure, otherwise, their rigid execution could cause unfair or uncoherent outcomes. The pathways for avoiding such outcomes to arise during the execution of the security.

Firstly, parties to a smart contract can attempt to resolve some of these challenges by coding specific contingencies into the contract itself during the formation phase. For instance, they could include provisions for recognizing and enforcing a force majeure event, impossibility of performance, or the need to relinquish the execution to a human oversight in certain situations.[50] Parties could code conditions that temporarily pause or terminate the contract execution if certain unexpected events occur, such as natural disaster, interruption of the market. This would mirror the traditional force majeure principle recognized under Luxembourg law further allowing human intervention if an unforeseen event arises, providing a layer of

50 Eric Tjong Tjin Tai, 'Force Majeure and Excuses in Smart Contracts' (2018) 26(6) *European Review of Private Law* 787–804, 799.

oversight over automated performance. Smart contracts should also rely on established oracle mechanisms, when external data feeds or third-party confirmations are required before executing certain key actions or to ensure that triggered events are appropriately recognized. The immutability of the smart contract thus calls for upstream and explicit encoding, to ensure the execution of the contract remains compliant with the general principles of contract law and the specificities of the security it is carrying**.** However, this requires a complex cooperation between lawyers and the encoders of the smart contract. This relationship is not yet defined under Luxembourg law; however, one could extrapolate from other regulations establishing an intersection of technologies and law. In such frameworks, the creditor relying on the smart contract needs to ensure that the development of such technology will ensure his compliance under the adequate rules of contractual relationship.

Secondly, a practical approach under Luxembourg law might lead to use hybrid smart contracts for the more complex securities, which combine traditional legal agreements with smart contract elements.[51] This would allow for greater flexibility by having certain parts of the contract governed by traditional rules intervention while automating less complex or mechanical aspects through smart contracts. Hybrid contracts allow parties to benefit from the efficiency of automation while retaining flexibility in areas that require human intervention or legal discretion. For example, if the loan is secured by cryptocurrencies, the smart contract automatically would transfer the cryptocurrencies to a multi-signature wallet controlled by both the creditor and the debtor. The debtor pledges cryptocurrency worth €500,000 as collateral. The smart contract automatically locks this collateral in a multi-signature wallet. The creditor disburses a €300,000 loan in the context of an over-collateralization, notably due to the volatility of the pledged asset. The debtor makes regular payments, which are automatically deducted from their account by the smart contract. Payments are recorded on the blockchain, and the smart contract updates the loan balance. Due to market fluctuations, the value of the cryptocurrency chosen as collateral falls to €250,000, triggering an automated margin call. The smart contract sends a notification to the debtor to provide additional collateral in order to match the original value. The debtor could contact the creditor and

51 Carlos Molina-Jimenez and Sandra Milena Felizia, 'On the Use of Smart Hybrid Contracts to Provide Flexibility in Algorithmic Governance' (2024) 6: e8 *Data & Policy Cambridge University Press* 1–11, 6.

requests more time to provide additional collateral. Thus, the creditor, using the hybrid contract's provisions, agrees to pause the liquidation process by activating the smart contract's kill switch allowing either party, or a neutral third party, to pause or halt the execution of the contract in certain circumstances Moreover, relying on more automaticity, a kill switch could be triggered by predefined events, such as one party notifying the other of a legal or technical issue. The smart contract would halt, and parties could address the matter through traditional legal channels before execution continues. After the borrower successfully repays the loan, the smart contract automatically releases the collateral back to the borrower. In case of default, if no agreement is reached, the smart contract automatically transfers the collateral to the lender according to the terms agreed upon in the traditional contract.

B. Regulatory perspectives under Luxembourg law

Due to the technical challenges associated with smart contract used for security purposes, one should consider an appropriate regulation. However, such regulation should take different paths in order to approach it adequately. Two levels can be considered. Firstly, a general and legal approach for ensuring a fundamental layer of principles applying to smart contracts used for security purposes. Secondly, a specific approach needs to be undertaken by the professionals when the drafting occurs,

Firstly, a further Blockchain law should open the recourse to DLT as a mean to form and execute security. This opening follows the legal logic that Luxembourg has established considering such technology and would also need to be accompanied with fundamental guarantees adequate for such smart contracts. This article highlighted the need to recourse to unavoidable features to ensure the coherence of the smart contract with security law. Among them, the recourse to an agreed oracle, upstream compliance with the general principles of contract law as stated in article 1108, and the technical recognition of contract execution principles, such as force majeure. What such legal regulation needs to consider, is the framing of principles needed for ensuring compliance between the general principles of contract law with smart contracts used for security purposes. Such an approach should also consider ensuring the parties have clearly established the triggering event and its dependence on an agreed oracle, to limit the risk of undue financial risk to the debtor.

Secondly, a more specific approach needs to be undertaken by the parties in order to fill the gap that would not be appropriate to address at the level of the law. This would mostly regard the choice of the appropriate security depending on the business relationship between the parties. It would be challenging for a law to address every technicality that the parties must have recourse to, given the freedom the parties should enjoy in drafting the smart contract. Nevertheless, these requirements should be addressed to offer technical safeguards. In this approach, compliance with the specific requirements of a specific security needs to be upheld by the parties foremost. They would need to ensure that the execution of the smart contract rigorously follows the requirements of the security they chose. Hence, such precise compliance at the technicality level mainly regards the securities enshrined in the law of 2005 for its requirements answer both formal and substance ones. In this phase, the legal requirements must be carefully implemented into technical standards by which the smart contract will abide. It is difficult to equate a general approach for this aspect, for it would be up to the parties to decide of the triggering event, the choice of the collateral, the terms of the security. What is crucial in this aspect, is to ensure that the will of the parties, and the execution of the security thereof, is ensured through the smart contract.

IV. Conclusion

While Smart contracts used for forming and executing securities can streamline the process of establishing and enforcing a security, its implementation can be challenging under Luxembourg law. Its features call for an upstream awareness from the creditor to answer both technical and legal requirements ensuring a compatibility with Luxembourg law. Moreover, while some securities can be more adapted for a full-fledged smart contract, the resort to hybrid ones should also be considered to avoid a full and unbeneficial automation of the contract leading to uncoherent outcomes. Creditors need to address what security would be adapted to either a smart contract or hybrid smart contract, the latest providing more room for human intervention than the first. The regulative perspective must spouse the flexibility offered by securities and the principles that should frame the formation and execution of smart contracts.

Creditors must also ensure of the opportunity of resorting to a smart contract, depending on the security they aim at integrating in its function-

ing. A coherence must occur during the drafting of the smart contract to ensure that it is adapted to the nature of the security and the business relationship with the debtor.

