jipitec

1| 2025

Volume 16(2025)
Issue 1 ISSN 2190-3387

Journal of
Intellectual Property,
Information Technology,
and Electronic Commerce
Law

www.jipitec.eu

2025

jipitec

Journal of Intellectual Property,
Information Technology and
Electronic Commerce Law

Volume 16 Issue 1 March 2025
www.jipitec.eu
contact@jipitec.eu

A joint publication of:

Prof. Dr. Thomas Dreier, M. C. J. (NYU)
KIT - Karlsruher Institut für Technologie,
Zentrum für Angewandte
Rechtswissenschaft (ZAR),
Vincenz-Prießnitz-Str. 3,
76131 Karlsruhe Germany

Prof. Dr. Axel Metzger, LL. M. (Harvard)
Humboldt-Universität zu Berlin,
Unter den Linden 6,
10099 Berlin

Prof. Dr. Gerald Spindler (†)
Dipl.-Ökonom, Georg-August-
Universität Göttingen,
Platz der Göttinger Sieben 6,
37073 Göttingen

Karlsruhe Institute of Technology,
Humboldt-Universität zu Berlin and
Georg-August-Universität Göttingen
are corporations under public law, and
represented by their respective presidents.

ISSN 2190-3387

Funded by
DGRI Deutsche Gesellschaft für Recht und Informatik e.V.

Table Of Contents

Editorial

Articles

Editorial

by **Axel Metzger**

Recommended citation: Axel metzger, Editorial, 16 (2025) JIPITEC 1 para 1.

1 With the issue published today, JIPITEC opens the 16th annual volume since its foundation in 2010. The issue brings together contributions from authors from various regions of Europe, from the Baltic States, Eastern and Central Europe to Italy and Portugal, thus demonstrating the truly European claim of the journal. The six articles have undergone a rigorous double-blind peer review process, with 18 further submissions rejected. This demonstrates the journal's high quality standards. The topics of the issue range from various aspects of the regulation of artificial intelligence (articles by Nuno Sousa e Silva, Mateus Correia de Carvalho and Stepanka Havlikova), to the legal coverage of health apps (Liga Svempe), to the Cyber Resilience Act and Open Source (Mattis van 't Schip) and to the regulation of push notifications under data protection law (Tristan Radtke). The issue closes with an Opinion of the European Copyright Society on Copyright and Generative AI. We hope you find this JIPITEC issue a stimulating read and invite you to submit manuscripts for the next issues!

Axel Metzger

The Artificial Intelligence Act: Critical Overview

by **Nuno Sousa e Silva** *

Abstract: This article provides a critical overview of the recently approved Artificial Intelligence Act. It starts by presenting the main structure, objectives, and approach of Regulation (EU) 2024/1689. Followed by a definition of key concepts, finally the material and territorial scope, as well asan examination of the timing of application, are analyzed. Although the Regulation does not explicitly set out principles, the main ideas of fairness, accountability, transparency, and equity in AI underly a set of rules of the regulation. This is discussed before looking at the ill-defined set of forbidden AI practices (manipulation and exploitation of vulnerabilities, social scoring, biometric identification and classification, and predictive policing). It is highlighted that those rules deal with behaviors rather than AI systems. The qualification and regulation of high-risk AI systems are tackled, alongside the obligation of transparency for certain AI systems, the regulation of general-purpose models, and the rules on certification, supervision, and sanctions. The text concludes that even if the overall framework can be deemed adequate and balanced, the approach is so complex that it risks defeating its own purpose of promoting responsible innovation within the European Union and beyond its borders.

Keywords: Artificial Intelligence; EU Law; AI Act

Recommended citation: Nuno Sousa e Silva, The Artificial Intelligence Act: critical overview, 16 (2025) JIPITEC 2 para 1.

A. Introduction

1 The rapid technological evolution of recent decades - generating a vast collection of digitized and accessible information (made possible by the Internet) and advances in terms of hardware and software - has allowed certain mathematical techniques (like machine learning) to become revolutionary. This is at the root of the dizzying developments in Artificial Intelligence that have taken place in the last few years.

2 However, despite the numerous advantages that this development brings,[1] a catastrophist tone has gained prominence.[2]

3 In the second decade of the 21st century, safety in Artificial Intelligence (hereinafter "**AI**") has established itself as an interdisciplinary branch of study, going beyond ethical considerations.[3] There

[*] Lawyer and Assistant Professor at the Portuguese Catholic University (Porto). E: nsilva@ucp.pt W: www.nss.pt.

[1] Among many others, the acceleration of drug development (J. Jumper et al., 'Highly accurate protein structure prediction with AlphaFold' *Nature* 596 (2021) pp. 583-589) and vaccines (A Sharma, et al. *Artificial Intelligence-Based Data-Driven Strategy to Accelerate Research, Development, and Clinical Trials of COVID Vaccine.* BioMed research international (2022)), the fight against climate change (J. Cowls, et al. 'The AI gambit: leveraging artificial intelligence to combat climate change-opportunities, challenges, and recommendations' in AI & Society 38 (2023) pp. 283-307) and the creation of new materials (Phil De Luna (ed.), *Accelerated Materials Discovery: How to Use Artificial Intelligence to Speed Up Development* (De Gruyter 2022)).

[2] Among the most influential works along these lines are Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (OUP 2014) and, earlier, Ray Kurzweil, *The Singularity Is Near: When Humans Transcend Biology* (Viking 2005). For a more balanced view, see Henry A Kissinger/Eric Schmidt/Daniel Huttenlocher, *The Age of AI: And Our Human Future* (Little, Brown and Company 2021).

[3] R.V. Yampolskiy, 'Artificial intelligence safety engineering: Why machine ethics is a wrong approach' in AAVV, *Philosophy and Theory of Artificial Intelligence* (Springer 2013)

are discussions regarding the transparency and explainability of decisions made by AI systems,[4] the potential for discrimination or injustice in the use of these systems,[5] and the challenges to control and align AI systems with human values.[6] There is a pressing need to guarantee the robustness and technical quality of AI.[7] The extractive practices of both data (some of it protected by intellectual property rights) and minerals and the energy consumption of AI are also a matter of concern.[8]

4 In recent years, lawyers and politicians have started to consider laws to deal with the multiple challenges of AI. The issues are complex and have a subatantial impact on fundamental rights (freedom, work and employment, privacy, equality and non-discrimination, democratic participation, access to justice, freedom of expression and information, political organization, environmental protection), civil and criminal liability, personal data protection, privacy and personality rights, intellectual property, competition law, environmental law, criminal law, tax law and administrative law.[9]

5 Although regulatory initiatives are taking place all over the world, the European Union has taken the lead.[10] On February 16, 2017, the European Parliament adopted a resolution with recommendations to the European Commission on civil law rules on robotics.[11] This resolution recognizes the dangers and opportunities of robotics and artificial intelligence and makes various suggestions for their regulation, urging the Commission to present a legislative

pp. 389-396.

[4] This is what is known as XAI (*explainable* AI). On the wider topic cf. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2016). Discussing the existence of a right to explanation under art. 22 GDPR, see the debate between Sandra Wachter / Brent Mittelstadt / Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' International Data Privacy Law, vol. 7(2) (2017) pp. 76-99 and Gianclaudio Malgieri / Giovanni Comande, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' International Data Privacy Law, 2017, Vol. 7(4) pp. 243-265. The majority of author seem to agree that under the GDPR there is no right to a detailed explanation of the decision, but only to a statement of its basic criteria and parameters (AAVV, *General Data Protection: art.-by-article commentary* (Hart C. H. Beck 2023) p. 541).

[5] Among many, Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Brown 2016); Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (NYU Press 2018); Meredith Broussard, *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech* (MIT Press 2023).

[6] See Brian Christian, *The Alignment Problem* (Atlantic Books 2020) and Stuart Russel, *Human Compatible: AI and the Problem of Control* (Penguin 2019).

[7] Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (Penguin 2017).

[8] Kate Crawford *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial* Intelligence (Yale University Press 2021).

[9] Books on the Law and/of AI have multiplied. Initially, the study (and the European Parliament's approach) focused mainly on robotics, and the general works include Ugo Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts* (Springer 2013); Alain Bensoussan/Jérémy Bensoussan, *Droit des Robots* (Larcier 2015) and Ryan Calo/Michael Froomkin/ Ian Kerr (eds), *Robot Law* (EE 2016). In fact, the tendency to focus analysis on robotics extended beyond law, as evidenced by Patrick Lin/Keith Abney/George A. Bekey (eds), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press 2011). These books mainly dealt with personality, crime, contracts and torts (liability). Others, such as Moisés Barrio Andrés (eds), *Derecho de los Robots* (Wolters Kluwer 2018), have gone further, also dealing with issues of employment law, financial and tax law, health law and its impact on the legal professions. Still under a perspective of Law and Robotics, but focusing on Artificial Intelligence, cf. Jacob Turner, *Robot Rules* (Palgrave 2019) and Ryan Abbott, *The Reasonable Robot* (Cambridge University Press 2020). In line with the more general trend, authors have come to prefer AI-centered analysis. More general books include Matt Hervey/Matthew Lavy (eds.), *The Law of Artificial Intelligence* (Sweet & Maxwell 2020); Woodrow Barfield / Ugo Pagallo, *Advanced Introduction to Law and Artificial Intelligence* (Edward Elgar 2020); Woodrow Barfield / Ugo Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar 2020); Jan De Bruyne / Cedric Vanleenhove (eds.), *Artificial Intelligence and the Law* (Intersentia 2021); Hoeren / Pinelli, *Künstliche Intelligenz - Ethik und Recht* (C. H. Beck 2022); and Charles Kerrigan, *Artificial Intelligence: Law and Regulation* (Edward Elgar 2022). Ebers/ Heinze/Krügel/Steinrötter, *Künstliche Intelligenz und Robotik* (C.H. Beck 2020) is noteworthy for its breadth and depth, with over a thousand pages of sectoral analysis. There are also empirical studies, critical theories and law and economics (e.g. Georgios Zekos, *Economics and Law of Artificial Intelligence* (Springer 2021)).

[10] Beyond the EU, on 17 May 2024 the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law was approved by the Council of Europe (“**CoE Convention**”). In the same month, the 2019 OECD guidelines (Recommendation on Artificial Intelligence) were revised (C/MIN(2024)16/FINAL). In the US, there is sectoral legislation, initiatives (e.g. USC 15 Chpater 19 - National Intelligence Initiative), state legislation and executive orders, but no general federal law has yet been passed. Some countries, such as Australia, Japan, Israel, Singapore and India, have followed *soft law* approaches, complemented by sectoral interventions. There have been some proposals for legislation, for example in Brazil and Canada. In July 2023, Peru adopted Law 31814 to promote the use of AI. For a follow-up on legislative and regulatory developments in this area, see. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker

[11] (2015/2103(INL)).

proposal on legal issues related to the development and use of robotics and Artificial Intelligence. Annexed to this document were recommendations on the content of such a proposal - including the definition of a robot, the creation of a registration system managed by a European agency, rules on civil liability, insurance and guarantee funds and the establishment of interoperability rules - and a "Robotics Charter", a voluntary code of conduct aimed at robotics researchers and *designers*. This 2017 resolution accelerated the discussion on legal issues related to artificial intelligence and robotics.[12]

6 In the following year, the Commission presented two communications "Artificial Intelligence for Europe"[13] and "Coordinated Plan for Artificial Intelligence"[14]. Resolutions, studies and reports followed and the "White Paper on Artificial Intelligence" presented by the Commission in February 2020 set the approach for the upcoming proposals.[15]

7 On October 20, 2020, the European Parliament adopted a resolution with recommendations to the Commission on the civil liability regime applicable to artificial intelligence.[16] This document contained the text of a draft regulation on liability for the operation of AI systems.[17] On September 28, 2022, the European Commission presented two proposals: a revision of the Product Liability Directive, which aims to replace Directive 85/374/EC[18] and a new Directive on the adaptation of non-contractual civil liability rules to artificial intelligence.[19] These are still under discussion.

8 However, the main regulatory approach to this phenomenon is the Artificial Intelligence Regulation, known as the ***AI Act***.[20] This regulation stems from a proposal presented by the European Commission in April 2021.[21] The proposal was the subject of intense negotiations (including a 36-hour marathon session between representatives of the European Commission, European Parliament and Council), far-reaching amendments and a *corrigendum* (of April 19, 2024), was approved on June 13, 2024, and published on July 12 under the number 2024/1689.[22]

9 This article aims to provide a critical overview of the main aspects of this Regulation. The critique is undertaken from a dogmatic perspective.[23] The goal is to present a general descriptive legal analysis of the Regulation in the wider context of EU law with a view towards ensuring logical consistency and a better understanding of the applicable rules (knowing what the law is). With that in mind, it will be possible to make an assessment of the possible impact of the AI Act and whether it achieves its self-proclaimed goal, i.e. to "*foster the development, use, and uptake of AI in the internal market that at the same time meets a high level of protection of public interests, such as health and safety and the protection of fundamental rights*" (recital 8).

10 After this introduction, the article analyzes the structure, objectives, and approach of the AI Act. It explains key concepts such as "artificial intelligence," "deployer," and "provider." The article then examines the scope of the AI Act, explores the core principles underpinning the Act, including fairness, transparency, and accountability, and the

[12] On the state of the subject at that time, see the text (in portuguese) Nuno Sousa e Silva, 'Direito e Robótica: Uma primeira *aproximação*' Revista da Ordem dos Advogados [2017] pp. 485-551.

[13] COM(2018)237 final of April 25, 2018.

[14] COM(2018)795 final, of December 7, 2018.

[15] With the subtitle "A European approach to excellence and trust" (COM(2020)65 final).

[16] 2020/2014(INL).

[17] Recital 9 of this proposal reads: "*Council Directive 85/374/EEC ("Product Liability Directive") has proven for more than 30 years to be an effective means of obtaining compensation for damage caused by a defective product. It should therefore also be used with regard to civil liability actions by a party suffering loss or damage against the producer of a defective AI system.*" In fact, this Directive does not apply well to *software* and, more generally, to digital content or goods with digital content (including artificial intelligence agents and autonomous robots). In addition, there are some restrictions and practical obstacles to obtaining compensation.

[18] COM(2022)495 final.

[19] COM(2022)496 final.

[20] This text is referred to as the "**Regulation**" or "**AIA**" and to which, unless otherwise indicated or contextualized, the rules quoted without further indication belong. The legislative basis used is twofold: arts. 16 (on data protection) and 114 (on the internal market), both of the TFEU.

[21] COM(2021)206 final. For a description of the background and main features of the evolution of the proposals up to 2023, see Nikos Th. Nikolinakos, *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies-The AI Act* (Springer 2023) and Carmen Muñoz García, *Regulación de la inteligencia artificial en Europa* (Tirant lo Blanch 2023).

[22] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 creating harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Regulation).

[23] As defined by Robert Alexy, *A Theory of Legal Argumentation* (OUP 1989) pp. 250-251 dogmatic refers to "a mixture of at least three activities: (1) that of describing the law in force, (2) that of subjecting it to conceptual and legal systematic analysis, and (3) that of working out proposals about the proper solutions to legal problems.". Along the same lines see the perspective of Roger Cotterrel, *The Politics of Jurisprudence* (Butterworths 1989).

specific practices it prohibits, such as manipulative uses of AI and social scoring. Furthermore, the article examines the classification of "high-risk" AI systems, such as those used in healthcare, the obligations for transparency in certain AI systems, and the regulations surrounding general-purpose AI models. Finally, it describes the mechanisms for certification, supervision, and enforcement of the AI Act, concluding with an evaluation of this important piece of EU legislation.

B. Structure, objectives, and approach

11 The Regulation is an example of the so-called "regulatory brutality" trend.[24] This piece of legislation is particularly complex, involving 68 definitions, 113 articles, 13 annexes and 180 recitals. The penalties are severe (up to 7% of the offender's global revenue or 35 million euros), the territorial scope of application is particularly broad, and supervision is carried out at national and EU level, establishing a new regulatory architecture, which includes the *EU AI Office*, the EU AI *Board*, an advisory forum and a scientific panel of independent experts (arts. 64 ff.) and, at national level, at least one national notifying authority and one national market surveillance authority (art. 70).

12 The AIA is made up of 13 chapters: 1) general provisions; 2) prohibited practices; 3) high-risk systems; 4) transparency obligations for certain types of systems; 5) general purpose models; 6) measures in support of innovation; 7) governance; 8) high-risk system database; 9) post-market monitoring, information sharing, and market surveillance; 10) codes of conduct and guidelines; 11) delegation of powers and Committee procedure; 12) sanctions and 13) final provisions.

13 The major division of the Regulation is based on a risk classification of AI systems. [25] This classification considers the uses or applications of AI systems. It is, therefore, a question of knowing what the system is designed for, the so-called "intended purpose," defined in art. 3/12 as "*the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation*". Thus, the same algorithms and software applied both in system A and system B can lead to a different risk classification.[26] The approach is not on the technology but rather on the goal of each system. Conversely, the provider can exclude the application of certain rules or even the Regulation as a whole if it is careful and explicit in the instructions and materials it makes available.[27]

14 There are two levels of risk: intolerable risk (which leads to the prohibition of certain practices or uses of AI systems - article 5)[28] and high-risk.[29] Most of the

[24] V. Papakonstantinou/Paul De Hert, 'The Regulation of Digital Technologies in the EU: The law-making phenomena of "act-ification", "GDPR mimesis" and "EU law brutality"' Technology and Regulation [2022] pp. 48-60.

[25] "Risk" is defined in art. 3/2 of the Regulation as " *'the combination of the probability of an occurrence of harm and the severity of that harm;*". On the risk-based regulatory approach *see* Giovanni De Gregorio / Pietro Dunn, 'The European risk-based approaches: Connecting constitutional dots in the digital age' Common Market Law Review vol. 59(2) (2022) pp. 473-500. Criticizing the notion of risk in the context of the regulation see Marco Almada / Nicolas Petit, 'The EU AI act: a medley of product safety and fundamental rights?' RSC Working Paper 2023/59 pp. 19-20.

[26] There will often be difficulty in determining what is the use in question - if the system has several possible applications and the Regulation applies to the entire value chain, could that system have different levels of risk along the chain? The answer must be yes. As noted, what matters for the classification is the intended use. When the system was designed for a given, low-risk use is actually being used for a high-risk application, art. 25 provides that this change of purpose can change the qualification of the person who made it, changing from "deployer" (the user) to "provider" (the person primarily responsible for ensuring compliance with the Regulation). In addition, the Regulation deals with general purpose models (art. 51 ff.), which can be used for many different purposes.

[27] art. 8. Even so, the Regulation obliges the producer of a high-risk system to have a risk management system, which includes (art. 9/2/b)) the estimation and assessment of the risks that may arise from "reasonably foreseeable misuse", defined as " *use of an AI system in a way that is not in accordance with its intended purpose, but which may result from reasonably foreseeable human behaviour or interaction with other systems, including other AI systems*" (art. 3/13).

[28] One might wonder if this approach makes sense. If the same application or practice took place without the use of AI systems, would it be legal? If the answer is no, then the association with AI systems is irrelevant. In fact, I submit that art. 5 is about regulating conducts and would not need to be AI-specific

[29] art. 50 does not refer to "low risk" or "limited risk", it applies in light of the use in question, regardless of the risk classification of the system. It is often pointed out that there are AI systems, such as video games and *spam* filters, which are not covered by the Regulation and would constitute another category of "no risk". think it would be better to just point out that these systems are not covered by the Regulation. Nevertheless, recital 27 hints at voluntary compliance. Marco Almada / Nicolas Petit, (n 25), pp. 8-9 mention three tiers: intolerable risk (art. 5), high-risk (covered by the Regulation) and other AI systems (which are not covered by the Regulation, but are subject i.a. to

rules are aimed at high-risk AI systems. As we shall see, art. 5 presents difficulties of interpretation and delimitation. It is therefore essential to look at art. 6, which defines high-risk systems, to understand the scope of the prohibited practices. If the Regulation considers a certain uses of the AI system to be high-risk, then it cannot be included in the prohibited practices. In other words, article 6 is particularly important to define the scope of article 5.

15 The Regulation also regulates so-called general purpose AI models, i.e. "*an AI model (...) that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market*" (art. 3/63), in particular those that present a systemic risk (arts. 51 ff.).

16 The approach taken in the Regulation is in line with legislation on product safety,[30] namely Regulation (EU) 2023/988 of 10 May 2023 on general product safety,[31] and sectoral regulatory instruments on toys,[32] cosmetics,[33] and medical devices.[34] AI systems are regarded as goods, and high-risk systems must bear a conformity mark (*CE* - short for *conformité européenne*) which confirms that there has been a verification and that the (high-risk) AI system complies with the applicable EU legislation (art. 48).[35] The simplest way to avoid ambiguities and interpretative difficulties will be to follow the standards and technical norms approved under the Standards Regulation,[36] thereby benefiting from a presumption of conformity (arts. 40/1 and 42/2).[37]

17 Nevertheless, the Regulation considers the complexity (and sophistication) of Artificial Intelligence. To use Laura Caroli's words, "[an AI system] is not a toaster". This is why the AI Act presents considerable deviations from classic product safety laws, namely by imposing duties on users of the systems (art. 26) and, in some cases, requiring an impact assessment on fundamental rights (art. 27). The Regulation is therefore a hybrid, combining an approach typical of rights-legislation such as the GDPR with another, typical of regulatory law. However, with the exception of the right to lodge a complaint (art. 85) and an explanation of the role of the AI system in certain decisions (art. 86), this Regulation does not establish subjective rights.

18 Although this is an "Artificial Intelligence" regulation, it seems to me that many of these practices and actions, especially the prohibited ones, would already be covered by the existing regulatory framework, namely the Digital Services Regulation,[38] the General Data Protection Regulation,[39] the rules

Regulation 2023/988 on general product safety).

[30] Michael Veale / Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach' Computer Law Review International (2021) p. 98. In this sense, the AI Act makes copious references to Regulation (EU) 2019/1020 on market surveillance and product conformity.

[31] In 2008 the EU adopted the so-called "New Legislative Framework", an updated legislative package of general rules for ensuring product safety and conformity, accompanied by special rules for certain categories (to date 26 categories, including elevators, construction material, explosives, radio, fertilizers, batteries, machinery and *drones*). The Regulation is now part of this category of legislation.

[32] Directive 2009/48/EC of the European Parliament and of the Council of June 18, 2009 on the safety of toys

[33] Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products.

[34] Regulation (EU) 2017/745 of the European Parliament and of the Council of April 5, 2017 on medical devices. As stated in recital 19 of this regulation, "*It is necessary to clarify that software in its own right, when specifically intended by the manufacturer to be used for one or more of the medical purposes set out in the definition of a medical device, qualifies as a medical device, while software for general purposes, even when used in a healthcare setting, or software intended for life-style and well-being purposes is not a medical device. The qualification of software, either as a device or an accessory, is independent of the software's location or the type of interconnection between the software and a device.*" Peter Feldschreiber (ed.), *The Law and Regulation of Medicines and Medical Devices* (OUP 2021).

[35] The CE marking rules are contained in Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products. The rules on standards are contained in Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardization.

[36] Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardization, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision 1673/2006/EC of the European Parliament and of the Council.