A Human-Centric Approach Through Rights: The *Missing Dialogues* between Rights, Risks, and Remedies under the EU's Artificial Intelligence

Elif Biber & Herwig C.H. Hofmann

Abstract: A human-centric approach to the regulation of AI must be built on individual rights, which in turn need to be enforceable through adequate remedies. This chapter argues that the EU's Artificial Intelligence Act (AI Act), despite its explicit commitment to a human-centric and rights-based approach to artificial intelligence (AI), fails to establish a meaningful relation between fundamental rights and the risk-based regulatory framework on which it is built. While the AI Act repeatedly invokes the protection of fundamental rights as a core objective, this article demonstrates that such references remain largely abstract and insufficiently translated into enforceable legal mechanisms capable of addressing the concrete societal impacts of AI systems and thereby hampering a truly human-centric orientation of the regulatory regime it built. This is at the heart of enforcement difficulties, weakening the regulatory power of the AI Act. The chapter develops this argument in three steps. First, it situates the AI Act within the broader evolution of the EU's digital acquis and its long-standing normative commitment to human dignity, democracy, and the rule of law. It shows that, notwithstanding this tradition, the AI Act adopts a predominantly technocratic governance model that tends to conflate the protection of fundamental rights with technical risk management and product-safety compliance. Secondly, the analysis examines the AI Act's mandatory requirements, focusing on human oversight and the Fundamental Rights Impact Assessment (FRIA). It argues that both instruments, while formally rights-oriented, are structurally ill-equipped to capture the contextual, systemic, and lived impacts of AI, and risk devolving into checklist-based or symbolic safeguards. Thirdly, the chapter analyses the Act's review and conformity-assessment procedures, highlighting how extensive reliance on self-assessment and private actors undermines accountability, weakens access to justice, and obscures the distinctive normative status of fundamental rights. Finally, the chapter advances concrete pathways to bridge this missing dialogue between fundamental rights and the legal architecture intended to secure their effective realisation in practice, underscoring the need to strengthen access to justice. Together, these elements are presented as indispensable to operationalising and giving substantial effect to the EU's human-centric vision of AI governance.

I. Human-Centeredness, Fundamental Rights and AI

European legal frameworks addressing digital technologies place strong emphasis on the centrality of the human factor and fundamental rights in

an era increasingly shaped by automation and artificial intelligence (AI).[1] Although still evolving, the EU's digital legal architecture demonstrates its commitment to European values and fundamental rights, which serve as the guiding constitutional principles for balancing interests through legislation and regulation. By contrast, compared to other leading regulatory models,[2] the EU explicitly seeks to place individuals and their fundamental rights at the centre of digital governance, in an attempt to reinforce its broader tradition of rights protection and European values. This orientation is not merely a regulatory preference but a normative stance, positioning Europe as an advocate of a rights-based, human-centric vision of digital transformation. The approach to equate the human-centric approach with fundamental rights is well established. The HELG identified it as the attempt to '*ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights…*'[3] However, the realisation of the human-centric vision based on rights remains work in progress. Based on the necessity to link a human-centric approach with individual rights, this chapter points at the need for robust review and enforcement mechanisms in order to ensure

1 The most relevant legal instruments are Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) [2016] OJ L 119/1; Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (Digital Services Act) [2022] OJ L 277/1; and the Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L 1689/1.

2 For a broad, comparative overview, see: Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford University Press 2023).

3 European Commission, 'Ethics Guidelines for Trustworthy AI', High-Level Expert Group on Artificial Intelligence (HLEG-AI), 8 April 2019, available at op.europa.eu/en/publication-detail/-/publication/d3988569–0434–11ea-8c1f-01aa75ed71a1 (accessed 13 March 2026) 37.The Guidelines define the human-centric approach as follows '*The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come.*'

realisation of what was articulated in e.g. the Commission's *European Declaration on Digital Rights and Principles for the Digital Decade* (of January 2022)[4] setting out a comprehensive vision of Europe's digital transition, framed around the principle of 'putting people at the centre.'[5] It reaffirmed that digital rights and freedoms must be respected and upheld online, with the same degree of protection as they are offline in the analogue world – thereby establishing a clear link between digital governance and the EU's constitutional traditions of human dignity, democracy, and the rule of law.[6]

Nonetheless, while this normative commitment is both ambitious and forward-looking, a substantial, and we might argue, widening gap remains in terms of the concrete implementation and operationalisation of the European human-centric rights-based approach.

Bridging this gap requires not only regulatory innovation and institutional capacity-building, but also procedural innovations that translate abstract values into enforceable legal principles.

This chapter explores the necessary steps to consolidate and concretise the European human-centric, rights-based digital framework, with a focus on how such a model can be rendered both practically effective and globally influential. It first focuses on what we refer to as the dynamics of the *missing dialogue*[7] between rights, risks and remedies under the EU's Artificial Intelligence Act (AI Act).[8] It then focuses on the legal pathways to bridge this missing dialogue.

4 European Commission, *European Declaration on Digital Rights and Principles for the Digital Decade* COM(2022) 28 final, 26 January 2022.

5 Elif Biber, 'Art. 2 The Principle of Solidarity and Inclusion' in Rossano Ducato, Alain Strowel and Enguerrand
Marique (eds), *The Declaration on European Digital Rights and Principles for the Digital Decade: A Commentary with a Legal Design Perspective* (forthcoming).

6 European Commission, *European Declaration on Digital Rights and Principles for the Digital Decade* COM(2022) 28 final, 26 January 2022, Preamble, para. 3 ('*The digital transformation should not entail the regression of rights. What is illegal offline, is illegal online*').

7 For an earlier discussion on the *missing dialogue* between rights and risks see Elif Biber, *A Rights-Based Inter-Legal Approach to Artificial Intelligence*, (Hart Publishing 2026).

8 See the AI Act in (n 1).

II. A Missing Dialogue between Rights, Risks, and Remedies

The EU's AI Act is primarily grounded in the risk-based regulatory framework, as also underlined in Recital 26 of the AI Act, which claims that proportionate and effective binding rules for AI systems require that 'the type and content of such rules' be tailored to 'the intensity and scope of the risks that AI systems can generate.'[9] Within the context of the AI Act, the risk-based framework is expressly tied to risks affecting health, safety, and fundamental rights. However, while the Act repeatedly invokes the notion of fundamental rights, its references to the fundamental rights remain largely generic, offering limited guidance on how specific rights are to be addressed within the regulatory scheme and by specific rules.

The engagement with fundamental rights does not appear to run counter to the emphasis on fundamental rights and a human-centric approach that predates the AI Act. For instance, in 2017, the European Council called on the EU to adopt a forward-looking, proactive approach to AI, emphasising the need to act urgently while safeguarding data protection, digital rights, and ethical standards. It further invited the European Commission to develop a distinctively European approach to AI.[10]

The Commission subsequently issued the draft and final versions of the Ethics Guidelines for Trustworthy AI in 2018 and 2019, also referring to AI systems as *socio-technical* systems rather than merely as products.[11] These Guidelines articulate a European, human-centric vision of AI governance, positioning the protection of human values as a central regulatory objective.[12] They underline the centrality of fundamental rights and human

9 Recital 26 of the AI Act in (n 1).

10 European Council, *European Council meeting (19 October 2017) – Conclusions* EUCO 14/17 (2017) 8 https://www.consilium.europa.eu/media/21620/19-euco-final-conclusions-en.pdf accessed 1 September 2025. Responding to this call, the Commission published a Communication in April 2018, which sought to lay the foundations for an ethical and legal framework for AI that reflects the Union's core values and aligns with the EU Charter of Fundamental Rights in European Commission, *Communication on Artificial Intelligence for Europe* COM(2018) 237 final, 25 April 2018; Charter of Fundamental Rights of the EU [2012] OJ C 326/391.