[37] Michael Veale / Frederik Zuiderveen Borgesius, (n 30), p. 105 point out that this will probably be the path followed by most producers.

[38] Regulation (EU) 2022/2065 of the European Parliament and of the Council of October 19, 2022 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Regulation).

[39] Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and

of fair competition and consumer protection, including Advertising Law and, more generally, the rules protecting personality rights and Fundamental Rights.[40] We must not forget that the regulation of Artificial Intelligence does not begin or end with this Regulation, despite its undeniable importance.[41]

C. Concepts

19 The Regulation has taken a maximalist approach to definitions, defining terms that are already part of the European acquis such as "personal data", "non-personal data", "profiling", "biometric data", enshrining unhelpful definitions such as "AI literacy" and terms that are self-explanatory such as "publicly accessible space", "training data" or "instructions for use".[42]

20 On the other hand, the concept "law enforcement" is important but not obvious. This term, which appears 98 times in the Regulation, is defined in art. 3/46 as "*activities carried out by law enforcement authorities or on their behalf for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including safeguarding against and preventing threats to public security*", with "*law enforcement authority*" being "*any public authority competent for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security; or any other body or entity entrusted by Member State law to exercise public authority and public powers for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security*" (art. 3/45). In other words, when the Regulation refers to law enforcement, it is essentially referring to police activity.

21 For a proper understanding of the regulation, it is necessary to understand the definition of an Artificial Intelligence system and analyze the various categories of subjects.

I. Artificial Intelligence System

22 The first challenge for regulation was to find a suitable definition of Artificial Intelligence. Many definitions associate intelligence with human intelligence, the ability to use reasoning to achieve goals. Other perspectives approach the concept through the programming techniques used.[43] After much discussion, the Regulation ended up adopting the definition of "Artificial Intelligence systems", which replicates the updated OECD definition: "*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*" (art. 3/1).[44]

on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[40] Cfr. Stefan Scheurer, 'Artificial Intelligence and Unfair Competition - Unveiling an Underestimated Building Block of the AI Regulation Landscape' GRUR Int vol. 70(9) (2021) pp. 834-845.

[41] Even in terms of product safety in the internal market, recital 166 points out that "*it is important that AI systems related to products that are not high-risk in accordance with this Regulation and thus are not required to comply with the requirements set out for high-risk AI systems are nevertheless safe when placed on the market or put into service. To contribute to this objective, Regulation (EU) 2023/988 of the European Parliament and of the Council (53) would apply as a safety net*" .

[42] See, respectively, art. 3(50), (51), (52), (34), (56), (44), (29), and (15). On the other hand, "widespread infringement" (art. 3/61) and "deep fakes" (arts. 3/60) are defined, but these terms are used only once in the Regulation (respectively arts. 73/3 and 50/4).

[43] This was the much-criticized original approach of the European Commission.

[44] The concept also corresponds to that used in art. 2 of the CoE Convention. On updating the OECD definition, see the *Explanatory Memorandum on The Updated OECD Definition of an AI System* (March 2024), which prefers to use the notion of AI *systems* for regulatory purposes. This perspective is in line with the US Executive Order on Artificial Intelligence (*Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)*, but has some notable differences. The Presidential Order, which generalizes the approach of Executive Order 13960 (that was directed only at federal agencies), is based mainly on cybersecurity requirements, monitoring and technical quality of systems and defines Artificial Intelligence. In section 3 b) of EO 14110 as "a *machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action*". The US Executive Order refers to the human definition of objectives, which is not required in the OECD and AI Act definitions. Luciano Floridi, 'On the Brussels-Washington Consensus About the Legal Definition of Artificial Intelligence'. Philosophy & Technology (2023) vol. 36 (87) The definition of AI used in the ISO/IEC 22989:2022 (2022) standard is similar: "*a technical and scientific field devoted to the engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of defined objectives*".

23 This notion seems particularly broad and almost coincides with the concept of software. The distinction lies in the existence of some degree of **autonomy** and the mention of **inferences**. In this sense, recital 12 explains that "*the definition should be based on key characteristics of AI systems that distinguish it from simpler traditional software systems or programming approaches and should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations*". To this end, it stresses that what is essential is the ability to make *inferences, i.e. the* possibility of processing or generating new data in contexts other than those in which the system was trained.[45] In other words, simple automations, formulas, static software or totally deterministic programming (*if x, then y*) are excluded.[46] As the notion is broad, in case of doubt the system analyzed should be considered an AI system.

24 It is important to stress that the regulation essentially concerns systems as a whole (including *hardware,* i.e. computers, sensors, peripherals and other software that does not constitute artificial intelligence). **Systems** must be distinguished from **models**. As pointed out in recital 97: "*Although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI systems. AI models are typically integrated into and form part of AI systems.*" While ChatGPT (from OpenAI) constitutes an AI system (including several layers of software, a graphical interface, servers, etc.), there are several models (which act as the system's "engine") that can integrate it (to date, and in the case of ChatGPT, three options are available: GPT-3.5, GPT-4, and GPT-4o). It is possible to use the same model to build systems with very different applications, purposes, and modes of operation.[47]

II. Subjects

25 The Regulation mentions several roles that form part of the AI value chain: the provider, the importer, the distributor, the authorised representative, and the deployer, all of whom are covered by the generic notion of "**operator**". As we shall see, the Regulation applies to any provision of the system in the EU, even if it is free of charge.[48]

26 The main target of the AIA is the *provider*, defined in art. 3/3 as "*a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge;*". The central feature that defines someone as a provider is the fact that they offer an AI system under their own name.[49] Providers, when not

[45] Recital 12: "*... The capacity of an AI system to infer transcends basic data processing by enabling learning, reasoning or modelling.*". As highlighted in the *Explanatory Memorandum on The Updated OECD definition*, there are also inferences in the training phase, especially in the case of *unsupervised machine learning*.

[46] *Robotic Process Automation* (i.e. a way of automating repetitive processes, usually in a business context) will have to be analyzed on a case-by-case basis. In some cases, AI agents may be involved; in others, it is mere deterministic programming. In any case, it seems that most cases of RPA will not fall within the material scope of the Regulation as they are unlikely to present a relevant risk.

[47] It is mostlyto this extent that the AIA is also concerned with models. There is, however, a definition of "general-purpose AI system" in art. 3/66: "*an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems*". This concept is only used by the Regulation to refer to the modification of such a system to serve a (specific) purpose classified as high-risk (art. 25/1/c)).

[48] The Regulation refers to the territory of the Union and does not cover other countries in the European Economic Area.

[49] This will also include so-called OEMs (*Original Equipment Manufacturers*), who may not have had any role in the development of the system, but who integrate it into their product and/or present the AI system as their own. The qualification cannot be circumvented, however, by arguing that the "producer" of the system merely provides technical means. Of course, there will be dubious situations: when company A provides *middleware* to allow its customers to develop AI models, applications, or even systems and/or allows these models and applications to run on its infrastructure (servers), who is the provider? I think we can consider company A's *middleware* system as an AI system and company A as the provider of that system. However, the systems developed by each of company A's clients and eventually made available to third parties will constitute separate systems of which company A's clients will be the providers. The situation can get complicated if company A provides a configurable AI system. In that case, considering art. 25, whether those customers remain deployers or become providers of a new system will depend on the extent of the modifications made and/or the branding of that customer on the system. If these changes are significative, company A (provider of the original system) must "*closely cooperate with new providers and shall make available the necessary information and provide the reasonably expected technical access and other assistance that are required for the fulfilment of the obligations set out in this Regulation*" (art. 25/2). There is also provision for "mandatory contracting". Pursuant to art. 25/4 "*The provider of a high-risk AI system and the third party that supplies an AI system, tools, services, components, or processes that are used or integrated in a high-risk AI system shall,* ***by written agreement****, specify the necessary information, capabilities, technical access and other assistance based on the generally acknowledged state of the art, in order to enable the provider of the high-risk AI system to fully comply with the obligations set out in this Regulation*". Although recital 88

established in the EU, must fulfill their obligations through **authorised representatives** established in the EU (defined in art. 3/5), as provided for in arts. 22 (in the case of high-risk AI systems) and 54 (general purpose AI models).[50]

27 The user, except for those who use the system as part of a personal, non-professional activity,[51] is the "*deployer*" (art. 3/4) and also has obligations of their own, namely, to supervise the operation of the system (cf. arts. 26 and 50/3 and /4).

28 Importers, i.e. those people located in the EU who place an AI system on the internal market (art. 3/6), will have certain obligations to verify and guarantee conformity, as well as to collaborate with the authorities (art. 23). The Regulation reserves the term "placing on the market" for the initial act making available of an AI system on EU territory (art. 3/9), with "making available" being defined as any supply in the context of a commercial activity (art. 3/10). Thus, importers carry out "placing on the market", while distributors (art. 3/7) are engaged in "making available on the market" following importation.[52] Distributors are subject to obligations of verification and cooperation with the authorities that are very similar to those placed upon importers (art. 24).

29 Another concept, which is not defined but is included in the concept of operator, is that of "product manufacturer" (referred to in art. 2/1/e)). Given that what is at stake is the joint provision of a product and an AI system under one's own name or brand, product manufacturers should be considered providers.[53]

30 A person can become a provider if they "*put their name or trademark on a high-risk AI-system already placed on the market*" (art. 25/1/a)), "*make a substantial modification to a high-risk AI-system that has already been placed on the market or put into service, in such a way that it remains a high-risk AI-system*" (art. 25/1/b)) or "*modify the intended purpose* (...) so that the *AI-system concerned becomes a high-risk AI-system*" (art. 25/1/c)).[54] Although the reverse is not expressly spelled out, changing the intended use of the AI system to one that is not considered high-risk will allow the modified system to escape the application of certain rules or even the Regulation as a whole.

D. Scope of application

31 Despite being a general regulation, the AI Act explicitly safeguards the application of the rest of the legal and regulatory framework (art. 2, paragraphs 5, 7 and 9)[55] and allows complementary national rules to be adopted in certain areas, such as more favorable standards for the protection of workers (art. 2/11) or rules on the use of remote biometric identification systems (art. 5/5 and /10). In addition, the application of some legislation (art. 2/2, referring to the list in Section B of Annex I) and sectoral supervision (arts. 72 and 74) is reserved. [56] There are also matters that depend on implementing measures at national level, in particular the designation of national authorities and the supervisory framework (arts. 70 and 74), as well as the sanctions regime (art. 99/2). On the other hand, the Commission has broad power to adopt delegated acts, complete and update the Regulation (arts. 7 and 97),and perform extensive evaluations and reviews (art. 112). The Commission will also draw up comprehensive guidelines on the Regulation (art. 96) and encourage the development of codes of practice (art. 56).

32 The AI Regulation is in line with the latest trend in digital single market regulation, having **extraterritorial** application.[57] According to art.

may give a different impression, I don't believe that art. 25/4 applies to those who merely provide models and I believe that the "mandatory" contracting provided for in this article should be interpreted restrictively (otherwise, even the supplier of cooling systems for the computers used to train an AI system or the provider of meals to data scientists could be covered).

[50] This obligation is similar to that laid down in art. 27 GDPR and typical of product safety legislation.

[51] I anticipate that this exception will be interpreted restrictively. Thus, my use of an AI system to generate images for a conference presentation as a teacher or lawyer would not be covered.

[52] This distinction will be more frequent when AI systems integrate hardware than with standalone software. In any case, there are often software distribution agreements, including resale agreements.

[53] In this sense, see art. 25/3.

[54] The notion of substantial modification is defined in art. 3/23 as "*a change to an AI system after its placing on the market or putting into service which is not foreseen or planned in the initial conformity assessment carried out by the provider and as a result of which the compliance of the AI system with the requirements set out in Chapter III, Section 2 is affected or results in a modification to the intended purpose for which the AI system has been assessed*". This definition is close to the idea of purpose change set out in art. 6/4 of the GDPR. A system subject to a substantial modification is treated in the Regulation as a new system (cf. art. 43/4).

[55] In addition to these provisions, there are rules, such as art. 87, which expressly refers to other European legislation.

[56] It should also be noted that the Regulation, in articles 102 to 110, amends various instruments of EU law.

[57] Christopher Kuner, 'Protecting EU Data Outside EU Borders under the GDPR' Common Market Law Review 60 (2023) pp. 77-106. This approach by the European Union has contributed to the so-called "Brussels Effect", a term coined

2/1, a minimum point of contact of the user or the result of the AI system with the territory of the Union is sufficient to trigger the applicability of the Regulation. Thus, if the result of an AI system is used in the EU or affects people located in the EU, this is enough for the Regulation to apply. On the other hand, the Regulation does not apply to anyone who develops AI systems in the EU, even for purposes prohibited by the Regulation, for use in third countries (i.e. there is no export control). Along the same lines, there is an obligation for providers established in third countries to appoint an authorised representative (arts. 22 and 54).

33 An important note in terms of **jurisdiction** concerns the decentralized nature of supervision. Except in the case of general-purpose AI models, which will be supervised by the European Commission, the competent national authorities will be responsible for dealing with all infringements that take place within their territory. Thus, the same provider and infringement may be subject to the concurrent jurisdiction of several national authorities.

34 Pursuant to article 2/6, research and development activities "in the laboratory" are excluded from the material scope of the Regulation (articles 57 ff. establish a complex set of rules for testing in a real world environment). Activities prior to the system being placed on the market or put into service are also not covered by the Regulation (art. 2/8).

35 The Regulation will also not apply to systems developed or used exclusively for military, national security or defence purposes (art. 2/3) or to use by public authorities of third countries and international organizations provided that these entities adequately safeguard fundamental rights (art. 2/4).

36 The topic of ***open source*** was the subject of much debate.[58] "Domestic" uses, i.e. "*in the context of a personal activity of a non-professional nature*", are excluded (art. 2/10). However, making software (including the parameters of a model) available under open-source licenses can also be done in a professional context.[59] The compromise solution is a limited exemption (art. 2/12).[60] The key to understanding this provision is the aforementioned difference between models and systems. The provision of open-source AI *models* enjoys certain exemptions under the Regulation. *Models* made available under open-source licenses are only required to comply with two obligations (copyright compliance policy and transparency regarding training data)[61] except in the case of general-purpose models with systemic risk (articles 25/4, 53/2 and 54/6). On the other hand, for AI *systems* covered by the Regulation (regardless of the level of risk), the fact that they are made available in *open source* is irrelevant. Simply put, the partial exemption is for models, not systems.

37 The application of these Regulations over time will be phased in. The Regulation entered into force on August 1, 2024, with the amendments to the legislation mentioned in articles 102 to 110 taking effect on that date. The general application of the Regulation is scheduled for August 2, 2026 (art. 113). There are, however, parts of the Regulation that will apply sooner. This is the case for the first two chapters (on prohibited practices), which will apply from February 2, 2025 (art. 113/a)), and the rules on the institutional framework, which will apply from

and described by Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (OUP 2020). This expression alludes to the influential power of the European Union's regulatory acquis in matters such as competition law, environmental law, digital law and data protection. In these areas, the EU has been a pioneer in regulation and is often followed as a model in other jurisdictions. In addition, multinational companies end up adopting European rules as a global compliance standard. However, in the specific case of the AIA, it is far from clear whether the approach taken at EU level will have this effect (cfr. Ugo Pagallo, *Why the AI Act Won't Trigger a Brussels Effect* (2023) in https://ssrn.com/abstract=4696148).

[58] On the notion and history of *open source* see Amanda Brock (ed), *Open Source Law, Policy and Practice* (OUP 2022). In the case of AI, the debate around open source occurs at various levels. Some advocate the need to restrict the circulation of information (and are proponents of what is known as *security through obscurity*), going so far as to compare the availability of code for certain systems to the availability of instructions for producing an atomic bomb. Others argue that openness is the most effective way of guaranteeing diversity, advancement, and even security. There is also considerable disagreement as to what is meant by *open source* in AI: whether it is enough to make the architecture and parameters of a model available (e.g. *open weights*) or whether the *dataset* used to develop it must also be made available. On the conceptual discussion in this area see Andreas Liesenfeld/Mark Dingemanse, 'Rethinking open source generative AI: open-washing and the EU AI Act' FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (June 2024) pp.1774-1787. In recitals 102 and 103, the Regulation seems to adopt a rather narrow notion of open source.

[59] In fact, in some contexts, only companies with a lot of resources will be able to develop certain models (e.g., the several LLama developed by Meta).

[60] In fact, the text of art. 2/12 is completely useless: the exclusion provided for does not apply to the three types of systems covered by the Regulation.

[61] As explained in recital 104, the fact that a model is *open source* does not mean that one will have access to the training data or that respect for intellectual property rights has been guaranteed.

August 2, 2025 (art. 113/b)).

38 On the other hand, the rules on high-risk systems that are safety components of harmonized products (art. 6/1) will have a *vacatio legis of* 36 months and will only apply from August 2, 2027 (art. 113/c)). More importantly, the rules concerning high-risk AI systems will only apply to AI systems placed on the market after that date. AI systems already placed on the market, when they are considered high-risk, are exempt from the rules of the Regulation unless they undergo significant changes (art. 111/2).[62] General purpose AI models placed on the market before August 2, 2025, will only be required to comply with the Regulation from August 2, 2027 (art. 111/3).[63]

E. Principles

39 Although not in the initial proposal, which was essentially aimed at determining prohibited practices and regulating high-risk applications, there was consideration of enshrining a set of general principles applicable to all operators and all AI systems subject to the Regulation.[64] In the final version, the only duty with such breadth is the obligation imposed on providers and implementers to ensure that people operating or using AI systems "*have a sufficient level of AI **literacy***" (art. 4).[65]

40 Nevertheless, those principles still underlie the requirements placed on high-risk systems (arts. 8 to 15) and their operators (arts. 16 to 27).

41 At issue is a set of concerns developed in the interdisciplinary field known as AI safety or *FATE (Fairness*, *Accountability*, *Transparency*, *Ethics) AI*, including concerns of control, transparency, alignment, non-discrimination, robustness, and security.

42 Some principles are hard to parse. Of course, we are all in favor of *fairness.* The great difficulty, which is the field of philosophy and then politics, translating into the committed choice of each society at a certain time and place through positive law, lies in defining what is just, equitable, and fair. This problem is both conceptual and technical-mathematical.[66] In practical terms, not much can be drawn from this principle.

43 There are similar difficulties with algorithmic bias. Some of the known problems result from the poor quality of the data used (namely lack of representativeness or quantitative or qualitative insufficiency) or programming errors.[67] On the other

[62] In that sense, it will no longer be the same system. It is unclear how the concept of "significant changes in their design" differs from "substantial modification" used in arts. 25 and 43/4. Recital 128 indicates that the concepts do not coincide. In any case, this rule, which gives a significant advantage to incumbent operators, is explained by the prohibition of retroactivity (what triggers the application of most of the Regulation's rules is the placing on the market). On the other hand, the prohibitions in art. 5, which refer to prohibited practices (and not system requirements) can and will be fully applicable to systems that are already on the market. In the case of certain "large-scale IT systems" of the European Union already in use, such as the Schengen IT system or the visa and travel information system (the list is in Annex X), which are already in operation, it is stipulated that they must be brought into conformity with the Regulation by December 31, 2030 (art. 111/1).

[63] On the other hand, models placed on the market after August 2, 2025 will have to comply with the rules "immediately" (art. 113/1/b)).

[64] In particular in art. 4a presented in May 2023 (COM(2021)0206 - C9 0146/2021 - 2021/0106(COD)), which set out the following principles: "a) human oversight and control; b) technical robustness and security; c) privacy and data governance; d) transparency; e) diversity, non-discrimination and fairness; f) social and environmental well-being". arts. 7 to 13 of the CoE Convention also set out the following principles: human dignity and autonomy, transparency and control, accountability and responsibility, equality and non-discrimination, protection of privacy and personal data, reliability and safe innovation. Many of these principles coincide with those listed in art. 5 of the GDPR, which will remain fully applicable whenever AI systems process personal data. Recital 27 of the AI Act mentions the "seven non-binding ethical principles" and encourages voluntary compliance with them.

[65] art. 20 of the CoE Convention also establishes a principle of promoting digital literacy. It should be noted that there is no sanction for the violation of the duty of promoting AI literacy.

[66] Sorelle A. Friedler / Carlos Scheidegger / Suresh Venkatasubramanian, 'The (Im)possibility of fairness: different value systems require different mechanisms for fair decision making' Communications of the ACM. 64 (4) (2021) pp. 136-143.

[67] Examples abound, such as Google Photos' facial recognition system classifying black individuals as gorillas (in 2015), Amazon's recruitment tool prejudicing women (2018) and, more recently, in 2023, the iTutorGroup tool, used in recruitment, automatically rejecting applications from women over 55 and men over 60. The problem of algorithmic discrimination is widespread and reaches a large scale, as demonstrated by Z. Obermeyer et al., 'Dissecting racial bias in an algorithm used to manage the health of populations' Science, (2019) 366(6464) pp. 447-453 on the health system in the USA. Hilde Weerts et al, 'Algorithmic unfairness through the lens of EU non-discrimination law: Or why the law is not a decision tree'. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (2023) pp. 805-816 and Philipp Hacker, 'Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies Against

hand, many problematic situations simply result from the system having been optimized to achieve a given beneficial or innocuous objective. For example, if an algorithm is designed to favor what an internet user pays more attention to, it could end up recommending alcoholic drinks (having indirectly detected that (s)he is an alcoholic) or promoting offensive or aggressive speech (since this is what most people will pay more attention to). These challenges, especially those posed by recommender systems, are already partially addressed in the Digital Services Act ("DSA").[68] In any case, the AI Act places significant emphasis on diversity and the prevention of discrimination and bias.[69] Putting an end to these occurrences is impossible, but there is an obligation to make adequate efforts to follow the best practices to prevent easily avoidable mistakes.

44 **Transparency** can be understood as referring to several different concepts.[70] One of them, employed in art. 50, refers only to the origin of a given content or agent as being or coming from AI systems. Transparency is also covered by the obligation to provide and maintain technical documentation (arts. 11, 18, 20 and Annex IV), record-keeping (arts. 12 and 19), the provision of information (art. 13), and cooperation with authorities (art. 21).

45 When transparency refers to the characteristics of the AI system, this concept can allude to the description of the human tasks of designing, configuring and making the system available, even if the system is itself (i.e. in its operation) opaque. Transparency is sometimes used to refer to **interpretability**, i.e. the ability to understand how an AI system works,[71] and/or *explainability*, i.e. the clarification of why a certain result was obtained by operating the system.[72] A system can be interpretable, but produce concrete results that are not explainable.[73] For example, we know the parameters used and the steps followed by the system to assign an insurance premium, but we can't explain why individual A has a higher premium than individual B. There are, however, artificial intelligence techniques that generate totally opaque systems (e.g. large language models, such as GPT);

Algorithmic Discrimination Under Eu Law' Common Market Law Review 55 (2018) pp. 1143-1186.

[68] The DSA defines a "recommender system" as "a *fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service or prioritise that information, including as a result of a search initiated by the recipient of the service or otherwise determining the relative order or prominence of information displayed"* (art. 3/s)) and imposes, only on online platform providers, obligations of transparency of such systems (art. 27). In the case of providers of online platforms or very large online search engines, there are also duties to assess systemic risk, including assessing the "*design of their recommendation systems and any other relevant algorithmic system*" (art. 34/2/a)) and adopting measures to mitigate the risks identified in these systems (art. 35/1/d)). Under art. 38 of the DSA, very large online platforms and very large online search engines must allow users to configure recommendation systems so that they do not carry out profiling (a concept defined in art. 4/4 of the GDPR). Providers of these systems are also required to explain to regulators "*the design, logic, operation and testing of their algorithmic systems, including their recommendation systems*" (art. 40/3 DSA). On the subject of recommender systems, see Sergio Genovesi / Katharina Kaesling / Scott Robbins (eds), *Recommender Systems: Legal and Ethical Issues* (Springer 2023) and Mireille Hildebrandt, 'The issue of proxies and choice architectures. Why EU law matters for recommender systems.' Frontiers in Artificial Intelligence 5 (2022): 789076.

[69] In particular in art. 10 on data governance and art. 15/4 on cybersecurity. The technical documentation required of providers of general purpose models also includes "*a detailed description of the elements of the model (...) and the relevant information on the development process, including (...) information on the data used for training, testing and validation, if applicable, including the type and provenance of the data and the curation methodologies (e.g. cleaning, filtering, etc.), the number of data points, their scope, etc, cleaning, filtering, etc.), the number of data points, their scope and main characteristics; how the data were obtained and selected, as well as all other measures to detect the inadequacy of data sources and methods to detect identifiable biases*, if applicable" (Annex XI, Section 1 (2)). On the other hand, the Regulation confers supervisory powers over high-risk AI systems on national public authorities or bodies that supervise or ensure compliance with obligations under Union law protecting fundamental rights, including the right to non-discrimination (art. 77).

[70] The GDPR also uses the concept of transparency in art. 5/1 and recital 58, referring to the clear communication of information. Noting the "marked polysemy" of the concept of transparency, see Lorenzo Cotino Hueso, 'Transparencia y explicabilidad de la inteligencia artificial y "compañía" (comunicación, interpretabilidad, integilibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad...). Para qué, para quién y cuánta.' in Lorenzo Cotino Hueso / Jorge Castellanos Claramunt (eds), *Transparencia y explicabilidad de la inteligencia artificial* (Tirant lo Blanch 2022) pp. 25 ff. In 2017 Zachary C. Lipton, *The Mythos of Model Interpretability*, arXiv:1606.03490 (2017) even stated: "*the term interpretability holds no agreed upon* meaning". The aforementioned 2020 technical standard used in this text seems to contribute to greater terminological certainty.

[71] This is the definition in the technical standard ISO/IEC TR 29119-11:2020(en), 3.1.42.

[72] See the definition used in the technical standard ISO/IEC TR 29119-11:2020(en), 3.1.31.

[73] art. 14/4/c) states that the system must allow a human being to "correctly interpret the results of the high-risk AI system, taking into account, for example, the available interpretation tools and methods". This wording seems to admit the use of so-called *black-box AI*, but in such cases there are no interpretation tools or methods available.

we know very little about their inner workings.[74] For these, interpretability and explainability are not technically possible. [75]

46 The AI Act does not impose a general obligation to generate explainable models or decisions. However, in the case of high-risk systems, it establishes a right to an explanation of the role of the system (arts. 13 and 86), and to understand the main principles of its operation and the decision taken (arts. 14 and 86). The text of art. 86 (and recital 171) is not entirely clear as to whether it is necessary to explain the specific decision or whether a general explanation is sufficient.[76] On the other hand, the references to the relevant technical capacities to explain the results (art. 13/3/b)/iv)) and "*where appropriate, information enabling those responsible for the deployment to interpret the results of the high-risk AI system and to use them appropriately*" (art. 13/3/b)/vii)) are made in the context of technical documentation, which seems to indicate that a generic and abstract explanation (*interpretability*) is at stake and not a real *explainability*. Furthermore, even if a right to an explanation of the specific decision were established, the protection of personal data, business secrets and other types of secrecy would act as a limit to the exercise of this right.[77] In this sense, in my opinion, AI techniques that do not allow explanations to be generated (e.g. deep learning neural networks or support vector machines) remain legally admissible, even in the case of high-risk systems.

47 **Supervision and human control** are reflected in the obligation for the provider to adopt a risk management system (art. 9), quality control (art. 17), to monitor its post-marketing operation (art. 72), to report serious incidents (art. 73) and to design high-risk systems in a way that allows for understanding and intervention in their operation (art. 14), namely the existence of a kill switch (art. 14/4/e)). These aspects intersect with cybersecurity and robustness concerns (art. 15) - to which an important legislative framework is associated, namely the NIS 2 Directive (Dir. 2022/2555 of December 14, 2022, on measures for a high common level of cybersecurity across the Union) - and with the GDPR rule restricting the possibility of automated decisions to certain cases (art. 22 GDPR).[78]

48 The implementation of these principles and of the Regulation will be densified to a large extent through *standards* and Commission guidelines, which will help to increase legal certainty.

F. Prohibited practices

49 At an early stage, the Commission proposed the establishment of four prohibited practices, said to pose an unacceptable risk, which could be summarily described as subliminal manipulation systems, systems that exploit vulnerabilities causing behavioral distortion and damage, social scoring systems and real-time biometric identification systems (e.g. facial recognition). These prohibitions had some exceptions and used particularly vague language.[79] After intense discussions and negotiations, the language has been refined, the list of prohibited practices has been extended, but the result is not much better. They now include:

- Manipulation and exploitation of vulnerabilities - art. 5/1/a) and b)
- General social scoring - art. 5/1/c)
- Predictive policing - art. 5/1/d)
- Creation of facial recognition databases - art. 5/1/e)
- Emotion recognition systems in the workplace or education - art. 5/1/f)

[74] This is an area of scientific research. Recently, a large group of Anthropic researchers published a paper "Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet" (https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html) in which the topic is discussed in detail and advances in the possibility of interpreting language models and using this technique for security purposes are demonstrated.

[75] Although recital 71 and, to a certain extent, art. 15/1/h) of the GDPR may give the impression that there would be a right to an explanation of automated decisions, this does not seem to be the most correct interpretation. See *supra* note 4 and also L. Edwards. / M. Veale, 'Enslaving the algorithm: From a "right to an explanation" to a "right to better decisions"?' IEEE Security & Privacy, 16(3) (2018), pp.46-54.

[76] The different language versions (in English "meaningful explanation", in Portuguese "explicação clara e pertinente", in Spanish "claras y significativas", in French "claires et pertinentes", in Italian "chiare e significative" and in German "klare und aussagekräftige") are not conclusive.

[77] In a similar vein, see art. 25/5. There are also duties of secrecy and confidentiality (art. 78). On the wider problem see Gianclaudio Malgieri, 'Trade Secrets v Personal Data: A Possible Solution for Balancing Rights' International Data Privacy Law, vol. 6(2) (2016) pp. 102-116.

[78] On this rule and the associated problems, see Federico Marengo, *Privacy and AI: Protecting Individual's Rights in the Age of AI* (2023).

[79] Michael Veale / Frederik Zuiderveen Borgesius, (n 30), pp. 98-99: "*In briefings on the prohibitions, the Commission has presented an example for each. They border on the fantastical (...) A cynic might feel the Commission is more interested in prohibitions' rhetorical value than practical effect*".

- Biometric classification of protected categories - art. 5/1/g)
- Special cases of real-time biometric identification - art. 5/1/h)

50 This list is not exhaustive. Other practices may be prohibited or unlawful on other grounds (art. 5/8). For example, systems that generate *deep fakes* are not normally seen as high-risk but are only subject to transparency obligations (art. 50/4). However, when such a system is configured or prepared to generate child pornography that will be a crime[80]

I. Manipulation and exploitation of vulnerabilities

51 A prerequisite for freedom in general, especially freedom of thought, choice, and expression, is an adequate perception/representation of reality. Private autonomy requires this. For this reason, national legal systems make legal transactions concluded on the basis of defects of will voidable and prohibit and punish unfair commercial practices and misleading advertising. The free will of each person, as a reflection of their dignity, is also reflected in the prohibition of experimentation on people and the requirement of free and informed consent, especially in the case of voluntary limitation of personality rights.

52 Some AI systems have the potential to manipulate and mislead, interfering with the free formation of thoughts, opinions and, thus, affect choices.[81] In this sense, the text of art. 5/1/a) of the Regulation prohibits "*the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harms.*" This wording uses indeterminate concepts and qualified language ("materially", "appreciably", "significant", "reasonably likely").[82] These qualifiers seem to indicate that not every advertising technique or hidden or misleading practice will be covered.[83] In fact, I believe that the criteria of advertising law and consumer protection will be less demanding, i.e., certain conduct qualified as aggressive or misleading advertising and/or unfair commercial practices will not fall under art. 5/1/a) of the Regulation. In such cases, the AI system will not be prohibited, but the activities in question, regardless of the use of an IT system, will be covered by the existing rules.

53 In turn, art. 5/1/b) prohibits "*the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm;*". Such behavior, in the context of legal transactions, is already prohibited by contract and consumer law. Here, too, it seems that the qualifiers used and the limitation to certain characteristics make the Regulation's standard more demanding than the legislation already in force, and, to that extent, the Regulation will have little impact.[84]

54 When thinking about personalized pricing that takes into account that a potential customer is in a situation that makes them willing to pay a higher price (e.g. their cell phone is low on battery or their biometric data indicates dehydration or fatigue),[85]

[80] Curiously, the same might not necessarily be true of so-called "face swap porn" of adults. E.g. in Portugal, this practice has no clear criminal framework to date. For minors, art. 176 of the Portuguese Criminal Code is sufficient if there is a "realistic representation of a minor", regardless of whether a forgery is involved. In the case of an adult, it is difficult to say that there is an offense against privacy (since there was no actual capture of real images). However, art. 5/1/b) of Directive 2024/1385 on combating violence against women and domestic violence seems to call for the criminalization of this practice.

[81] Art. 5/2 of the CoE Convention refers to the freedom to form opinions.

[82] There is controversy over the scientific basis of subliminal influence (i.e. that which falls below the threshold of conscious perception). Rostam J. Neuwirth, 'Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA)' Computer Law & Security Review, 48 (2023), proposes the use of the term transliminal (instead of subliminal), since manipulation usually takes place between the plane of consciousness and unconsciousness.

[83] This overlaps with the topic of *dark patterns* (forms of user interface that promote an action or choice that users would be unlikely to make or take otherwise). On the subject *see* Harry Brignull, *Deceptive patterns - exposing the tricks tech companies use to control you* (Testimonium Ltd 2023) and Inge Graef, 'The EU Regulatory Patchwork for Dark Patterns: An Illustration of an Inframarginal Revolution in European Law?' (2023) https://ssrn.com/abstract=4411537.

[84] Rostam J. Neuwirth, (n 82), pp. 6-7. Vera Lúcia Raposo, 'Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence' International Journal of Law and Information Technology vol. 30 (2022) pp. 93-94.

[85] On the subject, mainly from an economic perspective, see

I believe these situations would not fall within the scope of this article of the Regulation, although they could still be considered illegal on other grounds.

II. Social *scoring*

55 The practice of *scoring*, i.e. assigning numerical values to individuals, although not defined, is already covered by the GDPR, as it almost always involves profiling and frequently also an automated decision. This operation is often necessary so that computer systems can perform their functions. However, it raises concerns, especially considering what certain countries, such as India and China, have implemented: social classification systems, which take into account the generality of citizens' behavior in order to assign a classification that determines or influences their treatment in various contexts.[86]

56 The Regulation only prohibits AI systems "*for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following* (...) *detrimental or unfavourable treatment* (...) *in social contexts that are unrelated to the contexts in which the data was originally generated or collected* (...) [or] *that is unjustified or disproportionate to their social behaviour or its gravity*" (art. 5/1/c)). What is at stake is what is known as *general social scoring, i*.e. the overall assessment of a natural person's behavior.[87] On the other hand, AI systems that do more restricted *scoring,* such as those dedicated to credit scoring, solvency assessment or risk assessments and the pricing of life or health insurance, will be classified as high-risk (Annex III, 5/b) and c)).[88] Finally, systems that score for the purposes of detecting financial fraud or for setting prices in car insurances will not even be covered by the Regulation. Again, what determines the risk classification of the system is the purpose of the quantitative assessment and not the practice of scoring itself.

57 As has been pointed out, *scoring* is usually associated with an automated decision, which, when involving the processing of personal data and producing legal effects concerning or significantly affecting the personal data subject, may from the outset be prohibited under art. 22 GDPR.[89] However, it is important to note that art. 22 of the GDPR only applies to *fully* automated decisions.[90] Therefore, at least in the case of high-risk systems, where the regulation requires human supervision (art. 14 AIA), it is possible to escape the application of this GDPR rule.

Mateusz Grochowski / Fabrizio Esposito /Antonio Davola, *Price 'Personalization vs. Contract Terms Personalization: Mapping the Complexity* (2024) in https://ssrn.com/abstract=4791124. irective (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council in order to ensure better enforcement and modernization of Union rules on consumer protection, has imposed an obligation to provide information on whether prices are determined automatically.

[86] Cfr. Ralph Schroeder, 'Aadhaar and the Social Credit System: Personal Data Governance in India and China' International Journal of Communication vol. 16 (2022) pp. 2370-2386.

[87] Nizan Geslevich Packin, 'Disability Discrimination Using Artificial Intelligence Systems and Social Scoring: Can We Disable Digital Bias?' Journal of International Comparative Law (2021) p. 496: "*Social scoring, however, attempts to systematically rate people in their entirety (and not just their creditworthiness) based on social, reputational and even behavioral features (as opposed to credit history)*". On the phenomenon see Danielle Keats Citron / Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' Washington Law Review 89 (2014) pp. 1-33.

[88] Vera Lúcia Raposo, (n 84) p. 94 points out that the reference to "a certain period of time" will exclude episodic scoring.

[89] In C-634/21, *Schufa*, (EU:C:2023:957), §44-46 the Court of Justice adopted a broad concept of decision, saying that a *credit score* qualified as such.

[90] The standard requires " *...three cumulative conditions, namely, first, that there must be a 'decision', secondly, that that decision must be 'based solely on automated processing, including profiling', and, thirdly, that it must produce 'legal effects concerning [the interested party]' or 'similarly significantly [affect] him or her'..*" (C-634/21, *Schufa,* §43). The EDPB, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (2018) p. 21 point out that merely symbolic human intervention is not enough. A decision is not considered fully automated when there are organizational measures that ensure substantial and structured human involvement. In case law, see the decision of the Rechtbank Amsterdam of 11.III.2021 (ECLI:NL:RBAMS:2021:1018,) in which the requirement of a consensus between several people was at issue), the decision of the Rechtbank Den Haag of 11.II.2021 (NL:RBDHA:2020:1013) in which a right of veto was provided for and a decision of the Austrian Bundesverwaltungsgericht of 18.XII.2020 (AT:BVWG:2020:W256.2235360.1.00) in which there were training and guidelines for dealing with the recommendation produced by the system. For more case law see Sebastião Barros Vale / Gabriela Zanfir-Fortuna, *Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities* (Future of Privacy Forum 2022).

III. Biometric identification and classification, including sentiment detection

58 Biometric identification systems, especially those for facial and emotion recognition, generated significant controversy during the legislative process. From the outset, these systems constitute an attack on individual privacy and freedom, with a high discriminatory potential. In this sense, companies such as Clearview.AI, which systematically scraped the Internet (especially social networks) to generate a facial recognition database, had already been sanctioned for violating the GDPR.[91] In any case, the Regulation now expressly prohibits this practice (art. 5/1/e)).

59 The use of emotion recognition systems has been challenged on technical grounds. It is argued that expressions are variable at an individual level and depend on the social and cultural context, so these systems are not reliable. In addition, they have a high discriminatory potential.[92] Paradoxically, the Regulation only prohibits the use of these emotion recognition systems in the context of work and education.[93] In all other cases, emotion recognition systems are considered high-risk systems (Annex III/1/c). On the other hand, the ban does not cover "*AI systems placed on the market strictly for medical or safety reasons, such as systems intended for therapeutical use*". This will raise questions in cases where systems are used for safety or medical reasons in the areas of workplace and education institutions. In that scenario, the intention seems to be allowing the use of such systems. Automatic interview systems should be classified as high-risk (Annex III,4), unless they also include an emotion recognition component.[94]

60 On the other hand, the very notion of emotion recognition must be read restrictively. Recital 18 explains: "*The notion refers to emotions or intentions such as happiness, sadness, anger, surprise, disgust, embarrassment, excitement, shame, contempt, satisfaction and amusement. It does not include physical states, such as pain or fatigue, including, for example, systems used in detecting the state of fatigue of professional pilots or drivers for the purpose of preventing accidents. This does also not include the mere detection of readily apparent expressions, gestures or movements, unless they are used for identifying or inferring emotions. Those expressions can be basic facial expressions, such as a frown or a smile, or gestures such as the movement of hands, arms or head, or characteristics of a person's voice, such as a raised voice or whispering.*".

61 While biometric identification in public spaces can serve laudable purposes (e.g. finding missing persons or fugitives), its operation implies the compression of citizens' privacy and the creation of a state of constant surveillance, intolerable in a democracy with European values.[95] In this sense, in 2023, in a unanimous decision, the European Court of Human Rights confirmed that the use of facial recognition technology to identify, locate, and arrest an individual in an administrative offense proceeding was unlawful (in violation of art. 8 of the ECHR).[96]

62 The approach of art. 5/1/h) is to prohibit the use of these systems of "*real-time' remote biometric identification systems in publicly accessible spaces for the purposes of law enforcemen*t ", except when strictly necessary for one of three objectives: "(*i) the targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;* or *(iii) the localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences referred to in Annex II and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years*." In such cases, art. 5/2 requires a fundamental rights impact assessment (art. 27) and registration (art. 49), and art. 5/4 specifies that the relevant market surveillance and data protection authorities must be notified of such

[91] The company was subject to fines of 20 million euros in France (2021, there was also a penalty payment of five million in 2023), Greece (2022) and Italy (2022). In 2023, the Austrian authority also considered this company's activity to be in breach of the GDPR, but did not impose any fines or other measures. In 2021, the Swedish supervisory authority fined police authorities for using Clearview's services. On the other hand, in the UK, the same company succeeded, in a court decision of 17.X.2023, in overturning the fine imposed, based on a question of jurisdiction and applicable law, particularly in light of *Brexit* - [2023] UKFTT 00819 (GRC).

[92] See recital 44.

[93] Both teaching and work can be done remotely, but I believe these situations are covered by the ban. What matters is the context, not the location.

[94] With a very critical view of these systems see Ifeoma Ajunwa, 'Automated video interviewing as the new phrenology' Berkeley Technology Law Journal vol. 36 (2021) pp.1173-1225.

[95] This matter is already regulated by Directive 2016/680. On the subject cfr. Vera Lúcia Raposo 'Look at the camera and say cheese': the existing European legal framework for facial recognition technology in criminal investigations' Information & Communications Technology Law, 33(1) (2024) pp. 1-20.

[96] *Glukhin v. Russia*, 11519/20 (decision of 4.VII.2023).

use.

63 It should be noted that remote biometric identification for other purposes or on a delayed basis is not prohibited,[97] and is generally classified as a high-risk use, except in the case of simple identity recognition and verification systems (Annex III, 1, a)).[98]

64 The Regulation also deals with biometric *categorization*, which differs from biometric *identification*. While in identification the aim is to determine who the person is, starting from certain physical, psychological or behavioral characteristics (biometric data - art. 3/34) to arrive at an individual; biometric categorization aims to classify the subject - to know if someone has a given characteristic.[99] Thus, in biometric *identification*, the system will know from my face that I am Nuno Silva, in biometric *categorization,* from the way I walk, the system will determine whether I have a risk of developing Alzheimer's or, by analyzing my face, it will assess whether I am a dangerous anarcho-syndicalist.

65 According to the AIA, biometric categorization systems " *that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation*" are prohibited (art. 5/1/g)).[100] Thus, systems such as the controversial neural network that allegedly detected people's sexual orientation from photographs will not be admissible.[101] There is, however, a caveat for processing and categorizing biometric data in the field of law enforcement, which remains admissible.[102] On the other hand, "*AI systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics*" are not prohibited, but are classified as high-risk systems (Annex III, 1, b)).

IV. Predictive policing

66 The definition of profiles is based on the repeatability and standardization of behaviour. It is based on the idea that the past repeats itself in the future and that there are certain features of individuals that have predictive capacity. The application of these techniques in the criminal context raises special concerns, especially given the potential consequences of an error or injustice and the presumption of innocence.[103]

67 Thus, the Regulation prohibits predictive policing practices that use AI systems to assess the risk of a natural person committing a criminal offense "*based solely on the profiling of a natural person or on assessing their personality traits and characteristics*" (art. 5/1/d)).

68 However, "*this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity*." In other words, the system must consider the concrete behavior and particular traits of a specific person and not their membership in certain categories or groups. This exception recognizes the potential usefulness of AI in the context of criminal investigation and prevention while ensuring that the assessment is based on actual data and not *exclusively*

[97] art. 26/10 states that, in the case of post-remote biometric identification systems (defined in art. 3/43, as opposed to "real-time" systems defined in art. 3/42), "*the deployer* (...) shall *request an authorisation, ex ante, or without undue delay and no later than 48 hours, by a judicial authority or an administrative authority whose decision is binding and subject to judicial review, for the use of that system, except when it is used for the initial identification of a potential suspect based on objective and verifiable facts directly linked to the offence. Each use shall be limited to what is strictly necessary for the investigation of a specific criminal offence*" If authorization is rejected, use must cease and the data must be destroyed. It also prohibits indiscriminate use ("non-selective") and allows member states to adopt more restrictive legislation.

[98] See recitals 15, 17 and 52 and the definition of biometric verification (art. 3/36).

[99] Biometric categorization system is defined in art. 3/40 as " *an AI system for the purpose of assigning natural persons to specific categories on the basis of their biometric data, unless it is ancillary to another commercial service and strictly necessary for objective technical reasons.*" (for examples of ancillary categorization see Recital 16), while biometric identification concerns the "*automated recognition of physical, physiological, behavioural, or psychological human features for the purpose of establishing the identity of a natural person by comparing biometric data of that individual to biometric data of individuals stored in a database*" (art. 3/35).

[100] This provision can be criticized for being too restrictive in the "protected categories".

[101] The controversial original study has been replicated by John Leuner, 'A replication study: Machine learning models are capable of predicting sexual orientation from facial images' *arXiv:1902.10739* (2019), who argues that these models take into account other factors and not facial physiognomy/structure.

[102] See Recital 30.

[103] As can be read in recital 42: " *In line with the presumption of innocence, natural persons in the Union should always be judged on their actual behaviour. Natural persons should never be judged on AI-predicted behaviour based solely on their profiling, personality traits or characteristics, such as nationality, place of birth, place of residence, number of children, level of debt or type of car, without a reasonable suspicion of that person being involved in a criminal activity based on objective verifiable facts and without human assessment thereof.*"

on (necessarily speculative) profiling.