11 *Ethics Guidelines for Trustworthy AI* (n 3) 2 ('*... providing guidance on how such principles can be operationalised in socio-technical systems').*

12 European Commission, *Ethics Guidelines for Trustworthy AI*, High-Level Expert Group on Artificial Intelligence, 8 April 2019 https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 accessed 1 September 2025.

dignity throughout the development, deployment, use, and monitoring of AI systems, while also recognising the relevance of environmental sustainability and intergenerational equity. However, the AI Act left its practical implications largely underdeveloped with terms such as 'trustworthy AI', 'ethical AI' or 'human-centric AI' lacking contour and 'the grand formula' of human-centric, secure, trustworthy and ethical AI all ending up as more or less loose references to fundamental rights.[13] With Micklitz we thus find that the AI Act's proclaimed human-centrism suffers from important gaps. Not only is the meaning of the term human-centric underdetermined, as the European digital policy legislation does not explicitly confirm that 'human control over AI systems is a necessary requirement for protection of human dignity.' Also, there is a missing link between the human-centric formula and the real-world, engaging with the factual side, the concrete impact of AI systems on society.[14]

We argue in this chapter that one reason for the gap between the EU's human-centric vision and the concrete societal impact of AI systems is the Act's heavy emphasis on highly complex technical procedures and obligations. This focus leaves only a narrowly defined space for the effective protection of fundamental rights. It limits the extent to which the human-centric approach translates into meaningful safeguards in practice, obscures the dynamics underlying the societal impact of AI, and, hence, results in a lack of *dialogue* between rights and risks. In this chapter, we discuss how a more meaningful dialogue between rights and risks can only take shape once effective legal pathways are established – pathways capable of making the social impact of AI both visible and open to debate. This allows us to link rights, risks and remedies.

1. The Missing Dialogue – an Illustration

According to its Recitals, the protection of fundamental rights is identified as a primary objective of the regulation[15] referring to 'fundamental rights'

13 Hans-W Micklitz, *The Role of Standards in Future EU Digital Policy Legislation: A Consumer Perspective*, commissioned by ANEC and BEUC, July 2023.

14 ibid. In this regard, Micklitz underlines that *'the call for humans to have the last word requires defining red lines which cannot be crossed in technical standardisation'*.

15 See article 1 of the AI Act in (n 1), which states that *'the purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level*

more than ninety times, signalling a deliberate and sustained emphasis on rights-based concerns. This emphasis indicates that, even within a predominantly product-safety-oriented regulatory framework, the AI Act treats the protection of fundamental rights not merely as a symbolic reference, but as a normative commitment and a core regulatory aim. The limitations become particularly apparent in the Act's tendency to merge rights protection with safety standards.[16] An absence of meaningful dialogue does not stem from a failure to reference fundamental rights, but rather from a failure to accord them proper engagement within the regulatory framework. However, it is critical to recognise, that as the Ethics Guidelines emphasise, that AI systems are *socio-technical* systems rather than merely as products.[17]

By addressing matters predominantly in terms of narrowly defined technical risks, the AI Act underplays its problematic nature with respect to the core requirements of fundamental rights balancing and limitation as expressed under article 52(1) of the EU Charter of Fundamental Rights (CFR). The problem with this risk-based approach lies, in part, in the fact that the risks to a particular AI system's compliance with fundamental rights can seldom be fully identified at the time of its conception or even at deployment.[18] Technological risk evolves continuously as use cases diversify, tools are combined in novel ways, information flows diversify, and data can be recombined with increasing sophistication. The rising proliferation of non-binding standards and guidance from diverse bodies and sources as a remedy has not been able to address the gaps.[19]

---

of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation.' See also Elif Biber, 'Machines Learning the Rule of Law: EU Proposes the World's First Artificial Intelligence Act' (Verfassungsblog, 17 December 2025), verfassungsblog.de/ai-rol/ (accessed 3 January 2026).

16 ibid, 9. See also theories of fundamental rights in Ernst-Wolfgang Böckenförde, 'Grundrechtstheorie und Grundrechtsinterpretation' (1974) 27 *Neue Juristische Wochenschrift* 1529–1538 (presenting five fundamental rights theories, namely the liberal, the institutional, the democratic-functional, the welfare-state, and the value theory of fundamental rights).

17 *Ethics Guidelines for Trustworthy AI* (n 3) 2 (*'... providing guidance on how such principles can be operationalised in socio-technical systems').*

18 David Collingridge, *The Social Control of Technology* (St Martin's Press 1980) 19.

19 Lisette Mustert and Cristiana Santos, 'The European Data Protection Board – a (non)consensual and (un)accountable role?' (2025) 59 *Computer Law & Security Review* https://www.sciencedirect.com/science/article/pii/S2212473X25000896 accessed 11 January 2026.

But serious questions remain about whether the pronounced reliance on technical, highly complex procedures and obligations contributes to a meaningful dialogue between rights and risks. Instead, the risk-based and standard-oriented approach entangles regulatory enforcement, creates legal uncertainty about the nature and extent of obligations, and blurs responsibilities for compliance with requirements set out in the legal framework.

2. The Human Oversight and the Fundamental Rights Impact Assessment Obligations

The Act's tendency to conflate the protection of fundamental rights with technical safety standards also manifests in its mandatory requirements concerning human oversight and fundamental rights impact assessments.

Example One: Human Oversight

An example is the requirement of human oversight, set out in article 14 of the AI Act. This provision requires the implementation of mechanisms for human oversight with the explicit objective of safeguarding individuals' fundamental rights.[20] In this context, human oversight is conceived as a means to prevent or mitigate risks to health, safety, or fundamental rights that may arise from the deployment of high-risk AI systems, whether such systems are used in accordance with their intended purpose or under conditions of reasonably foreseeable misuse, particularly where such risks persist despite compliance with other regulatory requirements.[21] Furthermore, article 14(4)(d) explicitly recognises the significance of individual autonomy by empowering the person tasked with oversight to decide, in specific circumstances, not to deploy a high-risk AI system or to disregard, override, or reverse its output.[22] Notwithstanding the questionable technical feasibility of these specific requirements, the provision of this specific technical product-safety-based remedy does not adequately account for the fundamental cognitive differences between human decision-makers and

20 See the requirement of human oversight in Germany in Paul Nemitz and Eike Gräf, 'Artificial Intelligence Must Be Used According to the Law, or Not at All' (Verfassungsblog, 2022) https://verfassungsblog.de/roa-artificial-intelligence-must-be-used-according-to-the-law/ accessed 15 December 2025.