69 In fact, predictive policing can be geared towards predicting crimes, predicting or identifying criminals, and/or predicting or identifying potential victims of crime.[104] Most of these systems, when not based exclusively on profiling, will fall under the high-risk classification (Annex III, 6). In this vein, recital 42 clarifies that the prohibition of art. 5/1/d) does not cover "*AI systems using risk analytics to assess the likelihood of financial fraud by undertakings on the basis of suspicious transactions or risk analytic tools to predict the likelihood of the localisation of narcotics or illicit goods by customs authorities, for example on the basis of known trafficking routes*".

70 A well-known example of an AI system for predictive purposes in the criminal context is the COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) system, used in some US courts to calculate the risk of recidivism and, on that basis, define sentencing.[105] Tools like this, if they are not based exclusively on profiling, are not covered by the ban but are considered high-risk AI systems (Annex III, 6 d) and e) and 8)).

G. High-risk systems

I. Qualification

71 The definition of high-risk systems is made in article 6 by reference to two Annexes.[106]

72 Annex I includes legislation on certain categories of products (such as toys, vehicles, explosives, elevators, or medical devices) and, according to art. 6/1, when AI systems are used as safety components in these products (or the AI systems are themselves products[107]) subject to a conformity assessment obligation, this is a high-risk system.[108] It is important to note that within Annex I, there are two sections: section A (legislation issued under the new legislative framework) and section B (prior legislation). According to art. 2/2, the latter, i.e. systems under section B, are practically excluded from the scope of the AIA (although AIA provisions will still apply by reference).

73 In turn, art. 6/2 refers to Annex III, which specifies certain uses such as biometric identification, management of critical infrastructures, admission and classification in educational establishments, job interviews, monitoring of workers, access to and use of (public and private) essential services, use in border control, in a judicial context or by law enforcement agencies. As Philip Hacker points out,[109] more important than the context of use is the purpose - a system used for medical operations or triage does not carry the same risk as a system that manages medical appointments.

74 The law works with auto-classification, i.e. each operator will determine the risk classification of their system. It is important to read the various hypotheses carefully and consider the Regulation's recitals. The Commission will adopt guidelines specifying the practical application of this article "*together with a comprehensive list of practical examples of cases of use of high-risk and non-high-risk AI systems*" (art. 6/5).

75 The risk classification is based on the intended use, but there are some caveats. For example, remote biometric identification systems are generally high-risk, but there is an exclusion for identity verification systems (Annex III(1)(a)). Similarly, systems for

[104] Walter Perry et al, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (RAND Corporation 2013).

[105] The subject of much academic and judicial discussion. In the well-known *Loomis v. Wisconsin,* 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S. Ct. 2290 (2017) the Wisconsin Supreme Court rejected the appeal of an individual who had been considered by the software to have a high-risk of recidivism and thus sentenced to 6 years in prison. According to the court, *due process* had not been violated despite the fact that the sentence had been determined using COMPAS, whose algorithm and mode of operation is unknown. The discriminatory nature of this system was the subject of a controversial report by ProPublica.

[106] The reason for this definition being made by reference is to make it easier to update these annexes in the simplified procedure (delegated acts of the European Commission) provided for in arts. 6/6, /7 and /8 and 7. For a critical overview of this classification and the many ways in which it is narrowed, see Emilija Leinarte, 'The Classification of High-Risk AI Systems Under the EU Artificial Intelligence Act' Journal of AI Law and Regulation vol. 1(3) (2024) pp. 262-280. She highlights that art. 6(1) "covers a limited group of AI systems due to significant sectoral carve-outs, limitations to sector-specific definitions of products and safety components of a product and a significant harm condition." (p. 274), believes that art. 6(3) is likely to have a material impact (p. 278), concluding that "large categories of technology which pose ethical and fundamental rights concerns" (p. 279) are left out of the high-risk classification.

[107] This can happen namely with toys or medical devices.

[108] "Safety component" is defined in art. 3/14 as "*a component of a product or of an AI system which fulfils a safety function for that product or AI system, or the failure or malfunctioning of which endangers the health and safety of persons or property." This definition is broad, but it should be read using* a normality/predictability criterion (recital 46 seems to confirm this by requiring "significant harmful impact").

[109] *AI Regulation in Europe: From the AI Act to Future Regulatory Challenges* (2023) arXiv:2310.04072 p. 7.

assessing the creditworthiness of natural persons or credit scoring are high-risk systems, except when such systems are used for the detection of financial fraud (Annex III(5)(b)).

76 In addition to specific exceptions, there is a more general derogation. According to art. 6/3, it is possible to disregard the high-risk classification for a system whose foreseeable use is listed in Annex III "*if it does not 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*" and provided that it does not carry out profiling of natural persons (last paragraph of this art. 6/3). However, this does not mean that all profiling AI systems are deemed high-risk. All that it means is that *prima facie* high-risk systems that carry out profiling will not be able to invoke the exemption.

77 Art. 6/3 sets out circumstances in which AI systems, despite having a purpose set out in Annex III, will not pose a significant risk: a) when they perform a narrow procedural task; b) when they are intended to improve the result of a previously completed human activity; c) when they aim to detect decision-making patterns or deviations from previous decision-making patterns and are not intended to replace or influence a previously completed human assessment; or d) when the AI system is intended to perform (only) a preparatory task. For an AI system not to be considered high-risk, despite its intended purpose, it is sufficient to meet one of these points and not to carry out profiling (as defined in art. 4/4 GDPR).

78 Recital 53 gives some examples of such systems, including AI systems designed to improve the language, professional tone, or style in previously drafted documents, systems that are used to check whether a teacher may have deviated from their usual pattern of awarding marks, intelligent file handling solutions, or AI systems used for document translation.

79 In any case, anyone wishing to invoke this derogation must document this assessment (art. 6/4) and register it (art. 49/2). A market surveillance authority may, however, disagree and demand corrective action (art. 80).

II. Rules

80 In simple terms, the AIA requires high-risk systems to be well-made, properly maintained, and adequately controlled. The operators must have documentation to prove compliance with the Regulation's rules.

81 Machine learning systems are subject to data quality requirements, particularly in terms of representativeness and the application of measures to detect and mitigate *biases* (art. 10). Article 10(5) even creates a new basis for the lawful processing of sensitive data (in addition to those in art. 9 GDPR) by establishing that, under specific conditions, it will be possible to process special categories of personal data "*to ensure bias detection and correction* ".[110] On the other hand, most of the Regulation's provisions will legitimize the processing of non-sensitive data since this will occur in order to comply with legal obligations (art. 6/1/c) GDPR).[111]

82 Providers of high-risk systems are responsible for meeting the requirements of articles 8 to 15 (art. 16), as well as ensuring the existence of a quality management system (art. 17), keeping documentation for a period of 10 years after the system has been placed on the market or put into service (art. 18), and maintaining logs (art. 19). There is also a duty to cooperate with competent authorities (articles 20/2, 21 and 73), to adopt corrective measures (article 20/1), and perform post-market monitoring (article 72). This monitoring includes a duty to inform the authorities in the event of a serious incident (art. 73), defined in art. 3/49 as "*any incident or malfunctioning in an AI-system which, directly or indirectly, has any of the following consequences: (a) death of a person or serious harm to a person's health (b) a serious and irreversible disruption of the management or operation of a critical infrastructure, (c) infringement of obligations under Union law designed to protect fundamental rights, (d) serious harm to property or the environment*".

83 From a more bureaucratic point of view, in addition to a duty of documentation and record-keeping, providers of high-risk AI systems are obliged to identify themselves as such (art. 16/b)) and to follow a conformity assessment procedure (art. 43),[112] including drawing up a declaration of conformity (art. 47), using the CE marking (art. 48) and registering the high-risk system (arts. 49 and 71).[113]

84 Although the most important duties fall on the

[110] This may make it difficult to apply bias mitigation measures to systems that are not high-risk, since for these there will be no lawful basis for processing sensitive data (Michael Veale / Frederik Zuiderveen Borgesius, (n 30), p. 103). Additional processing of personal data is also provided for under certain conditions to safeguard the public interest (art. 59).

[111] There is no equivalent basis for sensitive data, hence the need for article 10(5) AIA.

[112] A derogation from this procedure is provided for, particularly in cases of urgency (art. 46).

[113] Taking into account the principles of country of origin and mutual recognition, this operation only needs to be carried out in one Member State.

providers of AI systems, their *users* ("deployers") are also subject to several obligations set out in art. 26. To the extent that they control the system, deployers will have to respect the instructions for use of the AI system, ensure its human supervision and the quality and appropriateness of the input data, collaborate with the authorities, keep records of the system's operation and inform natural persons that they are subject to the use of the high-risk AI system.

85 In some cases, bodies governed by public law or private entities providing public services, as well as banks and insurance companies, must carry out a fundamental rights impact assessment (art. 27). This assessment is not to be confused with the obligation to carry out a data protection impact assessment laid down in art. 35 of the GDPR, although the Regulation itself recognizes the existence of partial overlaps (art. 27/6 AIA).

H. Transparency obligations

86 Article 50 AIA, the only one in Chapter IV, deals with certain systems defined in the light of their purpose, imposing minimum transparency/information requirements.[114] The first two paragraphs of this article impose duties on providers, while paragraphs 3 and 4 concern the duties of deployers.[115] These duties apply to the AI systems mentioned in art. 50, regardless of their risk classification.

87 Article 50/1 regulates AI systems "*intended to interact directly with natural persons*", i.e. so-called chatbots or conversational systems. These systems must be designed in such a way that it is clear to natural persons "*that they are interacting with an AI system, unless this would be obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use* ".

88 Generative AI systems ("*generating synthetic audio, image, video or text content*") are addressed in art. 50/2. There is an obligation to identify such synthetic content with a digital "watermark" "*in a machine-readable format and detectable as artificially generated or manipulated*".[116]

89 Those responsible for implementing emotion recognition or biometric categorization systems are subject to a duty to disclose its use (art. 50/3).[117] Similarly, those who create *deepfakes* must "disclose *that the content has been artificially generated or manipulated*" (art. 50/4 1st paragraph). [118] This duty can be compressed "*where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or program*". In that scenario; it is sufficient that the disclosure is done in "*an appropriate manner that does not hamper the display or enjoyment of the work.*". The duty of disclosure also exists in the case of news ("*text which is published with the purpose of informing the public on matters of public interest* "), except when the " *AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content* " (art. 50/4/2nd paragraph).

I. General purpose models

90 When the European Commission presented the proposal for a Regulation in April 2021, there were already some AI models with diversified capabilities, but the term "foundational models", used to indicate those models trained with large amounts of data and with the potential for various applications, had not yet been coined. It wasn't until August 2021 that a paper by Stanford researchers used this notion for the first time.[119] The real explosion of foundational models, which include GPTs from OpenAI and competitors PALM, BERT and Gemini (Google),

[114] The duties of transparency/disclosure set out in art. 50 do not apply when the system is legally authorized "*to detect, prevent, investigate or prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties* ".

[115] It is unclear whether the manufacturers of these systems are covered by the exemption from liability in art. 6 of the DSA. From the outset, it is debatable whether we can classify providers of general-purpose or generative AI models or systems as an "intermediary service" (as provided for in art. 3/g) of the DSA). Recital 119 of the AI Act seems to point to a case-by-case assessment.

[116] Watermarking solutions must be "*effective, interoperable, robust and reliable as far as this is technically feasible, taking into account the specificities and limitations of various types of content, the costs of implementation and the generally acknowledged state of the art, as may be reflected in relevant technical standards* ". This obligation does not apply to editing support tools (such as a spelling checker) and in general those that "*do not substantially alter the input data provided by the deployer or the semantics thereof*" (art. 50/2).

[117] As we have seen, this type of system can be banned or classified as high-risk. In any case, as Michael Veale / Frederik Zuiderveen Borgesius point out, (n 30) p. 107, this duty does not seem to add anything to what already results from the GDPR.

[118] Michael Veale / Frederik Zuiderveen Borgesius, (n 30) point out that a teleological understanding of this obligation should except uses in contexts where there is no risk of deception (as in the case of generic images used for marketing or presentation purposes). Recitals 132 and 133 seem to support this interpretation.

[119] Rishi Bommasani et al, *On the Opportunities and Risks of Foundation Models*, arXiv:2108.07258 [cs.LG].

Claude (Anthropic), Luminous (Aleph Alpha), Mistral 7B and LlaMA (Meta) took place in 2023.

91 This technology has particularities that are especially challenging. On the one hand, those models have high development costs, which create considerable barriers to entry. Unlike the specialized systems for which the Regulation was initially intended, these models have a capacity for generalization and will often be made available through programming interfaces (APIs) so that third parties can optimize and adapt them to specific applications. In this sense, these models, as Andrej Karpathy explains,[120] are close to operating systems, generating considerable dependencies. These considerations are typically addressed by Competition Law,[121] but the AIA has dedicated a chapter to them. arts. 89/2 and 93 provide for the protection of downstream providers, i.e., those who integrate a general-purpose model or system into their system and who become dependent on a general-purpose system that they do not control.

92 On the other hand, these large general-purpose models are often opaque: they are a vast array of numbers (the so-called parameters and weights of a neural network) that interact in ways that are beyond human comprehension. This lack of understanding raises concerns of security, control, and alignment.

93 In addition, developing these models requires massive amounts of data, much of which is taken from the Internet and includes personal data and data protected by intellectual property rights. Furthermore, contrary to what was initially thought, these models retain some of the data in "memory".[122] This makes assessing the lawfulness of these uses of material protected by third party rights even more complex.

94 Finally, most foundational models have "creative" capacities and, thus, also fall into the category of generative AI covered by art. 50.[123]

95 The Regulation deals with general purpose AI models (in arts. 53 and 54) and imposes additional duties (in art. 55) for so-called general purpose AI models with systemic risk.[124] According to art. 51, systemic risk exists if the model has "*high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks* " (51/1/a)) or equivalent capabilities or impact taking into account the criteria set out in Annex XIII, on the basis of a decision by the Commission, *ex officio* or following a qualified alert by the scientific panel (51/1/b)). "High impact capabilities" is defined as "*capabilities that match or exceed the capabilities recorded in the most advanced general purpose AI models*" (art. 3/64). In other words, in this matter, the law maker essentially refers to technical-scientific criteria set out in Annex XIII, which will be fleshed out by the Commission in delegated acts (art. 51/3). In any case, art. 51(2) establishes a (rebuttable) presumption that the model has high impact capabilities when the cumulative amount of computation used for its training, measured in floating point operations per second (FLOPS), is greater than 10^{25}.[125]

96 Article 52 sets out the procedure for classifying a model as having systemic risk, in which the provider "*may present (...) sufficiently substantiated arguments to demonstrate that, exceptionally, although it meets this requirement, the general purpose AI model does not present, due to its specific characteristics, systemic risks and, therefore, should not be classified as a general purpose AI model with systemic risk*" (art. 52/2).

97 Providers of general-purpose AI models are essentially subject to four duties set out in art. 53: i) to maintain appropriate and up-to-date technical documentation (paragraph 1/b) and Annex XI); ii) to facilitate integration and interoperability with their system (paragraph 1/b and, Annex XII); iii) to apply a policy of respect for copyright (paragraph 1/c)), in particular ensuring that the system respects the reservation of rights provided for in art. 4 of Directive 2019/790 in the context of text and data mining[126] and (iv) make publicly available a summary

[120] This statement is made in several public lectures available on Youtube. I especially suggest the video "[1hr Talk] Intro to Large Language Models".

[121] Cf. Hou Liyang, 'The Essential Facilities Doctrine - What was Wrong in Microsoft?' IIC 43(4) [2012] pp. 251-271.

[122] Milad Nasr et al, *Scalable Extraction of Training Data from (Production) Language Models*, arXiv:2311.17035 [cs.LG].

[123] Not all generative AI systems are foundational models; there are a number of specialized applications for creating music, images, text, etc.

[124] As already mentioned, general purpose AI models are defined in art. 3/63. "Systemic risk", in turn, is defined as "*a risk specific to the high-impact capabilities of general purpose AI models that have a significant impact on the Union market due to their reach or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights or the society as a whole, that can be propagated at scale across the value chain*" (art. 3/65).

[125] Floating-point operations are defined in art. 3/67 as "*any mathematical operation or assignment involving floating-point numbers, which are a subset of the real numbers normally represented in computers by an integer of fixed precision scaled by an integer exponent of a fixed base*". In this context, this value is a measure of the performance and computational capacity of the *hardware* used to train a given AI model. The higher it is, the greater the complexity of the models and the corresponding training costs. Interestingly, the US Executive Order uses 10^26 FLOPS as the threshold, i.e. ten times more.

[126] The Regulation devotes recitals 105 to 108 to the subject of

of the content used to train the model (paragraph 1/d), which provides for a model of this summary to be drawn up by the AI Office). Furthermore, there is a general duty to cooperate with the authorities (art. 53/3).

98 In the case of models with systemic risk, in addition to the duties applicable to all general purpose models, art. 55(1) stipulates that the respective providers must: a) carry out tests and evaluations of the model with a view to identifying and mitigating systemic risks; b) assess and mitigate any of those risks; c) monitor, document and communicate relevant information on serious incidents and any corrective measures to resolve them; and d) ensure an adequate level of protection in terms of cybersecurity.

99 If the model providers are established in third countries (outside the EU), an authorized representative will carry out these duties, as established in art. 54.

J. Certification, supervision, and sanctions

100 The AIA establishes preventive and repressive measures, although it essentially focuses on the placing on the market or putting into service of high-risk AI systems. Although civil liability is not directly addressed,[127] some of the Regulation's rules if breached, could give rise to liability under national rules. In addition, there is a reference to the possibility of collective claims pursuant to Directive 2020/1828 (art. 110).

101 Since this is product safety legislation, articles 28 ff. provide for a certification and control scheme. There will be at least one national notifying authority[128] and a national market surveillance authority (art. 70), which will be the competent national authorities under the terms of the AIA.[129]

102 The notifying authority is the one that assesses, designates, and supervises the conformity assessment bodies: typically, independent private entities that carry out testing, certification, and inspection activities on the systems to ensure that they meet the requirements of the Regulation. Notified bodies are a special category of officially designated conformity assessment bodies with CE marking competence.[130]

103 As provided for in arts. 40 and 42, the European standardization organizations will develop standards that will be adopted by the European Commission under Regulation (EU) No 1025/2012. Following these standards in a high-risk AI system will give rise to a presumption of conformity (arts. 40/1 and 42).[131]

104 National market surveillance authorities will deal with complaints (art. 85) and serious incidents (art. 73) and exercise the powers provided for in Regulation 2019/1020 (art. 74), including risk assessments, imposing corrective measures (art. 79), detecting non-compliance (art. 83) and supervising tests in real conditions (art. 76). It is also expected that these will be the authorities with sanctioning powers.

105 At European level, the Commission, through its *AI Office* (arts. 3/47 and 64),[132] will supervise general-

copyright. See Alexander Peukert, 'Copyright in the Artificial Intelligence Act - A Primer' GRUR-Int vol. 73(6) (2024) pp. 497-509 and very comprehensive João Pedro Quintais, Generative AI, Copyright and the AI Act (v.2) (November 01, 2024). https://ssrn.com/abstract=4912701

[127] As mentioned, this issue is addressed in two Directives still at the proposal stage: COM(2022)495 final and COM(2022)496 final.

[128] The definition of notifying authority ("*the national authority responsible for setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring*") is set out in article 3/19.

[129] Arts. 3/48 and 74. Different Member States have taken different approaches. Some, like Spain, have created a new authority. Others have preferred a decentralized system, using only sectoral regulators. Some have sought to assign these powers to existing authorities, such as the supervisory authorities in the field of data protection or the digital services coordinators under the DSA. In the case of EU activities subject to the Regulation, the supervisory authority will be the European Data Protection Supervisor (art. 74/9), who will also have the power to impose fines (art. 100).

[130] See art. 3/21 and /22 and in more detail the *2022 Blue Guide on the application of EU rules on products* (2022/C 247/01). The European Commission maintains a list of notified bodies, known as NANDO (https://webgate.ec.europa.eu/single-market-compliance-space/#/notified-bodies).

[131] The European standardization bodies are the European Committee for Standardization (CN), the European Committee for Electrotechnical Standardization (CENELEC) and the European Telecommunications Standards Institute (ETSI). There is also provision for the Commission to adopt common specifications if these organizations fail (art. 41). The request to issue standards relating to this Regulation was already submitted by the Commission to the CN and CENELEC in May 2023 (C(2023)3215 - Standardization request M/593). On the process and the role of standards in the Regulation *see* Marta Cantero Gamito / Christopher T Marsden, 'Artificial intelligence co-regulation? The role of standards in the EU AI Act' International Journal of Law and Information Technology, vol. 32 (1) (2024).

[132] This department of the European Commission was created by Commission Decision of January 24, 2024 (C(2024) 390 final).

purpose AI models, functioning for this purpose as a market surveillance authority (art. 75), with extensive supervisory powers (arts. 88 to 94) and the power to impose fines (art. 101). In addition to the *AI Office*, there is also a *European AI Board* (art. 65), made up of a representative from each Member State, whose main function is to coordinate the application of the Regulation between the various States (art. 66). The AI Office and the AI Board will be assisted by an advisory forum (art. 67) and a scientific panel of independent experts (art. 68).[133]

106 Sanctions vary according to the type of infringement, must take into account the specific circumstances (art. 99/7), and may include warnings and non-pecuniary measures (art. 99/1).[134] There are fines of up to 7% of worldwide turnover or 35 million euros in the case of prohibited practices (art. 99/3), up to 3% of turnover or 15 million euros for general infringements (art. 99/4) and up to 1% or 7.5 million euros in the case of providing "incorrect, incomplete or misleading" information to notified bodies and competent authorities (art. 99/5).[135]

107 The fact that an entity is sanctioned under the Regulation does not prevent other fines from being imposed, namely for violating the GDPR or the DSA.

K. Conclusion

108 Based on this analysis, the Regulation contains generally balanced and reasonable solutions. However, given its length, complexity, and poor legislative quality, it will become difficult to implement.[136] There is, therefore, a real risk that the European Union will negatively affect innovation and investment in the field of Artificial Intelligence. It is also possible that there will be a reduction in the supply and/or divergence of products or services, with the European public receiving different and less advanced versions.[137] As Migel Peguera Poch writes,[138] the Regulation is a remarkably complex instrument with unpredictable effects.

109 The main hope lies in the use of standards, whose mass adoption could significantly reduce *compliance* costs and reduce the considerable uncertainty that this legislative instrument will inevitably generate.[139]-[140] Another contribution to overcoming the limitations of this piece of legislation will have to come from lawyers.

[133] On this institutional framework see Claudio Novelli et al, 'A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities' (2024) at https://ssrn.com/abstract=4817755.

[134] Although the Regulation does not expressly mention it, it seems that the broad understanding of "undertaking" from Competition Law, which has been used in digital regulation, namely in data protection law and digital platforms, especially for sanctioning purposes, should apply. The definition is "*any entity engaged in an economic activity, regardless of the legal status of that entity and its method of financing*" (see C-138/11, *Compass-Datenbank GmbH*, EU:C:2012:449, §35).

[135] For providers of general purpose AI models, the framework is the same (art. 101). Interestingly, in the case of European authorities, the maximum amount is only 1.5 million euros for prohibited practices (art. 100/2) and 750,000 euros in other cases (art. 100/3). More important is the possibility given to Member States to "define rules to determine the extent to which fines may be imposed on public authorities and bodies established in that Member State." (art. 99/8). In other words, as with the GDPR, it seems legally permissible to exempt public bodies from fines. The best example comes from above...

[136] Michael Veale / Frederik Zuiderveen Borgesius, (n 30)

[137] Luciano Floridi, (n 44) : "*fridges, dishwashers, washing machines and even vehicles may need to remain on the safe side of "artificial stupidity" to avoid having to comply with the AI Act (CP version). A scenario becomes plausible in which companies start dumbing down ("de-AI-ing") or at least stop smartening up their products in order not to be subject to the AI Act.*". This does not appear to be fiction - witness Apple's recent announcement not to offer AI technology ("Apple Intelligence") in the European Union for fear of violating the Digital Markets Regulation - Regulation (EU) 2022/1925 (https://www.theverge.com/2024/6/21/24183251/apple-eu-delay-ai-screen-mirroring-shareplay-dma) and Meta's announcement not to offer a more advanced model in view of the "too unpredictable nature" of the European regulatory environment (https://www.theverge.com/2024/7/18/24201041/meta-multimodal-llama-ai-model-launch-eu-regulations).

[138] 'La propuesta de Reglamento de AI: una intervencióin legislativa insoslayable en un contexto de incertidumbre' in Migel Peguera Poch (coord.), *Perspectivas Regulatios de La Inteligencia Artifical en La Unión Europea* (Reus 2023) p. 179.

[139] The Regulation itself acknowledges this in recital 121, which reads: "*Standardization should play a key role in providing providers with technical solutions that ensure compliance with this Regulation, in line with the state of the art, in order to promote innovation, competitiveness and growth in the single market*.". For a non-exhaustive list of *standards* applicable in this context *see* Federico Marengo, (n 78) pp. 196 ff. and Alessio Tartaro, 'Regulating by standards: current progress and main challenges in the standardization of Artificial Intelligence in support of the AI Act' European Journal of Privacy Law and Technologies (2023) pp. 147-174.

[140] Some authors, including Emilija Leinarte (above n 105) and Sandra Wachter, 'Limitations and loopholes in the EU AI Act and AI Liability Directives: what this means for the European Union, the United States, and beyond.' Yale Journal of Law and Technology 26.3 (2024) pp. 671-718, take the view that the AI Act is a watered-down version of what it should be, having a narrow scope. Athough I partially agree, that does not alter the significant uncertainty generated.

The Regulatory Landscape of Health Apps in the European Union

by **Liga Svempe** *

Abstract: Digital tools, including numerous health apps, have become integral to our daily lives. However, the fact that many of these solutions are unregulated raises concerns related to their quality and safety. The current Medical Device Regulation 2017/745 covers devices explicitly designed for medical purposes and does not extend its regulatory scope to wellness applications beyond its intended purpose. Due to the complexity of the regulation, many manufacturers choose to avoid the certification pathway and market their products as wellness apps. As a result of this regulatory stance, the responsibility for preventing harm to users primarily lies with developers, application marketplaces, and consumers themselves. This situation is coupled with increasing consumer skepticism towards the healthcare system and growing reliance on online information, paving the way for uncontrolled and potentially hazardous market development. Real-world examples demonstrate that these non-regulated apps can be harmful; with the market expanding, this issue is likely to worsen. This article investigates the legal framework governing health apps in the European Union. I identify regulatory gaps and associated risks for public health, and propose measures to mitigate these challenges. Policymakers are advised to introduce updates to the General Product Safety Regulation or adopt national-level regulation as a short-term measure. Additionally, the author proposes revising the role and increasing the responsibilities of app marketplaces to prevent harmful apps from entering or operating in the market. Regulatory incentives, such as government reimbursement schemes, are suggested at the national level unless EU initiatives are introduced.

Keywords: Digital Health; Health Law; Health Technology; Medical Device; Wellness Apps

Recommended citation: Liga Svempe, The Regulatory Landscape of Health Apps in the European Union, 16 (2025) JIPITEC 24 para 1.

A. Introduction

1 Technological advancements during the last decades have significantly changed many industries, and they have the potential to also transform the healthcare industry, bringing new digital solutions that were unimaginable in the 90-s when the Medical Device Directive 93/42/EEC[1] (MDD) was adopted. Digital health has emerged as a separate discipline. According to the European Commission, "digital health and care refers to tools and services that use information and communication technologies (ICTs) to improve prevention, diagnosis, treatment, monitoring, and management of health-related issues and to monitor and manage lifestyle habits that impact health".[2] As stated in IQVIA Institute report,[3] in 2021 there were over 350,000 health-related mobile apps for various goals. However, the rapidly evolving market introduces not only new opportunities but also new risks, especially when it comes to their clinical effectiveness and safety, data safety, and pri-

[*] Liga Svempe is a PhD Candidate and an Acting Researcher at the Faculty of Social Sciences of the Riga Stradins University in Riga, Latvia.

[1] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices [1993] OJ L 169/1 (Medical Device Directive).

[2] 'eHealth : Digital Health and Care' (Public Health) <https://health.ec.europa.eu/ehealth-digital-health-and-care_en> accessed August 26, 2024.

[3] IQVIA Institute, 'Digital Health Trends 2021' (2021) <https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021> accessed July 18, 2024, 2.

vacy issues.

2 This study focuses on digital health apps . "Health apps" is an umbrella term defining software programs on mobile devices that process health-related data on or for users to maintain, improve, or manage health.[4] Health apps include both wellness and medical apps, the latter known as software as a medical device (SAMD) and certified as a medical device (see Figure 1).

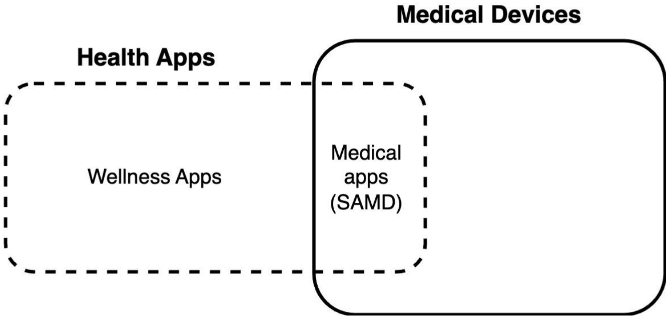


Figure 1: an overview of health app categories

3 International Medical Device Regulators Forum defines SAMD as "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device".[5] The Medical Device Regulation EU 2017/745 (MDR) similarly defines it as software intended by the manufacturer to be used for human beings for one or more specific medical purposes.[6] Such apps include CardioSignal for heart disease self-monitoring, remote care, and point-of-care diagnostics,[7] Kaia Health as a digital therapy for back pain,[8] and HelloBetter for various mental issues.[9] The key difference between a wellness app and an SAMD lies in its regulatory status - SAMD is certified as a medical device, whereas a wellness app lacks any certification or compliance with any regulations or quality standards related to healthcare. Wellness apps include, for example, BetterSleep to improve sleep quality,[10] Noom for weight management,[11] and Calm as a mental health app to help manage stress, calm anxiety, and improve sleep.[12] Today these wellness apps make up most of the health-related apps market. According to the EUDAMED database,[13] in August 2024, there were slightly over 1,900[14] software applications classified as medical devices, a small fraction of the total number of the 350,000 health apps mentioned above (the EUDAMED database is not yet fully functional therefore the actual number of SAMD would be higher).

4 However, during the last decades, numerous cases in the healthcare industry have highlighted insufficient regulatory oversight and harming the end-users (patients).[15] This along with the rapid technological advancements led to the adoption of Medical Device Regulation (EU) 2017/745,[16] whose main goal is to ensure "a high level of safety and health whilst supporting innovation".[17] However, the MDR does not currently regulate wellness apps that are not designed for medical purposes. As a result, the safety and efficacy can be poorly evaluated, potentially harming the end user.

[4] Maaß L and others, 'The Definitions of Health Apps and Medical Apps From the Perspective of Public Health and Law: Qualitative Analysis of an Interdisciplinary Literature Overview' (2022) 10(10) JMIR mHealth and uHealth e37980.

[5] International Medical Device Regulators Forum, 'Software as a Medical Device (SaMD): Key Definitions' <https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf> accessed August 9, 2024.

[6] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009, and repealing Council Directives 90/385/EEC and 93/42/EEC [2017] OJ L 117/1 (Medical Device Regulation), art 2.

[7] 'CardioSignal' (*CardioSignal*) <https://cardiosignal.com/> accessed November 26, 2024.

[8] 'Pain Relief in the Palm of Your Hand' (*Kaia Health*) <https://kaiahealth.com/> accessed November 26, 2024.

[9] 'Effective Psychological Online Courses' (*HelloBetter*) <https://hellobetter.de/en/> accessed November 26, 2024.

[10] 'BetterSleep' (BetterSleep) <https://www.bettersleep.com/> accessed November 26, 2024.

[11] 'Noom: Lose Weight and Keep It Off' (*Noom*) <https://www.noom.com/> accessed November 26, 2024.

[12] 'Experience Calm' (*Calm*) <https://www.calm.com/> accessed November 26, 2024.

[13] 'EUDAMED - European Database on Medical Devices' <https://ec.europa.eu/tools/eudamed/#/screen/search-device> accessed August 9, 2024.

[14] The search was conducted on August 9, 2024. Since the MDR transition period is ongoing, devices on the market are currently assessed either under the MDR or the MDD. Therefore, two separate searches were conducted: (1) search string included parameters "Applicable legislation: MDR (REGULATION (EU) 2017/745 on medical devices)" AND "Device types: Software" AND "Status: On the EU market"; this search returned 1392 records, (2) search string included parameters "Applicable legislation: MDD (Directive 93/42/EEC on Medical Devices)" AND "Device types: Software" AND "Status: On the EU market"; this search returned 513 records. Both searches return 1905 records in total.

[15] Such as Martindale V and Menache A, 'The PIP Scandal: An Analysis of the Process of Quality Control That Failed to Safeguard Women from the Health Risks' (2013) 106(5) Journal of the Royal Society of Medicine 173 and Cohen D, 'Faulty hip implant shows up failings of EU regulation' (2012) 345 BMJ e7163.

[16] Medical Device Regulation (n 6).

[17] Ibid, rec 1.

5 On the one hand, the absence of regulatory oversight might bring health risks to users and the latest data show that the quality of the apps is troubling (the risks are discussed and exemplified later in this paper); on the other hand, subjecting numerous digital solutions to the extensive medical device certification process, which entails significant time and financial resources, will slow their development.[18] While the exemption from regulatory scrutiny could be justifiable for applications posing minimal or no risk to human health, it simultaneously creates an open gateway for harmful applications, because the evaluation of the safety and efficacy of such applications are left to developers' discretion (developers' role in relation to marketplaces' role is discussed later in this paper).

6 The current consumer health decision-making process reveals several underlying challenges. First, there is a growing scepticism among consumers about the healthcare system, which was evident during the Covid-19 pandemic.[19] Furthermore, reliance on online information has surged, particularly among younger demographics,[20] though this varies depending on the health condition. For instance, research indicates that 65% of adolescents use online resources as their primary source for sexual advice, compared to just 8% seeking orthodontic treatment guidance. Cultural and national differences also influence the degree of reliance on online information.[21] Another study found that 56.6% of high school students had sought health information online rather than consulting a physician in person.[22] While some consideration is given to the credibility of sources, 51.9% of respondents admitted they rarely or never checked when the website was last updated or reviewed by a medical professional.[23]

7 Research indicates that the source of a message significantly influences how it is perceived, with endorsements from trusted sources enhancing the credibility of claims.[24] However, while such endorsements may change consumer attitudes, they do not necessarily translate into behavioural changes. In some cases, high-credibility labelling may have little to no impact on consumer health behaviour and, occasionally, may even have the opposite effect.[25] The author suggests further research into health decision-making, particularly within the context of digital health.

8 An increasingly important factor in health decision-making is the role of marketing, as individuals today can access information through a wide array of channels beyond traditional physician visits. Research highlights that marketing messages often include scientifically unfeasible health claims,[26] exploiting emotional vulnerabilities, which promote unrealistic consumer expectations and increase susceptibility to these misleading messages.[27] According to Pirsch et al.,[28] consumers can be categorized into three groups: the "smart consumer," who is educated, critical, and at a lower

[18] Svempe L, 'Exploring Impediments Imposed by the Medical Device Regulation EU 2017/745 on Software as a Medical Device' (2024) 12 JMIR Medical Informatics e58080.

[19] Shmerling MRH, 'What Happened to Trusting Medical Experts?' (Harvard Health, October 19, 2021) <https://www.health.harvard.edu/blog/what-happened-to-trusting-medical-experts-202110192621> accessed August 1, 2024.

[20] Gordon D, '33% Of Gen Zers Trust TikTok More Than Doctors, New Survey Shows' (Forbes, December 20, 2022) <https://www.forbes.com/sites/debgordon/2022/12/20/33-of-gen-zers-trust-tiktok-more-than-doctors-new-survey-shows/> accessed August 9, 2024; Evans N, 'Online Medical Advice: How Google and TikTok Are Shaping Patient Behaviors' *The Intake* (February 28, 2024) <https://www.tebra.com/theintake/medical-deep-dives/tips-and-trends/online-medical-advice-deep-dive-how-google-and-tiktok-are-shaping-patient-behaviors> accessed November 27, 2024.

[21] Park E and Kwon M, 'Health-Related Internet Use by Children and Adolescents: Systematic Review' (2018) 20 Journal of Medical Internet Research e120.

[22] Gazibara T and others, 'Searching for Online Health Information Instead of Seeing a Physician: A Cross-Sectional Study among High School Students in Belgrade, Serbia' (2020) 65 International Journal of Public Health 1269

[23] Park E and Kwon M (n 21).

[24] Parkinson TL, 'The Role of Seals and Certifications of Approval in Consumer Decision-Making' (1975) 9 Journal of Consumer Affairs 1; Ko Y and Phua J, 'Effects of Eco-Labels and Perceived Influencer Expertise on Perceived Healthfulness, Perceived Product Quality, and Behavioral Intention' (2024) 45 Journal of Current Issues & Research in Advertising 369.

[25] Griffiths M and others, 'Evaluating Source Credibility Effects in Health Labelling Using Vending Machines in a Hospital Setting' (2024) 19 PLOS ONE.

[26] Federal Trade Commission, 'Deception in Weight-Loss Advertising Workshop: Seizing Opportunities and Building Partnerships to Stop Weight-Loss Fraud' (2003) <https://www.ftc.gov/sites/default/files/documents/reports/deception-weight-loss-advertising-workshop-seizing-opportunities-and-building-partnerships-stop/031209weightlossrpt.pdf> accessed November 27, 2024; Sweney M, 'Olay Anti-Ageing Cream Ad Banned' *The Guardian* (March 4, 2009) <https://www.theguardian.com/media/2009/mar/04/olay-ad-banned> accessed November 27, 2024; Dodgson L and Hosie R, 'TikTok Said It Would Be a Haven for Body Positivity. Then It Took $4.3 Million to Push Weight-Loss Products' *Business Insider* (January 30, 2023) <https://www.businessinsider.com/tiktok-sold-ads-weight-loss-products-break-own-rules-2023-1> accessed December 27, 2024.

[27] Berzins LG, 'Protecting the Consumer Through Truth-in-Dieting Laws' (1999) 55 Journal of Social Issues 371.

[28] Pirsch JA, Landreth Grau S and Polonsky MJ, 'Lose 30 Lbs in 30 Days' (2013) 3 Journal of Social Marketing 56.

risk of being harmed; the “dumb consumer,” who is easily influenced and prone to impulsive decisions; and the vulnerable audience, who cannot recognize or protect themselves from persuasive tactics and face significant risks from deceptive marketing. However, even “smart consumers” are not immune to being misled in health-related decisions. The vividness and proximity of promised health rewards in marketing messages can narrow attention and induce impulsive behavior, overriding skepticism. Thus, many consumers, irrespective of their critical thinking abilities, are willing to trust unproven claims. Another study further shows that decisions often prioritize short-term, easily measurable outcomes, such as achieving thinness, over genuine long-term health benefits.[29] Although regulatory efforts to combat false claims have led to fines for manufacturers, these measures have not effectively eliminated misleading practices.[30]

9 A survey conducted by Blagec et al. provides insights into the perspective of manufacturers.[31] It shows that companies working in a business-to-business (B2B) model, serving hospitals and other large organizations, demonstrate a higher willingness to undergo certification. It is less appealing when the prospective buyer is a medical professional and lacks appeal when the buyer is an individual patient. However, this study relied on a convenience sample of just 21 respondents, limiting the generalizability of the results. The author recommends further, more in-depth research to explore the manufacturers’ perspective

10 To sum up, in the B2C market, consumers often prioritize emotional appeal and short-term outcomes over clinical evidence or long-term health benefits. Marketing messages frequently rely on emotionally engaging claims that influence consumer perceptions and behavior, even when such claims are unverified, and high-credibility endorsements may not be effective. At the same time, manufacturers often find the certification pathway unappealing. Therefore, regulatory intervention is suggested to ensure consumer protection.

11 This article examines the legal framework governing health apps within the European Union, with a focus on identifying regulatory gaps that may pose risks to user health and safety. The research scope is limited to industry-specific regulatory frameworks concerning product quality. It excludes data governance matters, as it represents a broad and complex subject that would be more appropriately addressed in a separate, dedicated study. The main target audience is policymakers, who are positioned to address these shortcomings and enhance public protection through regulatory action. Additionally, the findings aim to benefit the general public by raising awareness about the current limitations in their legal protections and encouraging more informed decision-making regarding the quality and reliability of health apps.

12 The article starts with a policy analysis to investigate the legal framework governing medical and wellness apps, highlighting the differences and shortages. The descriptive case study method is used to explore and provide examples of how individual countries can support manufacturers and promote quality assurance. The next section examines the regulatory framework for AI-based healthcare solutions. The following section examines marketplace policies for health apps, which are the final gateways for developers to enter the market. The final section investigates the quality of wellness apps using data from previous scientific studies and real-life examples from the media.

B. The Current Regulatory Status of Health Apps

13 The adoption of the MDR provides a clear definition of the SAMD concept. It defines that a medical device “means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes”.[32] Thus, the MDR specifically mentions that a medical device can be software if it is designed for a medical purpose such as diagnosis, prevention, monitoring, prediction, prognosis, treatment of a disease or injury, investigation or modification of a physiological or pathological process or state, or for the control or support of conception. In general, the MDR establishes specific risk-based requirements for the development and marketing of devices, ensuring product quality and clinical evaluation with the overarching aim of safeguarding patient health.

14 Additionally, the MDR extends its oversight to cover several groups of products without an intended medical purpose, including contact lenses, invasive products intended for cosmetic purposes, high-intensity electromagnetic radiation equipment,

[29] Calder RK and Mussap AJ, ‘Factors Influencing Women’s Choice of Weight-Loss Diet’ (2015) 20 Journal of health psychology 612.

[30] Pirsch JA, Landreth Grau S and Polonsky MJ (n 28).

[31] Blagec K and others, ‘Effects of Medical Device Regulations on the Development of Stand-Alone Medical Software: A Pilot Study’ (2018) 248 Studies in health technology and informatics.

[32] Medical Device Regulation (n 6), art 2(1).

and other products.[33] For the SAMD products, this extension of regulatory scope is irrelevant due to its tangible nature, however, this represents the intention to cover a more expansive array of products, considering their widespread usage and potential impact on human health, even when their purpose is non-medical. This intention suggests that the list of included products could potentially be expanded in the future if deemed necessary, considering that the previous regulation (MDD) did not include such a clause.

15 Compared to the MDD, the MDR requirements are more stringent, posing several challenges that threaten businesses for manufacturers. These include increased expenses, lack of regulatory expertise, constraints on product updates, and other issues. Consequently, this can lead to delays in market entry, withdrawal from the European market in favour of other regions, or even the discontinuation of devices.[34] Therefore, considering the complexity, some manufacturers decide to pursue the business strategy of positioning their products as wellness applications, not for medical purposes. This approach allows them to avoid the lengthy and expensive certification process, even though the actual functionality and use of the app could be regarded as medical. The possibility of this strategy is supported by the European Union Court of Justice ruling on Brain Products.[35] The decision clarified that if a manufacturer hasn't designed a product for medical purposes, the necessity for CE certification does not apply. This approach, however, can pose risks to consumers, as the products have not undergone a review process and can lack clinical evidence.

16 When a manufacturer opts for the wellness pathway, there are no mandatory quality standards or specific requirements to adhere to. The General Product Safety Regulation 2023/988,[36] which aims to ensure consumers' health and safety,[37] stipulates that only safe products may be marketed.[38] A "safe product" is defined as one that "does not present any risk, or only minimal risks compatible with the product's use, considered acceptable and consistent with a high level of protection of consumer health and safety". The term "health" here is interpreted according to the World Health Organization's definition: "a state of complete physical, mental, and social well-being, and not merely the absence of disease or infirmity".[39] Product safety can be demonstrated by assessing the product's characteristics,[40] its compliance with relevant European standards or national requirements,[41] or through other documents addressing product safety.[42] However, since no specific mandatory quality standards or safety metrics exist for digital health apps, determining whether a product meets the definition of a "safe product" is left to the manufacturer's discretion, allowing room for interpretation. This means there are no preventive legal measures to protect consumer health, potentially exposing them to low-quality or harmful products. A study by Singh et al. indicates that only a minority of health-related apps are likely to be useful.[43] This means that the consumers may waste money on a product with no health benefits; in the worst case, the product could harm their health. In such instances, consumer protection mechanisms were established by Product Liability Directive 85/374/EEC[44] and transposed into national legislation, which held manufacturers liable for damage caused by defects in their products. It established that the burden of proof lies with the injured party, who must demonstrate the defect and the causal relationship between the defect and the injury.[45] However, this can be challenging for regular consumers without specific knowledge, leaving many injury cases unaddressed. As per data from the Impact assessment report by the European Commission,[46] 77% of the public indicated moderate to significant difficulties in proving defects in technically complex or AI-based products. While only a limited number of software incorporates AI,

[33] Ibid, art 1(2).

[34] Svempe L (n 18).

[35] Case C-219/11 *Brain Products GmbH v BioSemi VOF and Others* [2012] ECR.

[36] Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC [2023] OJ L 135/1 (General Product Safety Regulation).

[37] Ibid, rec 4.

[38] Ibid, art 5.

[39] Ibid, rec 19.

[40] Ibid, art 6.

[41] Ibid, art 7.

[42] Ibid, art 8.

[43] Singh K and others, 'Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications' (2016) 5 Issue brief (Commonwealth Fund) 1.

[44] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products. [1985] OJ L 210/29 (Product Liability Directive).

[45] Ibid, art 4.

[46] European Commission, 'Impact Assessment Report Accompanying the Document Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products' (2022) <https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:52022SC0316> accessed August 22, 2024.

any digital health app can be considered a technically complex product requiring technical savviness. Therefore, in October 2024, the EU adopted a new directive on liability for defective products which replaces the directive 85/374/EEC.[47] It suggests a less stringent burden of the proof rule if "the claimant faces excessive difficulties, due to technical or scientific complexity".[48] This suggests that it will be easier for consumers to claim compensation in case a defective product has caused harm to their health.