21 Article 14(2) of the AI Act in (n 1).

22 Article 14/(4)(d) of the AI Act in (n 1).

AI systems.[23] Therefore, we are not only confronted with the questionable remedial power of article 14 AI Act's specific requirements, we are also confronted with a lacking link between the obligations under article 14 and the realisation of the possible limitations to individual rights arising from automated systems using AI.

This observation also arises from a temporary aspect. The faster technological development progresses, and the less foreseeable future technological innovations are, the more 'timeless' the regulatory framework must be to project regulatory power into the uncertain future. The 'half-life' of highly specific technological regulation will be all the shorter, the faster the technological innovation cycle turns. Therefore, to confront the unforeseen and unforeseeable, a return to basics, including regarding the rights to be protected, is preferable to highly prescriptive requirements. The AI Act's article 14 human oversight obligations date back to the early drafting of its regulatory content, nearly a decade ago. At that time, today's possibilities of AI were not understood, as little as we can expect to anticipate the specific risks that a given AI-based technology will pose a decade from now. We do know, however, now and in the future, that fundamental rights must be protected. This amounts to the real futureproofing of law and regulatory content through constitutional steering.

Translating this to human oversight under the article 14 AI Act, this means that, given contemporary technological developments and the increasing complexity of AI models, it is often unrealistic to expect human overseers to meaningfully comprehend, scrutinise, and, if need be, correct the outputs of AI-driven decision-making processes. The systemic nature of such systems renders comprehensive oversight impracticable: human operators are generally unable to monitor the full scope of data collection, mining, and processing activities, nor can they reliably interpret or validate the frequently opaque outputs generated by complex models.[24] Meaningful oversight would presuppose an extensive level of transparency and docu-

23 Joseph Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (W H Freeman and Company 1976) 7. See also a debate on this issue in Manuel Alfonseca et al, 'Superintelligence Cannot Be Contained: Lessons from Computability Theory' (2021) 70 *Journal of Artificial Intelligence Research* https://jair.org/index.php/jair/article/view/12202 accessed 13 September 2025.

24 On the inadequacy of human oversight see Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 *Computer Law & Security Review* 1–22 (demonstrating the limitations of individuals in performing the required oversight role and advocating for institutional oversight mechanisms). See also Ben Green and Amba Kak, 'The False Comfort of Human Oversight as an

mentation regarding data processing practices that are rarely available in practice today. It would also require human overseers to be able to engage in independent, de novo decision-making to assess AI-generated outputs against alternative outcomes and reconstruct the underlying logic and reasoning of automated decision-making processes – an expectation that is often unattainable in real-world settings.

Furthermore, the fundamental cognitive differences between humans and AI systems risk entrenching what the Act and parts of the literature refer to as 'automation bias', rather than mitigating it. The AI Act explicitly warns those responsible for oversight to remain 'aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system.'[25] This warning reflects concern about this phenomenon, namely an 'undue deference to automated systems by human actors that disregard contradictory information from other sources or do not (thoroughly) search for additional information.'[26] However, this characterisation implicitly locates responsibility at the level of individual cognition and judgment, thereby underestimating the structural and organisational conditions in which such deference arises. For instance, in many public and private institutional settings, the bias issue is more plausibly understood as systemic or organisational issues: humans are frequently deprived of the time, resources, technical access, or institutional authority necessary to meaningfully interrogate or reproduce automated decision-making processes. Moreover, organisational incentives may actively discourage deviation from algorithmic outputs, even where doubts exist.

If human oversight is to serve as a genuine safeguard rather than a symbolic one, it must therefore be made *de facto* feasible, not merely formally assigned. While the AI Act correctly acknowledges the risk of uncritical overreliance on automated outputs,[27] additional structural and procedural

---

Antidote to AI Harm' (Future Tense, 2021) https://slate.com/technology/2021/06/human-oversight-artificial-intelligence-laws.html accessed 13 September 2025.

25 Article 14(4)(b) of the AI Act in (n 1).

26 Saar Alon-Barkat and Madalina Busuioc, 'Human–AI Interactions in Public Sector Decision-Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice' (2022) *Journal of Public Administration Research and Theory* 7–8 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3794660 accessed 13 September 2025.

27 Over-reliance on the outputs of high-risk AI systems can give rise to significant harms. A stark illustration is provided by the Spanish government's deployment of the VioGén system, an algorithmic tool intended to assess the risk of recidivism in cases of gender-based violence. As documented by AlgorithmWatch, the system generated profoundly unreliable risk assessments. Of the fifteen women killed in

dimensions must be addressed if human oversight is to operate as an effective counterweight rather than a legitimising façade. Otherwise, it is not only foreseeable but likely that the excessive digitalisation, for example, of government services and the automation of decision-making will supplant the uniquely human capacity to exercise discretion informed by contextual sensitivity and measure.[28]

Example Two: Fundamental Rights Impact Assessments

Another example of a *missing dialogue* between rights and risks can be observed in article 27 of the AI Act, which imposes a mandatory obligation only on deployers – an obligation that may also extend to private entities that provide public services – to conduct a Fundamental Rights Impact Assessment (FRIA).[29] Impact assessment tools were initially developed as accountability tools for legislative and administrative procedures.[30] Injecting such an obligation into the AI Act's context of specific deployment of AI systems profoundly changes the nature of the tool. FRIA under

incidents of domestic violence in Spain in 2014 – each of whom had previously reported their aggressor – fourteen had been classified by VioGén as facing either a 'low' or 'non-specific' level of risk. See AlgorithmWatch, *Automating Society Report 2020* (2020) 7, https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf (accessed 13 September 2025). For a more detailed account of the system's design and operation, see José Luis González Álvarez and others, 'Integral Monitoring System in Cases of Gender Violence: VioGén System' (2018) 4(1) *Behavior & Law Journal* 29–40.

28 Sofia Ranchordás, 'Empathy in the Digital Administrative State' (2022) 71 *Duke Law Journal* 1341–1389 (examining the role of empathy as a key value in administrative law).

29 It is important to note that FRIA should not be understood in isolation, but rather as part of a broader ecosystem of impact assessment instruments, including Human Rights Impact Assessment (HRIA) and Data Protection Impact Assessment (DPIA) or Privacy Impact Assessment (PIA). The key features of the FRIA under the AI Act are (i) its ex ante nature, (ii) its mandatory character, (iii) its broad scope, which encompasses all relevant fundamental rights and not merely data protection issues, and (iv) the fact that it is complemented by the obligation on AI providers to assess fundamental rights impacts within the conformity assessment under article 9 of the AI Act. See also Recital 96 of the AI Act ('*Services important for individuals that are of public nature may also be provided by private entities*').