17 While consumers have the option to seek compensation for damages, the author suggests that it should not be the primary approach. The foremost objective should be to protect individuals' health before any harm occurs. Given the absence of specific quality measures for wellness apps, several voluntary codes of conduct have been discussed and established to promote best practices. Yet these codes are often siloed and country-specific, requiring greater policy coordination to ensure that standards are clear, comprehensive, and consistent on an international scale.[49] Furthermore, due to their voluntary nature, developers may disregard these codes.

18 Therefore, policymakers should establish a reasonable regulatory framework for wellness apps to ensure their quality and safety or find ways to support manufacturers in pursuing regulatory compliance. These two options are not mutually exclusive and can be pursued simultaneously to enhance consumer safety and benefit society.

19 Germany was the first country to introduce state support for digital health solutions thus promoting product quality and supporting manufacturers. At the end of 2019, the German parliament (Bundestag) adopted the Digital Healthcare Act (Digitale Versorgung Gesetz, DVG),[50] being a pioneer in introducing a government reimbursement scheme for lower-risk digital healthcare solutions (Class I and IIa). DVG allows an eased pathway for the manufacturers, who cannot yet provide clinical evidence of the positive healthcare effect of their digital health application (Digitale Gesundheitsanwendungen, DiGA), to apply for the provisional listing, allowing them to collect the necessary data in one year (or two years in exceptional cases).[51] The DiGAs have to be certified as medical devices, however, this way DVG promotes the certified pathway as more attractive for the manufacturers, as it opens a market of more than 70 million individuals (88% of the population[52]) using public health insurance.

20 In September 2024, there were 20 applications in the provisional listing and 35 applications in the permanent directory listing,[53] indicating that slightly over one-third of the applications have used the eased option in Germany. It can be considered as an incentive from the government, however, it is in favour of society as it nudges the manufacturers to stay on the regulatory track, focusing on quality and consequently ensuring users' safety, contrary to choosing the non-regulatory pathway of wellness apps. Worth mentioning that there were only 9 applications that have been removed since introducing the DVG (5 apps in 2022, 1 app in 2023, and 3 apps in 2024), suggesting that manufacturers can demonstrate the positive effects of their products. However, the manufacturers have already criticized the reimbursement scheme for its pricing model, low awareness and adoption, and insurers-related roadblocks.[54] This indicates that the processes still need improvement.

21 Germany was later followed by other European countries, introducing reimbursement schemes for digital medical devices. France, Italy, the Netherlands, Poland, Sweden, and the United Kingdom are now also reimbursing the digital solutions, while Belgium is reimbursing the entire clinical pathway which includes a digital health solution.[55]

[47] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC [2024] OJ L, 2024/2853 (New Product Liability Directive).

[48] Ibid, art 10(4).

[49] Ferretti A, Ronchi E and Vayena E, 'From principles to practice: benchmarking government guidance on health apps' (2019) 1(2) Lancet Digit Health e55-e57.

[50] 'Bundestag stimmt Digitale-Versorgung-Gesetz zu' (Deutscher Bundesstag, 2019) <https://www.bundestag.de/dokumente/textarchiv/2019/kw45-de-digitale-versorgung-gesetz-664900> accessed July 29, 2024.

[51] 'The Fast-Track Process for Digital Health Applications (DiGA) According to Section 139e SGB V. A Guide for Manufacturers, Service Providers and Users' (Federal Institute for Drugs and Medical Devices) <https://www.bfarm.de/SharedDocs/Downloads/EN/MedicalDevices/DiGA_Guide.html>.

[52] Blümel M and others, 'Germany: Health System Summary' (The European Observatory on Health Systems and Policies 2022) <https://eurohealthobservatory.who.int/publications/i/germany-health-system-summary-2022>.

[53] Federal Institute for Drugs and Medical Devices, 'DiGA-Verzeichnis' <https://diga.bfarm.de/de> accessed September 18, 2024.

[54] Nicol-Schwarz K, 'DiGA promised German digital health startups access to 73m patients — but slow insurers and poor adoption hold it back' <https://sifted.eu/articles/diga-promised-german-healthtechs-access-to-73m-patients-but-insurer-roadblocks-and-slow-adoption-are-limiting-its-potential> accessed September 16, 2024.

[55] van Kessel R and others, 'Digital Health Reimbursement Strategies of 8 European Countries and Israel: Scoping Review and Policy Mapping' (2023) 11 JMIR mHealth and

22 Another example is the Food and Drug Administration (FDA) in the USA easing compliance rules for mental health apps during the Covid-19 pandemic to address the increased psychological distress in society.[56] It allowed the manufacturers to market their apps without submission of premarket notification, waiving the requirement to submit clinical evidence and compliance with a few other requirements. The incentive allowed various companies to enter the market earlier. For instance, one of them – a Swedish manufacturer Orexo – in 2020 was able to launch three apps in the US market contrary to one planned app without the policy change.[57] Additionally, Mattioli[58] indicates that the relaxed ruling changed product marketing, and wellness apps started claiming more medical benefits. This would not be allowed under previous stricter regulations. While the FDA policy changes were temporary,[59] it provides real-world data for the policymakers. The experienced benefits would potentially allow to improve the existing regulations and incorporate the changes in the standard FDA procedures, while still ensuring safety and effectiveness.[60] Regrettably, so far, the procedures remain unchanged.

C. The Emergence of AI in Healthcare

23 2024 was a landmark year for the Artificial intelligence (AI) regulatory framework. In 2020, 7.2% of mobile health apps incorporated AI,[61] and it would be safe to say that the number of such solutions would only grow, especially with the arrival of generative AI. The use of AI in healthcare presents several challenges, including data-related issues such as privacy, collection, storage, quality, accuracy, and security. Ensuring fairness, preventing various biases and discrimination, and addressing health equity are critical concerns. Additionally, there is a need to ensure transparency, accountability, explainability, and interoperability, and manage potential errors and misdiagnoses.[62] While medical apps are regulated under the MDR to ensure safety, wellness apps currently face much fewer restrictions and their developers may overlook potential risks. According to De Freitas and Cohen,[63] preliminary findings indicate that generative AI can allow consumers to use wellness apps for health-related purposes which may pose health risks, suggesting the need to regulate the technology itself, even if it is not intended for medical purposes.

24 To address these AI challenges, it is essential to establish reasonable regulation and governance that supports innovation, promotes transparency and accountability, and protects society. It is also important to prioritize ethical considerations, as emphasized in the European Commission's Ethics Guidelines for Trustworthy AI.[64] Therefore in March 2024, the pioneering Artificial Intelligence Act[65] (AI Act) in the EU was passed. It employs a risk-based approach, setting requirements for development and transparency, mitigating risks, and prohibiting solutions with unacceptable risk levels.[66] It also applies to all health apps, regardless of their regulatory status as medical devices or wellness apps. Seemingly, as AI-based SAMD are considered at least class IIa under the MDR,[67] they correspond to being classified as high-risk AI systems under the AI Act.[68] Wellness apps at this point would rarely classify as

uHealth e49003.

[56] Office of the Commissioner, 'Coronavirus (COVID-19) Update: Daily Roundup April 15, 2020' (U.S. Food and Drug Administration, 2020) <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-daily-roundup-april-15-2020> accessed July 20, 2024.

[57] Simonite T, 'The Therapist Is In—and It's a Chatbot App' (Wired, June 17, 2020) <https://www.wired.com/story/therapist-in-chatbot-app/> accessed September 16, 2024.

[58] Mattioli M, 'Second Thoughts on FDA's Covid-Era Mental Health App Policy' (2021) 21 Houston Journal of Health Law and Policy 9.

[59] FDA Center for Devices and Radiological Health, 'Transition Plan for Medical Devices That Fall Within Enforcement Policies Issued During the Coronavirus Disease 2019 (COVID-19) Public Health Emergency' <https://www.fda.gov/media/155038/download>.

[60] 'Remarks by Commissioner Stephen Hahn, M.D. — The COVID-19 Pandemic — Finding Solutions, Applying Lessons Learned - 06/01/2020' (U.S. Food and Drug Administration) <https://www.fda.gov/news-events/speeches-fda-officials/remarks-commissioner-stephen-hahn-md-covid-19-pandemic-finding-solutions-applying-lessons-learned> accessed August 16, 2024.

[61] Stewart C, "mHealth Apps Share with Advanced and Standard AI Worldwide 2020" (Statista, October 20, 2020) <https://www.statista.com/statistics/1180814/mhealth-apps-share-incorporating-ai/> accessed August 16, 2024.

[62] Bouderhem R, 'Shaping the future of AI in healthcare through ethics and governance' (2024) 11(1) Humanities and Social Sciences Communications.

[63] De Freitas J and Cohen IG, 'The Health Risks of Generative AI-Based Wellness Apps" (2024) 30(5) Nature Medicine 1269.

[64] European Commission, 'Ethics Guidelines for Trustworthy AI. Shaping Europe's Digital Future.' <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed September 5, 2024.

[65] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [2024] OJ L, 2024/1689 (Artificial Intelligence Act).

[66] Ibid, art 1(2).

[67] Medical Device Regulation (n 6), annex VIII, ch III, 6.3.

[68] Artificial Intelligence Act (n 65), art 6(1).

high-risk AI systems: in case they use biometrics for emotion recognition[69] and if they "pose a significant risk of harm to the health, safety or fundamental rights of natural persons".[70] However, some wellness apps might be classified as low-risk AI systems,[71] for instance, chatbots.

25 Nevertheless, the AI Act is still new, therefore the full impact on the development of digital health solutions yet remains uncertain. Potential issues may arise where the AI Act intersects with the MDR in practical applications.

D. Marketplaces

26 A crucial component in any business is the marketplace, where manufacturers (supply side) meet consumers (demand side). It serves as the final checkpoint where regulatory requirements can be enforced before the product reaches the consumer. This section will explore how marketplaces function as the final gatekeepers to screen out potentially harmful apps.

27 Currently, the predominant platforms for accessing all health mobile applications are the Apple Store (for iOS) and Google Play (for Android). According to the MDR definition, these platforms act as distributors - "any natural or legal person in the supply chain, other than the manufacturer or the importer, that makes a device available on the market, up until the point of putting into service".[72] The regulation establishes the general obligations for distributors, mainly being responsible for verifying that the device conforms with the requirements of the regulations and prohibiting market access to non-conforming devices.[73] These rules apply only to medical devices, not wellness apps.

28 While both platforms implement guidelines to mitigate the risks associated with potentially harmful applications, it is important to note that these platforms do not serve as a screening checkpoint.

29 The Apple Store's review guidelines for developers[74] state that applications behaving in a way that poses physical harm risks to users may face rejection. It is recommended that these applications provide supporting data and methodology to substantiate the beneficial health claims made. Although there is no separate review process for health claims' legitimacy or mandatory data submission, developers can accomplish this by providing references to data sources within the application. Furthermore, the guidelines stipulate that medical applications having received regulatory clearance should submit a link to the corresponding documentation. However, the Apple Store does not assess the necessity of regulatory clearance during the review process.

30 Google Play policy[75] also states that harmful health applications are not allowed in the store. It explicitly declares that the developer is fully responsible for being compliant with the applicable regulations. The guidelines define SAMD and set policies that the developers must comply with. Additionally, their policy states that the manufacturer is obliged to acquire the regulatory clearance, and while it shall not be submitted to Google Play, it should be provided upon request.

31 Both app stores highlight that applications with medical functionality that use only the built-in device features or sensors are not permitted. These would include, for example, apps to measure blood pressure, glucose level, oxygen level, and such. This provision directly constrains the range of functionalities permissible. While the Apple store requires a validated methodology for the products, Google Play requires supporting external products that ensure the provided functionality.

32 In general, although neither platform assesses the necessity for regulatory clearance, their policies are oriented toward user protection, as evidenced by the requirement to provide data substantiating the manufacturer's claims. A facet that is noticeably absent in the MDR. However, the real-life effectiveness of these requirements is questionable, as the app stores rely on the information submitted by the manufacturer without reviewing its quality and completeness. The reliability of such self-declaration by developers is questioned in a study conducted by Huckvale et al.[76] The research examined apps certified by the UK NHS Health Apps Library as clinically safe and trustworthy but found that a significant portion of these apps failed to comply with data protection principles. This finding highlights the shortcomings of accreditation processes that heavily rely on developers' self-declarations, ultimately failing to achieve one of

[69] Ibid, annex III, cl 1.

[70] Ibid, art 6(3).

[71] Ibid, art 51.

[72] Medical Device Regulation (n 6), art 2(34).

[73] Ibid, art 14.

[74] 'App Review Guidelines' (Apple Developer) <https://developer.apple.com/app-store/review/guidelines/#physical-harm> accessed July 3, 2024.

[75] 'Health Content and Services' (Play Console Help) <https://support.google.com/googleplay/android-developer/answer/12261419?hl=en&sjid=13806500483766338070-EU> accessed July 3, 2024.

[76] Huckvale K and others, 'Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment' (2015) 13(214) BMC Medicine.

their primary goals – helping people to find trusted, safe, and secure health apps and serving as a mark of quality. Another example is a study by Tangari et al.[77] which examined the privacy practices of health apps on Google Play and discovered discrepancies between the declared and actual data protection measures, suggesting that Google Play may not be sufficiently safeguarding its users' privacy. These examples illustrate the challenges of enforcing regulations in practice, where stakeholders may intentionally or unintentionally overlook process gaps and minimize their efforts, therefore it underscores the need for increased attention and governance.

E. Apps Causing Harm

33 The overall growing concern is that with the emergence of new technology, bringing new digital health apps, individuals are increasingly placing trust in these apps for their health-related decisions.[78] This section will explore the available evidence to support the theoretical examination of the legal framework in the previous sections to demonstrate that the lack of clinical evaluation, misuse, or product underdevelopment can cause harm to the consumer.

34 According to the IQVIA Institute report,[79] a trend of specialization is noticeable – general health wellness apps lose the majority, while more and more health condition management apps are entering the market, and now mental health, diabetes, and cardiovascular disease-related apps account for almost half of disease-specific apps. In the following section, the mental health and diabetes apps will be explored.

35 Nowadays, when there is rapid technological advancement and development of generative AI, it has become very tempting to quickly develop a product, and various mental health apps arise. Also, the Covid-19 pandemic with limitations in social and professional life contributed to the rise of mental health app downloads.[80]

36 However, the quality of these apps is alarming. As per Sucala et al., the majority of the anxiety relief apps have been developed without involving psychology professionals and there is a lack of data on their efficacy and effectiveness.[81] The lack of evidence is also highlighted in Koh et al. umbrella research,[82] additionally pointing out that some applications do not even provide a therapeutic rationale or evidence behind their interventions. Wang et al.[83] point out that clinical efficacy does not correlate with the popularity (number of downloads) and apps' ratings, which underscores the effect of marketing and search engine optimization processes. The shortcomings of large language models (LLMs) have been explored by Heston,[84] suggesting that LLMs cannot properly detect and address hazardous mental states, consequently not being able to manage the condition safely. De Freitas and Cohen[85] pointed out that wellness apps featuring generative AI, while not intended for mental health purposes, can be used for that purpose thus creating health risks.

37 The lack of regulation and control over the conversations can be even lethal. In 2023 an eco-anxious Belgian man after a six-week-long conversation with an AI chatbot committed suicide to save the planet.[86] Recently, a 14-year-old boy committed suicide after forming a deep emotional attachment to a fictional character during conversations with an AI-powered chatbot.[87] Another case of a harmful application is chatbot Tessa, which The National Eating Disorders Association removed for giving dangerous advice about eating disorders.[88] While initially the service was provided by professionals, soon after replacing them with AI, the problems arose. These are examples where a

[77] Tangari G and others, 'Mobile health and privacy: cross sectional study' (2021) 373 BMJ n1248.

[78] Hogan NM, Kerin MJ, 'Smart phone apps: Smart patients, steer clear' (2012) 89(2) Patient Education and Counseling 360.

[79] IQVIA Institute, 'Digital Health Trends 2021' (n 3), 2.

[80] Wang X, Markert C, Sasangohar F, 'Investigating Popular Mental Health Mobile Application Downloads and Activity During the COVID-19 Pandemic' (2021) 65(1) Human Factors 50.

[81] Sucala M and others, 'Anxiety: There is an app for that. A systematic review of anxiety apps' (2017) 34(6) Depression and anxiety 518.

[82] Koh J, Tng GYQ, Hartanto A, 'Potential and Pitfalls of Mobile Mental Health Apps in Traditional Treatment: An Umbrella Review' (2022) 12(9) Journal of Personalized Medicine 1376.

[83] Wang X, Markert C, Sasangohar F (n 80).

[84] Heston TF, 'Safety of Large Language Models in Addressing Depression' (2023) 15(12) Cureus.

[85] De Freitas J and Cohen IG, 'The Health Risks of Generative AI-Based Wellness Apps'' (2024) 30(5) Nature Medicine 1269.

[86] Laura W, 'Belgian man dies by suicide following exchanges with chatbot' *The Brussels Times* (March 28, 2023) <https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt> accessed August 12, 2024.

[87] Roose K, 'Can a Chatbot Named Daenerys Targaryen Be Blamed for a Teen's Suicide?' *The New York Times* (October 23, 2024) <https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html> accessed December 7, 2024.

[88] 'NEDA Suspends AI Chatbot for Giving Harmful Eating Disorder Advice' (Psychiatrist.com, June 5, 2023) <https://www.psychiatrist.com/news/neda-suspends-ai-chatbot-for-giving-harmful-eating-disorder-advice/> accessed August 12, 2024.

lack of oversight and control over technology and generative AI poses imminent risks to consumers' health.

38 A similar situation is evident with the other large segment – diabetes apps. Research shows that the overall quality of apps is moderate, and most of the self-management apps lack rationale – only 8% of apps had any evidence behind their program.[89] There is a low number of randomized controlled trials on diabetes apps, a small number of proven long-term benefits, and even limited high-quality short-term data.[90]

39 All these factors affect the trust and credibility of the technology and jeopardize the health app market development in the long term. Losing consumers' trust will decrease the adoption of the new technology in general, also of the apps that are clinically validated and actually do provide positive healthcare effects. And having low-quality apps in the market poses additional risks to the well-being of the individuals who already seek help.

40 From the legal perspective, the wellness applications do have a disclaimer in their terms of service and the app that it does not provide medical advice, and in case of any health concerns, the consumers shall consult with healthcare professionals. However, research[91] has demonstrated the common tendency of the average consumer to overlook the details in the fine print.

F. Conclusion

41 The current regulatory approach, which focuses only on the official intended use of the application, poses evident risks, as the oversight is not extended to wellness apps even though their functionality may resemble medical purposes. This is especially evident in the case of generative AI-based solutions. Hence, to ensure consumer safety, it is important that these apps have also undergone safety, quality, and efficacy evaluation, and are continuously monitored during their operation time as required by the MDR in the post-market surveillance process for medical apps. One potential solution is to expand Annex XVII of the MDR, which lists products without an intended medical purpose to which the regulation applies, to include wellness apps. However, the author believes that this extension of scope would be unreasonable because wellness apps typically pose minimal, if any, health risks and the current complex regulatory framework could potentially have a detrimental impact on the digital wellness market.

42 Thus, a specific regulation for wellness apps is suggested. Considering the arguments mentioned in this article, it is important to find a balance and introduce a fair regulatory framework. Policymakers need to make sure that they do not overregulate the sector, making it difficult for manufacturers to meet the requirements and unappealing to work in the market at all. However, reasonable and feasible requirements should be implemented to ensure the software is science-backed and safe. The author proposes adding a new clause to Article 6 of the General Product Safety Regulation, specifically addressing products intended for health-related use. The clause would reference an annex detailing the requirements necessary to ensure the quality of such products. For instance, it would mandate the involvement of relevant experts in product development: a mental health app should include input from professionals such as psychologists or psychiatrists, while a diet app should involve a qualified nutritionist. For generative AI solutions, regulatory requirements could consist of built-in limitations on the scope of advice provided. Additionally, recognizing that disclaimers and fine print are rarely read, it should be mandatory for users to be referred to health professionals during their interaction with the app, particularly when it resembles medical use or when the user is in a potentially harmful situation.

43 Until a proper European approach is established, the EU member states may implement such safety and quality requirements at the national level. However, this could lead to a fragmented internal market and is therefore recommended only as a temporary measure until unified European requirements are adopted.

44 Given the high health risks associated with AI in unregulated wellness apps, the author further proposes expanding the list of high-risk systems in Annex III of the AI Act to include products that would fall under the proposed new clause in Article 6 of the General Product Safety Regulation.

45 There is also unused potential in the collaboration between app marketplaces and regulatory bodies. While marketplaces currently provide guidelines for health apps to be accepted in the stores, the

[89] Geirhos A and others, 'Standardized evaluation of the quality and persuasiveness of mobile health applications for diabetes management' (2022) 12(1) Scientific Reports.

[90] Fleming GA and others, 'Diabetes Digital App Technology: Benefits, Challenges, and Recommendations. A Consensus Report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group' (2019) 43(1) Diabetes Care 250.

[91] Bakos Y, Marotta-Wurgler F, Trossen DR, 'Does Anyone Read the Fine Print? Consumer Attention to Standard-Form Contracts' (2014) 43(1) The Journal of Legal Studies 1.

responsibility of quality, compliance, and evidence behind claims is assigned primarily to the developers. Moreover, the app stores are minimally engaged in quality monitoring during apps' operation time. However, enhancing their role as gateways and assigning greater responsibility to them would be advantageous for society, mitigating health risks. For instance, if the additional clause to the General Product Safety Regulation mentioned above is adopted, gatekeepers should be required to actively verify whether the technical requirements have been implemented in the app, rather than relying solely on developers' self-declarations.

46 Additionally, attention should also be directed towards encouraging manufacturers to opt for the certification pathway, as it ensures the validation of apps for safety and quality. The current complex regulatory framework is unattractive to developers of lower-risk products. Hence, it is recommended to introduce incentives aimed at supporting the manufacturers. While the implementation of such programs at the EU level may require considerable time and effort, it is advised for individual countries to introduce incentives at the national level to support their med-tech companies. For example, national governments could establish reimbursement schemes for digital healthcare solutions. This is particularly crucial for small and medium-sized enterprises with limited financial resources.

It will be what we want it to be: Sociotechnical and Contested Systemic Risk at the Core of the EU's Regulation of Platforms' AI Systems

by **Mateus Correia de Carvalho** *

Abstract: The EU regulates AI systems of large digital platforms using a risk-based approach developed primarily through the Digital Services Act (DSA) and the AI Act (AIA). The existing literature highlights two main challenges to this regulatory strategy: the potentially unconstrained discretion and informational power of regulated tech companies, and the limited predictive value of risk regulation for less quantifiable forms of harm. This paper describes and systematises how EU law intends to address these challenges and ensure effective AI risk management processes. Through doctrinal analysis of the DSA, AIA, and their implementing laws and soft law, it lays out the integrated risk management framework these regulations establish for platforms' AI systems. It argues that this integrated framework has three main normative commitments: (i) AI systemic risks should be framed sociotechnically, (ii) their management should be methodologically contextual, and (iii) and civil society should be actively involved in identifying and mitigating AI systemic risks. On this last commitment, however, the mechanisms for civil society participation remain especially unclear. This paper thus offers an overview of all formal and informal spaces of participation in this risk management framework, differentiating them by their institutional setup, rationales for civil society intervention, types of expertise sought, and actors involved. Overall, this paper advances the dialogue on the EU's risk-based approach to platform and AI regulation, offering a possible baseline for critique and empirical inquiry into its implementation.

Keywords: AI Systems; Digital Platforms; Systemic Risk; Participation; EU Law

Recommended citation: Mateus Correia de Carvalho, It will be what we want it to be: sociotechnical and contested systemic risk at the core of the EU's regulation of platforms' AI systems, 16 (2025) JIPITEC 35 para 1.

A. Introduction

1 The emergence and increasing integration of AI-driven recommender systems[1] and generative AI[2] on digital platforms create risks of harm to persons' fundamental rights, health, and safety.[3]

[*] Doctoral Researcher, European University Institute (EUI), Fiesole, Italy. I would like to sincerely thank Deirdre Curtin, Rachel Griffin, Marta Maroni, Estela Lopes, Pankhudi Khandelwal, Renan Bodin, Alex Schuster, Johannes Müller, Marco Almada, Julia Galera Oliva, Mary Greenshields, and Timo Zandstra for commenting on earlier versions. This paper was presented at the Young Digital Law Conference 2024, hosted by Sciences Po (Paris) where it benefitted from the precious feedback of Ilaria Buri, Paddy Leerssen, Jenny Orlando-Salling and Louise Bartolo. I also greatly appreciate the peer review comments which improved the paper greatly. All remaining mistakes are mine.

[1] Since this article mainly looks at the EU's Digital Services Act (DSA), it defines AI recommender systems per its Article 3(s) as fully or partially algorithmically driven systems "used by an online platform to suggest" and/or prioritise specific information "in its online interface".

[2] Defined as "advanced machine learning models that are trained to generate new data, such as text, images, or audio", which makes them "distinct from other AI models, only designed to make predictions or classifications or to fulfil other specific functions" in Philipp Hacker, Andreas Engel and Marco Mauer, 'Regulating ChatGPT and Other Large Generative AI Models', *2023 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2023) 1113 <https://dl.acm.org/doi/10.1145/3593013.3594067> accessed 20 January 2024.

[3] Recitals 81 and 83 DSA and 15-16 AI Act (AIA); Kate Crawford, 'Can an Algorithm Be Agonistic? Ten Scenes from Life in Calculated Publics' (2016) 41 Science, Technology, & Human Values 77, 83–85; Brent Daniel Mittelstadt and others, 'The

Platforms' AI systems[4] also pose broader societal risks to democracy and civic discourse, as they have the potential to manipulate individuals' perception of reality,[5] mediate a significant part of their social interactions,[6] and, therefore, shape how they relate to one another in society.[7] Specifically, they may contribute to increasing polarization of public opinion,[8] and affect the integrity of electoral processes,[9] interfere with people's free access to and exchange of information,[10] and perpetuate long-standing patterns of discrimination and marginalisation of certain individuals and communities.[11]

2 But these are, in the end, just risks. What are, and will be, the specific negative impacts of digital platforms' AI systems on individuals and societies? Even if we may have some idea, no one can claim to know for sure the answer to this question. Indeed, AI's technical complexity and opacity,[12] coupled with its rapid development and varied integration in digital platforms,[13] make it very hard for regulators to gauge the harms it might cause and adopt suitable strategies to address them.[14]

3 In order to cope with these uncertainties and dynamically regulate AI systems, the EU has adopted a risk-based approach.[15] Specifically, it applies

Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society 9–10.; Brent Daniel Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society 9–10.

[4] Any reference to the 'AI systems' of digital platforms made henceforth should be understood, unless a more specific distinction is made, as comprising the two different types of algorithmic systems mentioned in footnotes 1 and 2: (i) algorithmic recommender systems; and (ii) generative AI models (hereinafter 'genAI').

[5] Recitals 67 DSA and 16 AIA; Rostam J Neuwirth, *The EU Artificial Intelligence Act: Regulating Subliminal AI Systems* (Routledge 2022).

[6] Jennifer Cobbe, 'Algorithmic Censorship by Social Platforms: Power and Resistance' (2021) 34 Philosophy & Technology 739, 739–743.

[7] Recital 79, DSA; Daniel Yudkin, Stephen Hawkins and Tim Dixon, 'The Perception Gap: How False Impressions Are Pulling Americans Apart' [2019] More in Common 6, 49, 51.

[8] Smitha Milli and others, 'Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media' (arXiv, December 2023) 6–7 <http://arxiv.org/abs/2305.16941> accessed 23 September 2024. Polarization, like many other effects of platforms' AI systems is a product of the entanglement between the latter, platforms interfaces, associated devices and technical infrastructure, individuals, and other social systems. See, to this effect, Sinan Aral, *The Hype Machine: How Social Media Disrupts Our Elections, Our Economy, and Our Health–and How We Must Adapt* (Crown Currency 2021) 3, 56–93; Cass R Sunstein, *#Republic: Divided Democracy in the Age of Social Media* (Princeton University Press 2018) 59–97.

[9] European Commission, Directorate-General for Communications Networks, 'Digital Services Act: Application of the Risk Management Framework to Russian Disinformation Campaigns' (Publications Office of the European Union, 2023) 59–63; 'Consultation on Guidelines for Providers of Very Large Online Platforms and Very Large Online Search Engines on the Mitigation of Systemic Risks for Electoral Processes' (European Commission, 2024) paras. 1, 3, 25, and 26, including cited sources.

[10] Recital 82 DSA; Rishi Bommasani and others, 'On the Opportunities and Risks of Foundation Models' (arXiv, 2022) 137 <http://arxiv.org/abs/2108.07258> accessed 13 December 2023; Paul Bouchaud and others, 'The Amazing Library: An Analysis of Amazon's Bookstore Algorithms within the DSA Framework' (AI Forensics; Check First 2023) 38 <https://checkfirst.network/wp-content/uploads/2023/12/AIF%20x%20CF%20-%20The%20Amazing%20Library_final.pdf> accessed 23 September 2024.

[11] Beatriz Botero Arcila and Rachel Griffin, 'Social Media Platforms and Challenges for Democracy, Rule of Law and Fundamental Rights' (Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, PE 2023) 10; Benjamin Laufer and Helen Nissenbaum, 'Algorithmic Displacement of Social Trust' (Knight First Amendment Institute 2023) 5 <https://s3.amazonaws.com/kfai-documents/documents/a29f3e5731/1.23.24-SocialTrust-Draft.pdf> accessed 16 February 2024.

[12] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big data & society.

[13] Stefan Larsson, Jockum Hildén and Kasia Söderlund, 'Between Regulatory Fixity and Flexibility in the EU AI Act' 3–5 <https://portal.research.lu.se/en/publications/between-regulatory-fixity-and-flexibility-in-the-eu-ai-act> accessed 15 March 2024; Paddy Leerssen, 'Embedded GenAI on Social Media: Platform Law Meets AI Law' (*DSA Observatory*, 16 October 2024) <https://dsa-observatory.eu/2024/10/16/1864/> accessed 22 October 2024; Mathias Vermeulen and Laureline Lemoine, 'From ChatGPT to Google's Gemini: When Would Generative AI Products Fall within the Scope of the Digital Services Act?' (*Media@LSE*, 12 February 2024) <https://blogs.lse.ac.uk/medialse/2024/02/12/from-chatgpt-to-googles-gemini-when-would-generative-ai-products-fall-within-the-scope-of-the-digital-services-act/> accessed 20 February 2024.

[14] Larsson, Hildén and Söderlund (n 13).

[15] Giovanni De Gregorio and Pietro Dunn, 'The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age' (2022) 59 Common Market Law Review 473, 476; Margot E Kaminski, 'The Developing Law of AI: A Turn to Risk Regulation' (2023) 3 <https://papers.ssrn.com/abstract=4692562> accessed 31 January 2024. For a discussion of other reasons for the adoption of risk-based regulation in digital governance matters, not often stated in policy documents and official communications, see, for example, Rachel Griffin, 'What Do We Talk about When We Talk about Risk? Risk Politics in the EU's Digital Services

such an approach to the regulation of AI systems of very large online platforms and search engines (hereinafter referred to as 'platforms', 'digital platforms' or 'VLOP/SEs'[16]) through the recent Digital Services Act (DSA)[17] and AI Act (AIA).[18]This approach frames AI's potential negative impacts as future risks of harm. It also mandates that private entities responsible for AI systems related to platforms establish processes for the iterative management of these risks.[19] The setting up and implementation of those risk management processes are then overseen by public supervisory authorities.[20]

4 The literature has pointed out that, like all risk regulation, the risk-based approach to AI regulation adopted in the DSA and AIA will face two main challenges. The first is conceptual: risk is often conceived in an actuarial and individual fashion, i.e., it focuses on quantitatively identifying and assessing risks of harm caused to specific individuals or entities.[21] These dominant conceptions of risk, while easier to calculate, fail to fully capture less quantifiable and intangible AI risks – e.g., to democracy, fundamental rights, civic discourse, or of gender-based violence – whose perceptions are contestable and highly subjective but that both the DSA and AIA aim to address.[22] The second challenge is institutional: risk regulation affords significant discretion to private regulated actors to set up risk management processes and strategies,[23] which might lead to ineffective and insufficient risk assessment and mitigation.[24]

5 Against this background, this paper aims to address how the DSA and AIA envision the creation of an effective risk regulatory regime applicable to the AI systems of digital platforms. Answering this question is, first and foremost, a descriptive exercise based on the legal doctrinal method. It requires reviewing the applicable legal sources and systematically describing the AI risk management schemes they institute.[25] In this case, it is important not only to describe the AI risk management provisions of the DSA and AIA, but also the legal acts and soft law instruments that concretise them. These are:

- the Commission Delegated Regulation (EU) 2024/436 of 20 October 2023 laying down rules on the performance of audits for very large online platforms and very large online search engines (hereinafter, the 'Delegated Regulation on Audits', or 'DRA');
- the Commission Implementing Regulation (EU) 2023/1201 of 21 June 2023 on detailed arrangements for the conduct of certain proceedings by the Commission pursuant to the DSA (hereinafter, the 'Implementing Regulation 2023/1021');

Act' (*Digital Services Act Observatory*, 31 July 2024) <https://dsa-observatory.eu/2024/07/31/what-do-we-talk-about-when-we-talk-about-risk-risk-politics-in-the-eus-digital-services-act/> accessed 2 September 2024.

[16] In this paper, I rely on the definition of 'digital platforms' used in the DSA's risk management provisions. Therefore, in accordance with art. 33(1) DSA, whenever this paper mentions 'digital platforms', 'platforms', or 'VLOP/SEs', these terms should be understood as referring to very large "online platforms and online search engines which have a number of average monthly active recipients (...) in the Union equal to or higher than 45 million".

[17] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277 2022.

[18] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), PE/24/2024/REV/1, OJ L, 2024/1689 2024.

[19] Gregorio and Dunn (n 15) 476; Daniela Stockmann, 'How Will the European Union Govern Social Media Platforms under the Digital Services Act?' (*Hertie School Centre for Digital Governance*, 16 June 2023) <https://www.hertie-school.org/en/digital-governance/research/blog/detail/content/how-will-the-european-union-govern-social-media-platforms-under-the-digital-services-act> accessed 26 December 2023; Margot E Kaminski, 'Regulating the Risks of AI' [2023] Boston University Law Review 1347.

[20] Fiona Haines, 'Regulation and Risk' in Peter Drahos (ed), *Regulatory theory: Foundations and applications* (Australian National University Press Acton, ACT, Australia 2017) 188–192; Martin Husovec, 'The Digital Service Act's Red Line: What the Commission Can and Cannot Do About Disinformation' (2024) 1, 7 <https://papers.ssrn.com/abstract=4689926> accessed 16 January 2024.

[21] Kaminski (n 19) 1390–1391; Kaminski (n 15) 14–16.

[22] See, e.g., recitals 44d AIA and 75 DSA. See also Kaminski (n 19) 1392–1393; Marco Almada and Nicolas Petit, 'The EU AI Act: A Medley of Product Safety and Fundamental Rights?' (2023) SSRN Paper 18–19; European Commission, 2023 (n 9) 11, 15-18.

[23] Julia Black and Andrew Douglas Murray, 'Regulating AI and Machine Learning: Setting the Regulatory Agenda' (2019) 10 European journal of law and technology 4–7; Kaminski (n 19) 1379; Gregorio and Dunn (n 15) 483–488.

[24] Kaminski (n 19) 1379-1380; Niklas Eder, 'Making Systemic Risk Assessments Work: How the DSA Creates a Virtuous Loop to Address the Societal Harms of Content Moderation' (2023) 13 <https://papers.ssrn.com/abstract=4491365> accessed 31 October 2023.

[25] Jan M Smits, 'What Is Legal Doctrine? On the Aims and Methods of Legal-Dogmatic Research' in Rob van Gestel, Hans-Wolfgang Micklitz and Edward L Rubin (eds), *Rethinking Legal Scholarship: A Transatlantic Dialogue* (Cambridge University Press 2017) 207, 210.

- the Commission Decision of 24 January 2024 establishing the European AI Office, C(2024) 390 final (hereinafter, the 'AI Office Decision');

- the Commission Draft Delegated Regulation laying down the technical conditions and procedures under which providers of very large online platforms and of very large online search engines are to share data pursuant to Article 40 of Regulation (EU) 2022/2065 (hereinafter, the Access to Data Delegated Regulation);

- the Commission DSA draft guidelines for platforms on mitigating risks for electoral processes (hereinafter, the 'DSA risk mitigation guidelines');[26]

- the 2023 Commission study applying the DSA's risk management framework to Russian disinformation campaigns (hereinafter, the 'DSA Russian disinformation study');[27] and

- the 2022 Strengthened Code of Practice on Disinformation (hereinafter, 'Disinformation Code of Practice').[28]

6 Because it conceives of law as a system, the doctrinal method is adequate to both (i) provide coherence to the many different provisions applicable to a given regulated matter and (ii) extract from those legal texts their normative meaning as ascribed to them by the legislator. In this paper, I thus use the doctrinal method to structure the DSA and AIA's risk management frameworks into a coherent system, all the while trying to understand the broader internal value-based logic that underpins it. This means that, besides simply describing their legal norms and competent institutions, I will also provide an interpretative analysis of the two regulations' own normative commitments and aspirations regarding how their risk management schemes *should* be enforced. Structuring the EU law regime of platform and AI risk management in this way will enable its future intra and extra-legal critique . For one, clearly stating the normative commitments and aspirations of the DSA and AIA's risk management regimes will allow, in time, for a critique of their implementation on the regulations' own terms.[29] In addition, highlighting those normative ambitions can also enable their own critique from extra-legal viewpoints that uncover and scrutinise the interests they serve, produce and help reinforce (and at the expense of whom they do so).[30]

7 Section B. argues that the DSA and AIA were conceived as instituting two different but complementary AI risk management schemes. After clarifying the relationship between the two regulations, I will separately describe the two risk management regimes they establish for platforms' AI systems.

8 Then, in Section C., I distil the commonalities between the two regulations' AI risk regimes that ultimately unify them into, I argue, one integrated EU AI risk governance framework applied to digital platforms. To do so, I combine the legal analysis of the DSA and AIA with insights taken from their *travaux préparatoires*, related policy documents and relevant literature. Particularly, I highlight three overarching normative commonalities of the two regulations' AI risk management schemes: they frame (at least some) AI risks as 'systemic' (C.I.); require that the assessment and mitigation of those risks be socially contextualised (C.II.) and expect civil society actors to be involved in those risk management processes (C.III.) Admittedly, the choice to focus on these three normative objectives of the DSA and AIA has, itself, a certain underlying normativity: it implies that the focus of the analysis of these risk management regimes is put not on their market regulation objectives but instead on their non-market, protective aims.[31] Simply put, I identify the DSA and AIA's three main normative commitments regarding how risk management procedures should be shaped in order to protect such values as fundamental rights, democratic processes, and public health and safety.

9 Of the three highlighted normative ambitions of the DSA and AIA's risk management frameworks, one has a particularly unclear path towards operationalisation: civil society involvement. Notably, the concrete procedures for civil society

[26] European Commission, 2024 (n 9).

[27] European Commission, 2023 (n 9).

[28] 'European Commission, The 2022 Strengthened Code of Practice on Disinformation', <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>.

[29] Martijn Willem Hesselink, 'Knowing EU Law : How Epistemic and Ontological Commitments Shape Different Understandings of European Law and Why It Matters' (European University Institute 2024) Working Paper 15 <https://cadmus.eui.eu/handle/1814/76827> accessed 3 May 2024.

[30] Ioannis Kampourakis, 'Bound by the Economic Constitution: Notes for "Law and Political Economy" in Europe' (2021) 1 Journal of Law and Political Economy 301; Hesselink (n 29) 15–19.

[31] As De Gregorio and Dunn (n 15) put it, these EU risk-based regulations seek to find a balance between market objectives, such as technological innovation, and non-market protection, such as fundamental rights protection. For a broader distinction between EU secondary law's market and non-market aims, see Bruno de Witte, 'Non-Market Values in Internal Market Legislation' in Niamh Nic Shuibhne (ed), *Regulating the Internal Market* (Edward Elgar Publishing 2006).

participation in this risk management framework are not defined in law . References to different civil society actors are scattered through the provisions of the two regulations, which mention different rationales and forms of civil society interventions in AI risk management. In addition, legal mobilisation literature points to the fact that civil society might intervene informally - and not just through formal avenues of public participation - in the implementation of EU legislation.[32] However, it is not clear what informal avenues of civil society involvement could be used to influence the implementation of the DSA and AIA's risk governance regimes. Although a full answer to these questions necessarily requires an empirical analysis, this paper takes a necessary first step. Section D. maps (i) what are, in the abstract, the formal and informal avenues of civil society participation in EU's AI risk governance, (ii) which civil society actors are empowered to participate therein, (iii) under what type of institutional setting, and (iv) with what aims. Section E. offers concluding remarks.

B. The DSA and AIA as an integrated risk management framework of platforms' AI systems

10 The development of the EU's Digital Strategy has led to the adoption of numerous new instruments of secondary law updating or adding to the existing EU law *acquis* governing digital governance matters.[33] When it comes to the regulation of the AI models and systems integrated into or whose output is diffused through digital platforms, two regulations are primarily relevant: the DSA and the AIA.[34]

11 In this section, I conceive the DSA and the AIA as instituting an integrated EU risk management framework applicable to platforms' AI systems. Before separately describing the relevant provisions of the DSA (B.I.) and AIA (B.II.), it is important to systematise how they relate to one another. Indeed, the two regulations' AI risk-based regimes apply to digital platforms' AI systems in different but complementary ways. This complementarity is highlighted by the AIA's *travaux préparatoires* and the DSA risk mitigation guidelines, which stress the need to ensure a consistent implementation of the two regulations.[35] But what does that exactly mean?

12 The AIA answers that question by stating that, to the extent that AI systems and models are embedded into VLOP/SEs, the latter should manage the systemic risks of those systems and models through the DSA's framework. Compliance with this framework means that corresponding systemic risk management obligations of the AIA "should be presumed to be fulfilled". The AIA's systemic risk management regime will nonetheless come into play if "significant systemic risks" not covered in the DSA are identified in platforms' AI systems and models.[36]

13 Two conclusions can be inferred from the foregoing. First, the DSA is the primary instrument that governs the risks posed by the AI systems of digital platforms. This means, in essence, that platforms must assess and mitigate emerging AI systemic risks at least once a year and in any event prior to launching any new AI-driven or AI-related feature or functionality of their services (art. 34-35 DSA). Second, the AIA functions as a residual regime for new emerging systemic risks that do not fit the DSA's mould.[37] In view of this communication between the DSA and AIA, some scholars have proposed that systemic risk analyses under the two regulations draw inspiration from each other or, even further, be done in integration, i.e., in one analysis considering platform-specific risks that the DSA focuses on, AI-specific risks addressed by the AIA, and also those risks produced by the entanglement between AI and digital platforms' architecture.[38]

[32] Elise Muir, Mark Dawson and Monica Claes, 'A Tool-Box for Legal and Political Mobilisation in European Equality Law' in Dia Anagnostou (ed), *Rights and Courts in Pursuit of Social Change: Legal Mobilisation in the Multi-Level European System* (2014); Lisa Conant and others, 'Mobilizing European Law' (2018) 25 Journal of European Public Policy 1376.

[33] European Commission, 'Communication from the European Commission: Report on the State of the Digital Decade 2024' (2024) 7–8.

[34] This is not to the exclusion of other previously adopted and still relevant EU digital regulations such as the General Data Protection Regulation. In this sense, see, for example, Margot E Kaminski and Gianclaudio Malgieri, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' (2021) 11 International Data Privacy Law.

[35] European Commission, 2024 (n 9), para. 58; European Commission, AI Act proposal, COM (2021) 206 final, 2021/0106 (COD), 5.

[36] Recital 118 AIA.

[37] For a similar argument and a broader analysis of the intersections between the DSA and AIA, see Leerssen (n 13).

[38] Natali Helberger and Nicholas Diakopoulos, 'ChatGPT and the AI Act' (2023) 12 Internet Policy Review 4; Philipp Hacker, 'The AI Act between Digital and Sectoral Regulations' (Bertelsmann Stiftung 2024) 17–19 <https://www.bertelsmann-stiftung.de/doi/10.11586/2024188> accessed 28 January 2025.

I. The core: the Digital Services Act and systemic risk management

1. Risk assessment

14 As outlined above, the DSA is the main instrument that platforms should take into account when managing the risks of their AI systems in accordance with EU law. The first step that VLOP/SEs need to take in this respect is to engage in risk assessment (art. 34 DSA). In essence, they must identify and assess the impact of any systemic risks in the Union stemming from the design, functioning, or use of their services and, amongst others, related algorithmic systems.[39]

15 For a certain risk to be identified and assessed in the DSA's risk management framework, that risk must be qualified as 'systemic'. Crucially, the DSA contains no clear definition of systemic risk; that much is a consensus of the early literature and research work on the regulation's risk management scheme.[40] I posit, however, that the DSA still gives us several helpful hints to flesh out this concept.

16 To begin with, art. 34(2) DSA lists possible sources of systemic risks and that list includes the design of platforms' recommender systems and 'any other relevant algorithmic system' integrated into or used within platforms' services.[41] In any case, recitals 79 and 84 show that platforms should consider not only the design of their algorithms, but also the latter's functioning and use, and especially so where they lead to the amplification of harmful information. In addition, platforms should be mindful of the 'inauthentic use of their service', namely through the generation and dissemination of synthetic content that is either illegal or may contribute to disinformation campaigns.[42] Such synthetic content nowadays is increasingly produced by genAI[43] and that should also be considered in platforms' risk assessments.

17 Furthermore, art. 34(1) DSA helps determine what negative effects may result from AI systemic risks. There, the EU legislator lists those possible negative effects of AI systemic risks that should always be part of VLOP/SEs' risk assessments. Platforms can, in their risk assessments, uncover other systemic risks but they must, in any case, consider all actual or foreseeable:

- dissemination of illegal content (art. 34(1)(a));
- fundamental rights violations (art. 34(1)(b));
- negative effects on civic discourse, electoral processes, and public security (art. 34(1)(c)); and
- gender-based violence, negative effects on minors' public health, as well as serious negative consequences on any person's physical or mental well-being (art. 34(1)(d)).