30 See e.g. Herwig C.H. Hofmann, *Automated Decision-Making (ADM) in EU Public Law*, in: Herwig C.H. Hofmann, Felix Pflücke (eds.), *Governance of Automated Decision Making and EU law* (Oxford University Press 2024) 1–32; Colin Kirkpatrick and David Parker (eds.), *Regulatory Impact Assessment – Towards Better Regulation?* (Edward Elgar 2007); Claire A. Dunlop, Claudio M. Radaelli (eds.), *Handbook of Regulatory Impact Assessment* (Edward Elgar 2016).

the AI Act constitutes an *ex ante*, rights-based assessment process aimed at identifying, analysing, and addressing the potential impacts of an AI system on fundamental rights before and during its deployment, and it extends beyond data protection considerations to assess how a specific AI use case, situated within its concrete social, institutional, and operational context, may restrict or adversely affect a broad range of fundamental rights and freedoms. In this sense, FRIA has been incorporated into the AI Act to serve as a preventive and iterative governance mechanism, requiring expert-based evaluation, contextual analysis, and reasoned documentation to inform design choices, deployment conditions, and risk management measures throughout the entire lifecycle of AI systems.[31] Notwithstanding these strengths, the FRIA framework under the AI Act relies predominantly on internal assessments conducted by AI deployers. No obligations to ensure the meaningful participation of rights-holders, affected communities, or civil society actors exist,[32] despite this participatory and exploratory screening being a central tenet of impact assessments as they emerged in their original public-law setting. This significantly limits the value of an impact assessment, since it relies on vastly reduced information.

This structural limitation risks reducing FRIA to a predominantly technocratic exercise, insufficiently informed by lived factual experiences, social power asymmetries, and contextual forms of vulnerability. Therefore, indirect, cumulative, or systemic impacts on fundamental rights may remain under-identified or inadequately addressed. As a result, although FRIA purports to strengthen *ex ante* accountability, the necessarily one-sided nature of this internal exercise constrains its ability to fully capture the societal dimensions of AI-related risks and ultimately undermines its potential as a meaningful, human-centric instrument of AI governance.

The concept of conformity assessment procedures, as referred to in the AI Act, applies to high-risk AI systems.[33] The AI Act combines internal assessment through impact assessment activities with external private monitoring. For example, as also examined below, Art 3(20) of the AI Act identifies 'conformity assessment' as demonstrating whether the requirements set

31 See a detailed analysis of this process in Alessandro Mantelero, 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template' (2024) 54 *Computer Law & Security Review* 1–18.

32 Margot E Kaminski and Gianclaudio Malgieri, 'Impacted Stakeholder Participation in AI and Data Governance' (2025) 27(1) *Yale Journal of Law & Technology*, 265.

33 Articles 3(20), 6(1)(b), and 43 of the AI Act n (1). Compare with articles 19 and 43 of the AI Act Proposal of 21 April 2021.

out in various standards and obligations relating to a high-risk AI system 'have been fulfilled'. That can be controlled (Art. 3 (21)AI Act) by a 'conformity assessment body', i.e. a 'body that performs third-party conformity assessment activities, including testing, certification and inspection' and needs to be confirmed by a conformity assessment declaration under article 47 AI Act. Here, self-regulation and self-assessment are combined with a form of private certification. But the problem arises that the amount of intervening norms – from standards to guidelines to assessment criteria risks to give false sense of compliance and risks producing a regulatory haze with lacking clarity of which norms are to be complied with, what is binding and what is only guiding, and whether compliance with the law and its objectives or predominantly compliance with certification procedures is the objective. As a result, the oversight framework assigns a particularly prominent role to private actors, not only with respect to the FRIA but more generally as to compliance with the safety requirements. This raises, on the one hand, questions regarding the legitimacy of delegating such extensive oversight authority to private entities, and on the other hand, concerns relating to their capacity to conduct sufficiently rigorous and comprehensive assessments of complex AI systems. These limitations are especially salient in relation to the evaluation of compliance with fundamental and human rights, an area in which private actors may lack both the mandate and the expertise to ensure adequate protection.[34]

3. The Missing Dialogue in Enforcement Procedures: Conformity Assessment Procedures and the 'Rights' under the AI Act

This discussion leads to the question of implementation and finally enforcement. The AI Act establishes a number of significant oversight bodies, including the 'notifying authority', 'market surveillance authority', 'conformity assessment body', 'post-marketing monitoring system', 'AI Office', and 'European Artificial Intelligence Board'. Pursuant to article 3(48), the first two of these are classified as 'national competent authorities'. The question of whether the creation of such a multiplicity of bodies constitutes an opti-

34 It should additionally be noted that these actors may commit categorical errors in identifying the right or rights affected, particularly where the legal characterisation of harms requires normative judgment beyond technical assessment. For a detailed analysis see Elif Biber, *A Rights-Based Inter-Legal Approach to Artificial Intelligence* (Hart Publishing 2026).

mal institutional design, and whether the procedural interactions among them are appropriately structured, falls outside the scope of this article.

However, we do consider that article 43(2) and Annex VI of the AI Act, which establishes the mechanisms for assessing compliance with the mandatory requirements, weakens the overall robustness of the regulatory framework. Under these provisions, the majority of compliance reviews are conducted through internal 'conformity assessment procedures' performed by the providers themselves – an approach adapted from EU product safety law.[35] In practical terms, this regulatory design entrusts providers with the responsibility of self-assessing whether high-risk AI systems satisfy the essential requirements laid down by the AI Act, with the exception of high-risk AI systems deployed in the field of biometrics, which are subject to a conformity assessment carried out by an independent 'notified body'.[36]

A similar concern arises with regard to standardisation bodies, as explicitly acknowledged by the European Commission, which underlines that, 'given the fundamental rights and data protection implications of the European standards and European standardisation deliverables requested, compliance with Union law on fundamental rights and Union data protection law should also be guaranteed.'[37] The use of private self-assessment mechanisms, together with the involvement of private certification bodies ('notified bodies'), creates conditions that are vulnerable to the abuse of power.

35 The conformity-assessment procedure is defined in article 2(12) of Regulation (EC) No 765/2008 as 'the process demonstrating whether specified requirements relating to a product, process, service, system, person or body have been fulfilled'. See Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 [2008] OJ L 218/30. See also European Commission, *The 'Blue Guide' on the Implementation of EU Product Rules 2016* (2016) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016XC0726(02)&from=EN accessed 13 September 2025, clarifying the conformity-assessment procedure and the CE marking. The AI Act similarly defines the procedure in article 3(20) as 'the process of demonstrating whether the requirements set out in Chapter III, Section 2 relating to a high-risk AI system have been fulfilled'. The conformity-assessment procedure is linked to the CE marking, defined in article 3(24) of Regulation (EU) 2024/1689 (n 1) as 'a marking by which a provider indicates that an AI system is in conformity with the requirements (…)'.

36 Recital 125 of the AI Act in (n 1).

37 European Commission, *Commission Implementing Decision of 22 May 2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence* C(2023) 3215 final, recital 15 https://ec.europa.eu/transparency/documents-register/detail?ref=C(2023)3215&lang=en accessed 17 December 2025.

Similar regulatory approaches have been applied in other policy areas, equally subject to critical review.[38] Ultimately, the delegation of significant oversight responsibilities to private companies raises legitimate concerns regarding independence and the risk of conflicts of interest.

A further issue that reinforces pre-existing structural imbalances of power concerns the lack of a clearly defined and accessible avenue of redress for individuals affected by AI-driven systems. While the AI Act explicitly positions itself as a regulatory instrument designed to uphold European values and protect fundamental rights amid rapid technological development, there remains a substantive risk that persons harmed by the deployment of prohibited or high-risk AI systems may, in practice, be deprived of effective remedies.

Notably, the AI Act contains only two provisions that explicitly frame entitlements as 'rights': *'right to lodge a complaint with a market surveillance authority*' under article 85 and *'right to explanation of individual decision making*' under article 86. article 85 grants both natural and legal persons the ability to submit a complaint where they have grounds to believe that the Regulation has been infringed, providing that *'any natural or legal person having grounds to consider that there has been an infringement of the provisions of this Regulation may submit complaints to the relevant market surveillance authority.*'[39] Given that the Act is primarily addressed to AI providers and deployers and is modelled on the conceptual architecture of EU product safety law, the formal recognition of such a right constitutes, at least in principle, a noteworthy step towards accountability. However, the practical effectiveness of this mechanism remains an open question. The highly technical and complex nature of the AI Act may significantly impede individuals' ability to identify, interpret, and substantiate what constitutes an 'infringement' of its provisions.[40] This informational and

38 See the discussion of these approaches in pharmaceutical regulation and the regulation of medical devices: Sylvia Kierkegaard and Patrick Kierkegaard, 'Danger to Public Health: Medical Devices, Toxicity, Virus and Fraud' (2013) 29(1) *Computer Law & Security Review* 14 (arguing that the EU has prioritised the health of the single market over the health of the community). See also Hannah van Kolfschooten, 'EU Regulation of Artificial Intelligence: Challenges for Patients' Rights' (2022) 59 *Common Market Law Review* 106 (arguing that the Proposal focuses on companies rather than patients).

39 Article 85 of the AI Act in (n 1).

40 This provision should be examined by comparison to the GDPR. Article 57 GDPR assigns to each supervisory authority the task *'to handle complaints lodged by a data subject, or by a body, organisation or association in accordance with article 80, and in-*

epistemic asymmetry risks rendering the right to lodge a complaint largely abstract, particularly for those lacking technical expertise or access to legal and computational resources.

In this context, Recital 58 assumes particular significance, as it explicitly acknowledges the potential of certain AI systems to undermine core procedural guarantees and emphasises the need for accountability and effective redress. Although formulated specifically with respect to law enforcement, the considerations can be generalisable in that it links 'the exercise of important procedural fundamental rights, such as the right to an effective remedy and to a fair trial…' to the requirement that AI systems be 'sufficiently transparent, explainable and documented' in order 'to avoid adverse impacts, retain public trust and ensure accountability and effective redress.'[41]

In addition, article 86 of the AI Act introduces a right to explanation of individual decision-making for persons affected by decisions taken on the basis of outputs generated by certain high-risk AI systems.[42] Where such decisions produce legal effects or similarly significant impacts on an individual's health, safety, or fundamental rights, the deployer is required to provide 'clear and meaningful information' regarding the role of the AI system in the decision-making process and the main elements underpinning the decision. However, the practical value of this right has been widely questioned in academic discourse.[43] Most notably, the obligation to provide explanations is imposed exclusively on the deployer, while the AI provider – who possesses critical knowledge concerning the system's design, training

vestigate, to the extent appropriate, the subject matter of the complaint and inform the complainant of the progress and the outcome of the investigation within a reasonable period, in particular if further investigation or coordination with another supervisory authority is necessary'. In addition, the GDPR explicitly enshrines both the *'right to lodge a complaint with a supervisory authority'* and the *'right to an effective remedy against a supervisory authority'* in articles 77 and 78.

41 Recital 58 of the AI Act (n 1).

42 The right to explanation under the GDPR has been extensively discussed in legal literature. See early works in Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) *International Data Privacy Law* 76; Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 243.

43 Margot E Kaminski, Gianclaudio Malgieri and Paul De Hert, 'The Right to Explanation in the AI Act' (8 March 2025) U of Colorado Law Legal Studies Research Paper No 25–9 https://ssrn.com/abstract=5194301 accessed 8 January 2026.

data, and operational logic – is excluded from the explanatory framework.[44] Generally speaking, the knowledge about the information that was used to generate the decision, both in terms of what was taken into account and what was discarded, as well as the criteria under which such information was evaluated, remains a central element. General explanations about 'main factors' will generally be of only limited value in terms of the enforcement of rights.[45]

This situation leaves deployers, particularly in the public sector, unable to offer more than 'box-ticking' explanations – often just a brief note that a human reviewed the AI's output. Such explanations are unlikely to enable affected individuals to meaningfully challenge the decision or to assess whether the AI system contributed to discriminatory or otherwise unlawful outcomes. Furthermore, the right applies only to a narrow category of high-risk systems listed in Annex III and is expressly subordinated to existing rights under Union law, significantly curtailing its scope of application. Taken together, these limitations risk rendering the right to explanation largely symbolic, offering weak procedural protection rather than an effective remedy in practice. Such a situation, in fact, further exacerbates the existing lack of meaningful dialogue between rights-based considerations and risk-based assessments.

III. Bridging the Gaps : How to Concretise the European Human-Centric Approach to Artificial Intelligence

We must therefore bridge the gaps that lead to what we describe here as the missing dialogue regarding rights and their protection. Bridging the gaps is essential in order to fulfil the promise inherent in the human-centredness of human self-determination. What this means arises not just from legal thinking in terms of rights, but also of the preconditions of rights protection. Arendt, for example, reminds us that human rights emerge

44 Melanie Fink and Simona Demkova, 'The Constitutional Foundation of Explanation Rights in EU Digital Regulation: A Contextual Approach' (November 2025) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5923283.

45 Herwig C H Hofmann, 'Assessing Cyber Delegation in EU Public Law' in Herwig C H Hofmann and Felix Pflücke (eds), *Governance of Automated Decision Making and EU Law* (Oxford University Press 2024) 33–52.

predominantly 'through the capacity to act in a shared public realm',[46] making them not an inner state but a practice within a web of human relationships.[47] This understanding has profound implications for digital environments,[48] especially for their regulation in line with standards of human-centeredness. If the rights depend on appearance, contestability, and – we would argue due to the reality of the relation between many users and generally big providers, collective action – then the structures through which digital power operates inevitably acquire legal and political significance. Where not only the architecture of advanced digital technologies systematically restricts the individual and collective capacity to act, but also the law regulating digital actors, through self-regulation and abstract risk terminology, limits individuals' capacity to remedy the wrong. In this context, human-centeredness will necessarily remain out of reach.

On the other hand, a meaningful bridging of the gaps between risks, rights and remedies can emerge only once effective legal pathways are established – pathways capable of rendering the social impacts of AI visible and open to contestation and debate. These pathways preserve the accessibility of justice by making public participation and the protection of rights explicit, rather than allowing them to be obscured within a 'legal black box' of technical obligations produced by the legal system itself.

Furthermore, given that AI systems are still often characterised as black-box systems – producing outputs without transparent or comprehensible explanations[49] – the role of law becomes critically important in safeguard-

46 Charles Barbour, 'Between Politics and Law: Hannah Arendt and the Subject of Rights', in Marco Goldoni, and Christopher McCorkindale (eds) *Hannah Arendt and the Law* (Hart Publishing House 2012) 307.

47 Hannah Arendt, *The Human Condition* (The University of Chicago Press 1958) 204, available at https://archive.org/details/dli.ernet.528547/page/n1/mode/2up accessed 13 March 2026 ('Power preserves the public realm and the space of appearance, and as such it is also the lifeblood of the human artifice, which, unless it is the scene of action and speech, of the web of human affairs and relationships and the stories engendered by them, lacks its ultimate *raison d' etre*. Without being talked about by men and without housing them, the world would not be a human artifice but a heap of unrelated things (…).')

48 See Elif Biber, 'A Challenge for a *Primavera Digitale*: The *Phantom Influence* of Artificial Intelligence Systems', International Journal of Constitutional Law's Blog, *ICONnect* (New York University School of Law) 3 February 2026, available at: https://www.iconnectblog.com/a-challenge-for-a-primavera-digitale-the-phantom-influence-of-artificial-intelligence-systems/ (accessed on 3 February 2026).

49 Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).

ing the rule of law, ensuring accountability, and preventing opacity from undermining legal principles. Although the development and evaluation of AI require technical expertise, its far-reaching social consequences make empowering individual engagement indispensable[50] – an involvement that substantive and procedural legal frameworks must actively enable rather than marginalise in order to make a rights-based human-centredness possible. Courts play an essential role in developing an ideally broadly defined legal framework to new such challenges. One example for this, which has become pathbreaking in many ways, was the clarification of both the socio-technical dimensions and the legal aspects of personal data processing and the human involvement in decision-making processes[51] in the famous CJEU case *Google Spain v AEPD and Mario Costeja González*[52] – widely discussed in scholarship on digital constitutionalism as an example of the normative power of judicial interpretation in the digital context.[53] The case demonstrates how courts can shape the scope and content of fundamental rights in response to technologically mediated forms of decision-making and control, and adapt obligations to an ongoing technological situation and its interactions with societal needs and requirements.

In order to ensure this across the board, however, collective redress and a heightened role of NGOs in representing individual rights is necessary. Developing the procedural right of effective access to justice is therefore necessary to address the potential of societal harms posed by AI systems towards a real-life tool.[54] As with harm caused by physical persons, AI-

50 Lucile Ter-Minassian, 'Democratizing AI Governance: Balancing Expertise and Public Participation' (16 February 2025) arXiv https://arxiv.org/abs/2502.08651 accessed 6 January 2025.

51 Elif Biber, 'Between Humans and Machines: Judicial Interpretation of the Automated Decision-Making Practices in the EU' in Herwig C H Hofmann and Felix Pflücke (eds), *Governance of Automated Decision-Making and EU Law* (Oxford University Press 2024).

52 Case C-131/12 *Google Spain SL and Google Inc v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* EU:C:2014:317.

53 Giovanni De Gregorio, 'The Rise of Digital Constitutionalism in the EU' (2021) 19 *International Journal of Constitutional Law* 41–70; Giovanni De Gregorio, *Digital Constitutionalism in Europe: Reframing Rights and Powers in the Algorithmic Society* (Cambridge University Press 2022); Elif Biber, 'Digital Constitutionalism in Europe: Exploring a Constellation of Legalities' in Giovanni De Gregorio, Oreste Pollicino and Peggy Valcke (eds), *The Oxford Handbook of Digital Constitutionalism* (Oxford University 2025).

54 Natalie Smuha, 'Beyond the Individual: Governing AI's Societal Harm' (2021) 10(3) *Internet Policy Review*.

based data processing or decision-making must be subject to meaningful avenues for individual redress to remedy harm and hold relevant actors accountable.[55] Fundamental rights that individuals enjoy in the physical world must be equally safeguarded in online and digital environments.[56] This equivalence between the analogue and the digital world does not exclude taking into account and reflecting in procedural provisions and the distinctive features of advanced technological systems, as well as the structural and systemic forms of harm they may generate. The digital environment differs in fundamental ways from the physical one and will therefore require adapted interpretations of existing rights,[57] and possibly, where that is not sufficient, also the creation of new rights and freedoms tailored to AI's specific architectural conditions. For example, the first preliminary reference to the Court of Justice of the EU (CJEU) concerning the AI Act, submitted in November 2024, centred on the right to an explanation of individual decision-making, which arises from individual freedom, privacy and data rights but was also explicitly articulated in the AI Act itself.[58]

These developments go hand in hand with the increasing role of procedural rights in the CJEU case law, where, for example, the right to access to justice has received increasing prominence in the jurisprudence.[59] Given the realities of the provision of services and AI embedded services in software is often undertaken by large companies which creates uneven playing fields, ensuring judicial review of individual rights via collective redress is requirement to not only in theory but also in practice realise, develop and uphold human-centredness through the recognition and enforcement of

55 See a recently published blog post on the issue in Vladimir Apraxine, 'Complementing the AI Act with a New Right to Access to Justice: A Discussion on Participatory Governance and Redress in the AI Act' (23 December 2025) https://www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/complementing-the-ai-act-with-a-new-right-to-access-to-justice-a-discussion-on-participatory-governance-and-redress-in-the-ai-act accessed 6 January 2026.

56 Dafna Dror-Shpoliansky and Yuval Shany, 'It's the End of the (Offline) World as We Know It: From Human Rights to Digital Human Rights – A Proposed Typology' (2021) 32(4) *European Journal of International Law* 1249.

57 See e.g. Diana-Urania Galetta, Herwig C.H. Hofmann, *Evolving AI-based Automation – Continuing Relevance of Good Administration*, (2023) 48 European Law Review 617–635.

58 CJEU, Case C-806/24 *Yettel Bulgaria* (pending).

59 See Simona Demková, Herwig C.H. Hofmann, *General Principles of Procedural Justice*, in: Katja S. Ziegler, Päivi J. Neuvonen and Violeta Moreno-Lax (eds.), *Research Handbook on General Principles in EU Law*, (Elgar Publsihing, Cheltenham 2022), 209–226.

individual rights.[60] Lessons for the digital field especially arise from *ASG-2*, which demonstrated that the Court in Grand Chamber, reaffirming the importance of effective judicial protection under article 47 of the Charter, held that national rules render the exercise of the EU right to full compensation practically impossible or excessively difficult, were in violation of EU law, thus requiring collective redress norms in certain conditions. This was confirmed in the *Apple* Case, which was also a competition law case concerning digital markets. The case, brought by two Dutch consumer organisations against Apple concerned the commissions charged on its iOS-based App Store.[61]

Both cases underline the particular relevance of representing non-governmental organisations (NGOs) in representing and defending the public interest and the interests of numerous individual rights holders.[62] Through the publication of detailed analyses on AI governance and regulation, NGOs have brought visibility to well-documented instances of AI-related misuse and abuse. The relevance of NGO engagement is illustrated by the Reclaim Your Face campaign, a Europe-wide civil society initiative advocating for a prohibition on indiscriminate or arbitrarily targeted applications of biometric technologies.[63] Similarly, the importance of civil society action is underscored by the landmark *SyRI* judgment, widely regarded as one of the most significant technology-related court decisions globally, which was initiated through the rigorous efforts of human rights activists and non-governmental organisations.[64] This recognition reflects an understanding of NGOs as institutional intermediaries capable of enhancing accountability where individual access to justice is limited. In light of their demonstrated capacity to act effectively on behalf of individuals and in the public interest, especially in addressing societal harms arising from AI systems, it is imperative that AI governance frameworks incorporate

60 The role of collective redress is well recognized in the enforcement of competition law and the protection of rights therein (see e.g. recently Case C-253/23 *ASG 2 Ausgleichsgesellschaft für die Sägeindustrie Nordrhein-Westfalen GmbH v Land Nordrhein-Westfalen* EU:C:2025:40) also with respect to digital markets (see e.g. Case C-34/24 *Stichting Right to Consumer Justice and Stichting App Stores Claims v Apple Distribution International Ltd and Apple Inc* EU:C:2025:936).

61 *Apple* (n 60) paras 13–21.

62 Center for AI and Digital Policy, *Artificial Intelligence and Democratic Values Index* (2022) https://caidp.org/reports/aidv-2021/ accessed 8 January 2026.

63 Reclaim Your Face https://reclaimyourface.eu/ accessed 3 January 2026.

64 *The Hague District Court*, Case C/09/550982/HA ZA 18–388 (2020) paras 6.1 – 6.118.

adequate procedural mechanisms to enable and support such collective representation.

V. Concluding Remarks

The AI Act, despite its explicit and repeated commitment to fundamental rights and human-centric AI governance, currently requires developing a meaningful dialogue between risks, rights and remedies. While the AI Act represents an ambitious and unprecedented attempt to regulate AI through a binding, risk-based framework, using plenty of rights-language, but in with insufficient legal mechanisms through which those rights are made visible, contestable, and enforceable in practice.

The analysis has shown that this missing dialogue is particularly evident in the AI Act's mandatory requirements, review, and enforcement architecture. Human oversight, while formally recognised as a safeguard for fundamental rights, is framed in ways that underestimate the cognitive, organisational, and structural asymmetries between human decision-makers and complex AI systems. Similarly, the FRIA is a potentially powerful governance tool in the hands of public regulators, but its predominantly internal design within the private FRIA under the AI Act limits its capacity to capture the lived, cumulative, and systemic impacts of AI on individuals and communities.

The enforcement model of the AI Act further reinforces this disconnect. By relying heavily on self-assessment and conformity procedures derived from product-safety law, the Regulation entrusts private actors with substantial responsibility for evaluating compliance with requirements that extend far beyond technical safety and deeply into the domain of individual fundamental rights, not just to health and safety but also to data protection, freedom of expression, non-discrimination and many other more. This approach risks conflating conformity with accountability and obscuring the distinct legal status of fundamental rights as values whose limitation and balancing require justification, independent oversight, and effective avenues of redress. In doing so, the AI Act exacerbates power asymmetries and renders access to justice increasingly difficult for those affected by AI-driven decisions.

Against this background, the chapter has argued that linking rights and risks requires constructing explicit legal pathways that translate abstract human-centric commitments into enforceable legal realities. Strengthening

access to justice in the AI context, recognising the institutional role of civil society organisations, and enabling rights-sensitive judicial interpretation emerge as key elements in bridging the gap between what is and what should be possible. These mechanisms do not seek to replace technical regulation but to complement it by embedding democratic participation, public contestation, and normative reasoning within AI governance.

Ultimately, a genuinely human-centric approach to AI cannot be realised through technical compliance alone. It requires the construction of a regulatory regime that acknowledges AI systems as *socio-technical* phenomena with profound societal consequences and that places fundamental rights at the centre of governance, not merely as objectives, but as operative legal standards. As emphasised by the Fundamental Rights Agency, effective protection of fundamental rights requires establishing robust mechanisms, rather than relying predominantly on self-assessment checks conducted by providers.[65] Without such a recalibration, the promise of a human-centred AI regulation within the AI Act risks remaining aspirational, leaving the dialogue between rights and risks unresolved and the protection of fundamental rights insufficiently secured in the digital age.

65 EU Agency for Fundamental Rights, *Getting the Future Right: Artificial Intelligence and Fundamental Rights* (2020) https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf accessed 8 January 2026.

Contributors

Anna Moraiti, PhD, specialises in AI and criminal legal theory.

Arthur BIANCO, PhD, is an expert in technology law, tax law, and administrative law.

Elif Biber, PhD, Legal scholar in European public law and digitalisation at the University of Luxembourg, Head of Digital Rights at the Digital Constitutionalist, former academic visitor at the Oxford Institute for Ethics in AI, and author of *A Rights-Based Inter-Legal Approach to Artificial Intelligence* (Hart, Oxford 2026).

Grazia Bruzzese, PhD-candidate specialising in investment funds and financial criminal law.

Herwig C.H. Hofmann, Professor of European and Transnational Public Law, University of Luxembourg.

Isabella Lorenzoni, PhD, Luxembourg Competition Authority, is a legal expert in digital economy and regulation.

Kalliopi Terzidou, PhD, specialises in reforms of the judicial administration of EU-based courts through the adoption of AI systems and their impact on access to justice. She also works as a Legal Documentalist in the EUR-Lex and Legal Information Unit at the European Commission.

Théo Antunes, PhD, Audiovisual Independent Authority of Luxembourg (ALIA), specialises in law and artificial intelligence.

Zahra Yusifli, PhD, specialises on the intersection of labour law, artificial intelligence, EU digital regulation and regulating AI in the workplace. She has previously served as a Labour Law Officer at the International Labour Organization (ILO) in Geneva.

The contributors were supported by the Luxembourg National Research Fund (FNR) supported **Digitalisation Law and Innovation** (**DILLAN**, FNR PRIDE19/14268506).