18 Despite the above, the DSA still does not answer the questions of (i) what the threshold for a risk to be considered systemic is; and (ii) how a systemic risk could be deemed to exist in concrete cases.[44] Such crucial questions have not, to date, been settled; nor was it the purpose of the DSA to answer them right away, as is implied by its risk-based approach. Indeed, as is common with risk regulation, the DSA does not purport to provide a substantive definition of AI systemic risks, but, differently, institutionalises risk assessment procedures whose output will be the iterative definition of that concept.[45]

19 Such an iterative and process-based definition of systemic risk should be framed by some guiding principles . First, the DSA prescribes that VLOP/SEs take into consideration the severity and probability of the identified risks in their respective risk assessments.[46] This emphasis on the combined effects of the potential negative impacts of a risk (severity) and the likelihood that those negative impacts materialise (probability) is a characteristic of so-called actuarial risk frameworks. These frameworks define risk as the product of quantifiable variables that are measured through a scientific

[39] Such risk assessments should be continuous – done at least once a year per art. 34(1) DSA - and iterative, i.e., they should build upon each other and show the evolution of previously identified systemic risks (recital 85 DSA).

[40] See, e.g., Anna-Katharina Meßmer and Martin Degeling, 'Auditing Recommender Systems Putting the DSA into Practice with a Risk-Scenario-Based Approach' (Stiftung Neue Verantwortung (SNV) 2023) 14 <https://shorturl.at/viWyd> accessed 17 January 2024; Jason Pielemeier and David Sullivan, 'Unpacking "Systemic Risk" Under the EU's Digital Service Act' (*Tech Policy Press*, 19 July 2023) <https://techpolicy.press/unpacking-systemic-risk-under-the-eus-digital-service-act> accessed 16 May 2024; Oliver Marsh, 'Researching Systemic Risks under the Digital Services Act' [2024] Algorithm Watch 5–7 <https://algorithmwatch.org/en/wp-content/uploads/2024/08/AlgorithmWatch-Researching-Systemic-Risks-under-the-DSA-240726.pdf> accessed 26 August 2024.

[41] Art. 34(2)(a) DSA.

[42] Recital 84 DSA; European Commission, 2024 (n 9) para. 25.

[43] European Commission, 2024 (n 9) para. 25.

[44] European Commission, 2023 (n 9) 15.

[45] Kaminski (n 19) 1402; Stockmann (n 19); Husovec (n 20) 7; Griffin, 'What Do We Talk about When We Talk about Risk?' (n 15).

[46] Recital 79 and art. 34(1) DSA.

or technical frame (usually cost-benefit analyses and/or mathematical assessments multiplying the intensity of the effects of a given harm by its likelihood).[47] Many scholars have pointed out the limited predictive value of actuarial risk frameworks, noting that they fail to fully capture less quantifiable and more socially-dependent risks, instead reducing them to mere technical and mathematical variables.[48]

20 The DSA's risk conception is not, however, purely actuarial. On the contrary, one can find in its articles, recitals, and implementing guidelines several references to the need to contextualise risk assessments by taking into account social and cultural factors that influence the risk perceptions of affected individuals and communities.[49] Although not conclusive, this emphasis on socially and culturally dependent risk assessments is useful to begin gauging the meaning of the 'systemic' in 'systemic risk'. Primarily, it makes clear that the assessment and consequent definition of systemic risk must necessarily extend beyond exclusively quantitative calculations. It should be based on contextual methodologies that locate assessments of risk in their specific social and cultural context.[50] Particularly, platforms must take into account regional and linguistic factors that might affect perceptions and, therefore, assessments of risk,[51] as well as the specific legal, societal and political contexts where systemic risks manifest themselves.[52]

21 In addition, many references to the effects of 'systemic risks' in the DSA suggest that this concept requires a framing that goes beyond identifying isolated instances of harm caused by AI systems. In particular, the DSA's qualification of risk as 'systemic' suggests a reference to the propagation at scale of the negative effects potentially caused by AI systems and digital platforms. That would be only natural since the negative effects of platforms' AI systems are inherently disseminated through the online audiences of large digital platforms.[53] Therefore, the DSA refers to the systemic risks of platforms (including those stemming from their AI systems) by emphasizing collective (as opposed to individual) forms of harm: it mentions, e.g., "societal concerns" and "societal and economic harm", such as risks to the "shaping of public opinion and discourse" through "coordinated disinformation campaigns", as well as of negative effects for "democratic processes", the (non-individualised) "exercise of fundamental rights", and online safety and trade.[54]

22 In the DSA Russian disinformation study, the Commission made the only official attempt to date to densify the concept of systemic risk. It did so through one of the variables of any actuarial risk framework: severity; but gave a distinct sociocultural flavour to that concept. In the Commission's words, for a risk to be systemic its actual or foreseeable negative effects must be 'severe enough'. And a systemic level of severity should be measured as a function of both qualitative and quantitative indicators. Specifically:

> "[s]everity is a function of the relationship between the qualitative assessment of the risk posed by the content in context and a quantitative measure of the reach and/or intensity of exposure of audiences to that content. It follows then that a risk may reach a systemic level in different ways. The higher the level of risk inherent in the content *in context*, the smaller the audience required to reach a systemic level. And by contrast, the lower the level of risk inherent in the content *in context*, the larger the audience required to reach a systemic level."[55]

23 This approach of the Commission to defining 'systemic risk' is not, by the Commission's own admission, set in stone.[56] In any case, it reinforces this paper's argument – to be developed in Section C. - that the DSA normatively aspires to a definition of AI systemic risk that is (i) socially contextualised; and (ii) refers to forms of harm that are propagated at scale and have, therefore, a distinctive collective nature.

2. Risk mitigation

24 After identifying and assessing the systemic risks stemming from their AI systems, platforms must proceed to the second step of the DSA's risk management framework: risk mitigation (art. 35 DSA). As the term 'mitigation' suggests, the endpoint of this stage of risk management is not to necessarily eliminate identified risks, but instead to reduce

[47] Haines (n 20) 183-184; Kaminski (n 19) 1392-1393.

[48] Jeroen van der Heijden, 'Risk Governance and Risk-Based Regulation: A Review of the International Academic Literature' [2019] State of the Art in Regulatory Governance Research Paper Series 25–26 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3406998> accessed 13 January 2024; Kaminski (n 19) 1354.

[49] Recital 79, 90, and art. 34 (2) DSA; European Commission, 2023 (n 9) 10, 13, 15; European Commission, 2024 (n 9) paras. 11-13.

[50] European Commission, 2023 (n 9) 15, 63.

[51] Art. 34(2) DSA.

[52] E.g., European Commission, 2024 (n 9), para. 31, where the Commission stresses the need for platforms to develop election-specific risk profiles in their assessments of systemic risks to electoral processes.

[53] Recital 80 DSA; European Commission, 2023 (n 9) 15, 17.

[54] Recitals 69, 79-83, art. 34(1) DSA.

[55] European Commission, 2023 (n 9) 15.

[56] ibid, 13.

their expected impact to acceptable levels.[57] But acceptable to whom? In other words, who is the ultimate decision-maker of what an acceptable AI systemic risk is and, consequently, of which measures are adopted to mitigate that risk? The answer is clear: VLOP/SEs. Similarly to risk assessment, it is for platforms to decide and put in place "reasonable, proportionate and effective mitigation measures" tailored to reduce the impact of previously assessed systemic risks (art. 35(1) DSA).[58] Articles 35 and 45 DSA contain a list of several possible risk mitigation measures which VLOP/SEs may choose from. Some of these measures are specifically relevant to the mitigation of AI risks, namely:

- Adapting the overall design and functioning of platforms' services and their online interfaces, which may in whole or in part be AI-driven (arts. 25 and 35(1)(a) DSA);
- Testing and adapting platforms' AI systems (art. 35(1)(d) DSA), with an emphasis on interventions related to the design of AI systems and finetuning of their parameters;[59]
- Ensuring that fake and deceptive AI-generated content (so-called deepfakes) is distinguishable as such (art. 35(1), k) DSA);[60]
- Adhering to codes of conduct - whose drawing up is promoted by the Commission - containing specific risk mitigation measures (art. 45(2) DSA).[61]

25 Similar to risk assessments, the Commission has also highlighted the need to contextualise risk mitigation measures. Specifically, it acknowledges that harmful high-risk content is not evenly distributed on platforms and might vary in time and/or between some audience segments.[62] Consequently, some individuals and communities might experience more severe levels of risk in certain moments in time. Therefore, the Commission recommends that risk mitigation measures be tailored to specific audiences[63] and time-specific contexts, such as elections/electoral campaigns.[64]

3. Risk management controls in the DSA's ecosystem

26 If, as shown above, VLOP/SEs are the ultimate decision-makers when it comes to systemic risk assessment and mitigation, how does the DSA ensure their accountability for those risk management choices? The response is threefold: platform compliance is monitored through (i) the internal compliance divisions of platforms themselves; (ii) independent audits contracted by platforms; and (iii) Commission or civil society adversarial audits based on DSA-mandated access to information.

a.) Internal compliance function

27 VLOP/SEs should, first and foremost, monitor compliance with the DSA from the inside. According to art. 41 DSA, they should establish an internal compliance division that is independent from their operational functions. This internal compliance division shall be headed by an "independent senior manager" who reports directly to the management body of VLOP/SEs (art. 41(2) DSA). Amongst the many tasks entrusted to it in art. 41(3) DSA, it is relevant in this case to highlight that the internal compliance function shall ensure that systemic risks are properly assessed in line with art. 34 DSA, subsequently reported, and appropriately mitigated in accordance with art. 35 DSA.[65]

b.) Independent audits contracted by platforms

28 In addition to having a compliance division tasked with internally monitoring compliance with the DSA, platforms shall "be subject, at their own expense and at least once a year, to independent audits" that assess their compliance with, amongst others, their

[57] Florian M Neisser, 'Riskscapes and Risk Management - Review and Synthesis of an Actor-Network Theory Approach' (2014) 16 Risk Management 88, 90; Kaminski (n 19) 1395, 1397.

[58] European Commission, 2024 (n 9) para. 8-10; De Gregorio and Dunn (n 15) 487-488. See also, in art. 35(1) DSA, "[s]uch measures *may* include (…)".

[59] Recital 88, DSA; European Commission, 2023 (n 9) 22-23.

[60] European Commission, 2024 (n 9), paras. 26, 28, 38.

[61] Recital 104 and art. 45 DSA. Interestingly, one voluntary code of conduct pre-dating the DSA, the Disinformation Code of Practice, should be made an official DSA code of conduct. This code of conduct could be particularly relevant in the context of mitigating risks of AI-generated content that is used in coordinated disinformation campaigns. See, to this effect, Recitals 84 and 106, DSA; European Commission, 2022 (n 28), commitments 14-16, p. 15-18; European Commission, 2023 (n 9) 12, 23; European Commission, 2024 (n 9), para. 58.

[62] For a similar, related, argument relating to uneven 'online visibility' of certain communities of users, see Rachel Griffin, 'The Law and Political Economy of Online Visibility: Market Justice in the Digital Services Act' (2023) 2023 Technology and Regulation 69, 71–73.

[63] European Commission, 2023 (n 9) 21-22.

[64] European Commission, 2024 (n 9) paras. 11-12, 37.

[65] Art. 41(3)(b) DSA.

risk assessment and mitigation obligations (art. 37(1)(a) DSA).[66] Within the DSA Framework, independent audits are considered an important tool for assessing platform compliance and, consequently, "meaningfully inform regulatory supervision".[67]

29 Independent audits may either be holistic, i.e., looking at how audited platforms assessed and managed all possible systemic risks listed in art. 34 DSA; or granular, i.e., focusing only on certain specific types or sources of systemic risks.[68] An audit might be more granular if it focuses only on how platforms have managed systemic AI risks stemming from the design and functioning of platforms' algorithms (recital 3 and art. 10(5)(b) and (c) DRA); or, even more specifically, those risks posed by a specific type of AI model or system (e.g., recital 25 DRA talks about auditing large language models).[69] They may also focus on certain types of systemic risks, e.g. those posed to fundamental rights.[70] Conversely, an audit may also have a more holistic focus if it examines how AI systems interact with a platform's overall design and thus contribute, in general, to the emergence of the different systemic risks covered by the DSA.

30 An example of a more holistic approach is the DSA Russian Disinformation study that was carried out by the Commission.[71] Despite not constituting a fully-fledged DSA audit (it was carried out in anticipation of the DSA's entry into force and so still with limited access to information),[72] this study is a good example of an analysis of platforms' systemic risks that holistically examines all sources of those risks in a specific context (i.e. Russian disinformation campaigns), including those risks stemming from AI systems.[73]

31 The output of each audit will be a report containing main findings, an overall opinion of the auditor on the platforms' compliance with the DSA,[74] and, if need be, operational recommendations for platforms to fully achieve compliance with the DSA.[75] These operational recommendations do not have to be necessarily followed by the audited platform, who has the discretion to determine the risk management measures they will implement (art. 37(6) DSA).[76]

c.) Commission and civil society adversarial audits

32 It is not particularly groundbreaking to state that, despite the effort of the DSA and DRA to secure the independence of the auditors contracted by platforms,[77] the risk of regulatory capture and/or ineffectiveness of audits still remains.[78] Indeed, there is huge potential for conflicts of interest and the development of pro-platform biases to surface in a scheme where auditors are contracted by platforms and will be, for a set period of time, contacting and collaborating directly with the personnel of VLOP/SEs .[79] Hence, it is worth exploring whether similar auditing exercises can, in the framework of the DSA, be carried out in a setting that is institutionally

[66] These audits must be carried out by independent auditing organisations with proven expertise in the area of risk management, as well as objectivity and professional ethics (art. 37(3) DSA). These organisations are contracted by platforms (art. 2(1) DRA) and will most likely be private consulting companies. See Giovanni De Gregorio and Oreste Pollicino, 'Auditing Platforms under the Digital Services Act' [2024] Verfassungsblog <https://verfassungsblog.de/dsa-auditors-content-moderation-platform-regulation/> accessed 25 September 2024; Alexander Hohfeld, 'DSA: Risk Assessment & Audit Database - First Round' (*Google Docs*, November 2024) <https://shorturl.at/STVQe> accessed 5 December 2024; Petros Terzis, Michael Veale and Noëlle Gaumann, 'Law and the Emerging Political Economy of Algorithmic Audits', *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2024) 1262–1263 <https://dl.acm.org/doi/10.1145/3630106.3658970> accessed 27 January 2025.

[67] Recital 1 DRA.

[68] The dichotomy between holistic or granular audits is not one of mutual exclusion, but rather of gradation. In simpler words, audits might be more or less granular.

[69] See recital 29 and arts. 2(17) and (18), 13(2) and 14(2) DRA for the types of audit exercises that auditors should carry out, namely so-called 'tests' and 'substantive analytical procedures'.

[70] European Commission, 2023 (n 9) 1, 12-13, 34, 48, 59-63.

[71] This audit was not carried out by an auditing organisation but the Commission stated its ambition to set, with this study, a baseline analytical framework to be used and iteratively improved by researchers and auditors; see European Commission, 2023 (n 9) 13.

[72] ibid 1, 12.

[73] ibid 34, 48, 59-63.

[74] This opinion might be 'positive', 'positive with comments' recommending specific but not major improvements, or 'negative'.

[75] Art. 37(4) DSA and art. 8 DRA.

[76] They may follow the operational recommendations; or, conversely, justify the reasons not to do so and set out other alternative measures. These choices must be featured in an implementation report produced by platforms within one month of receiving the audit report.

[77] Art. 37(3)(a) DSA; recital 2 and art. 4 DRA.

[78] De Gregorio and Pollicino (n 66).

[79] Meßmer and Degeling (n 40) 36; Martin Senftleben, 'Human Rights Outsourcing and Reliance on User Activism in the DSA' (*Verfassungsblog*, 21 February 2024) <https://verfassungsblog.de/human-rights-outsourcing-and-reliance-on-user-activism-in-the-dsa/> accessed 21 February 2024.

independent from platforms.

33 In this sense, one can find several hints in the DSA and related implementing law towards the possibility of other audits beyond those contracted out by platforms. I qualify those as 'adversarial audits', meaning more or less issue-specific risk audits or audit-like review exercises carried out by public authorities or civil society actors on the basis of publicly available or legally accessed information. Adversarial audits aim to scrutinise platforms' systemic risk management policies, actions and choices. In the DSA framework, I argue, adversarial audits can be conducted by (i) the Commission alone; (ii) the Commission in collaboration with civil society researchers ('collaborative adversarial audits'); or (iii) by civil society organisations and/or researchers themselves.

34 Firstly, conducting a risk adversarial audit could conceivably be one of "the necessary actions to monitor the effective implementation and compliance" with the DSA that the Commission may take in art. 72 DSA.[80] Interestingly, the Commission is able to appoint external experts and auditors to support the exercise of the aforementioned supervisory tasks (Art. 72(2) DSA, and recital 3 and art. 3(5)-(7) Implementing Regulation 2023/1021).[81] One can, therefore, deduce from these provisions the possibility of both Commission adversarial audits and collaborative adversarial audits.

35 Similarly, civil society organisations and independent researchers may, by themselves, conduct adversarial audits using information on platforms' risk management choices accessed through the mechanisms established in arts. 40 and 42 DSA.[82] By virtue of art. 40(4) and (8) DSA, certain researchers may be approved by national DSA supervisory authorities (so-called Digital Service Coordinators or 'DSCs') as 'vetted researchers'. These vetted researchers must, upon a request approved by a Digital Service Coordinator (art. 40(8) and (9) DSA), have access to data stored by VLOP/SEs that are needed for research that contributes to the detection, identification and understanding of systemic risks in the Union. This research may prove crucial to assess platforms' compliance with the risk assessment and mitigation obligations of arts. 34 and 35 DSA (recitals 96-98 and art. 40(4) DSA). It may point, for example, to certain emerging systemic risks overlooked by platforms, or to the insufficiency of their risk mitigation actions. Crucially, the output of vetted researchers' work must, per art. 40(8)(g) DSA, be made publicly available free of charge.[83]

36 To obtain vetted researcher status and have access to VLOP/SEs data, applicants must, in essence, (i) be affiliated with a research organisation within the meaning of art. 2(1) of Directive 2019/790 and (ii) have a project whereby they conduct research on platform-related systemic risks in the Union. The interpretation of these requirements and, therefore, the access of civil society actors to vetted research status depends on the case-by-case decisions of DSCs. These decisions will, in practice, determine to a significant extent who gets a meaningful possibility to carry out adversarial audits in the context of the DSA, meaning who gets sufficient access to data for in-depth systemic risk research.[84]

[80] Although information about the Commission's monitoring actions related to the DSA's systemic risk management scheme is not widely available to the public, one can see some references to Commission audits of platforms' compliance with such risk management scheme in, for example, 'Commission Opens Formal Proceedings against Facebook and Instagram under the Digital Services Act' (*European Commission*, 30 April 2024) <https://ec.europa.eu/commission/presscorner/detail/en/IP_24_2373> accessed 7 May 2024; 'Commission Opens Proceedings against TikTok under the DSA' (*European Commission*, 22 April 2024) <https://ec.europa.eu/commission/presscorner/detail/en/IP_24_2227> accessed 7 May 2024; 'Commission Sends Requests for Information to YouTube, Snapchat, and TikTok on Recommender Systems under the Digital Services Act' (*European Commission*, 2 October 2024) <https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-youtube-snapchat-and-tiktok-recommender-systems-under-digital> accessed 29 October 2024.

[81] See another form of collaboration between the Commission and individual experts in reviewing platform compliance under the DSA in 'Commission Sends Preliminary Findings to X for Breach of DSA' (*European Commission*, 12 July 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761> accessed 12 July 2024: "Based on an in-depth investigation that included, among others, the analysis of internal company documents, interviews with *experts* [N.B. emphasis added by author], as well as cooperation with national Digital Services Coordinators (...)". Similarly, but mentioning "third parties" and not "experts", see 'Commission Opens Formal Proceedings against Temu under DSA' (*European Commission*, 31 October 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_5622>.

[82] A similar argument pointing out this possibility is made in De Gregorio and Pollicino (n 66): (...) civil society organisations, which, considering the lack of reference in the delegated acts [N.B. on platform-contracted audits], are now looking more into the possibility of participating in this process and more generally to policy involvement in the DSA, also accessing data from online platforms".

[83] This research output might, interestingly, feed into the independent audits contracted by platforms, as it is one of the information sources that those auditing organisations must take into account per arts. 13(4) and 14(4) DRA.

[84] In Marsh's words, DSCs' decisions may also become a sort of "quasi case law" regarding both what are 'systemic risks'

37 Another hint towards the possibility of civil society adversarial audits is contained in recital 1 DRA, which alludes to the "enhanced scrutiny" of transparency reports of platforms. Indeed, per art. 42 DSA, platforms must make certain reports publicly available, including the results of their risk assessment and mitigation processes, as well as their audit reports and implementation reports (art. 42(4) DSA).

38 All in all, one can wonder whether the data accessed by vetted researchers, coupled with the data contained in transparency reports and other DSA access to information mechanisms,[85] may provide an overall level of insight into platforms' risk management processes that would allow the Commission and civil society to meaningfully scrutinise platforms' AI risk management. Similar regulatory scrutiny in tech regulation is often hampered by informational asymmetries between regulated actors and the public that favour the former and which they seek to preserve by citing trade secrecy and other commercial interests.[86] The enhanced access to information that arts. 40 and 42 DSA provide to public authorities and civil society actors is, therefore, key to concretising the unique public oversight promise of this regulation, as it may decisively tilt the regulatory balance towards information disclosure and consequently allow for evidence-based scrutiny of platforms' AI risk management.

39 Early reports on researcher access to information suggest that the corresponding DSA mechanisms might take a long time to be implemented properly.[87] But even those reports underline the potential of such access to information provisions: without them, researchers and civil society organisations will have a hard time auditing platforms' risk management based on high-quality and up-to-date information.[88]

II. Filling in the gaps: the AI Act

40 The AIA complements the DSA's systemic risk management framework. It does so in two distinct ways: through the institutionalisation of a residual risk management regime and by setting obligations that are relevant to how VLOP/SEs manage systemic risks related to the dissemination of inauthentic AI-generated content.

1. A residual risk management regime

41 Even if compliance with the DSA's systemic risk management framework creates a presumption that the corresponding AIA obligations have been fulfilled, the AIA is still relevant to manage newly identified "significant systemic risks" of platforms' AI systems and models that are not covered in the DSA (Recital 118, AIA). It is, therefore, useful to understand what AI systemic risks are not covered by the DSA and can, *a contrario*, be managed through the AIA. These will be, in essence, those systemic risks posed by AI systems and models that are not a possible source of systemic risks per the DSA.

42 In this respect, it should be clarified that the DSA *prima facie* applies to the algorithmic systems of platforms and any related systems (art. 34(1) DSA), meaning both the AI systems that are embedded in a platform's service and are thus behind its operation; and AI systems which *are* the service's digital infrastructure, which is typically the case of AI-powered search engines such as ChatGPT or Google

under art. 34 DSA and what is a 'vetted researcher' for these purposes; see Marsh (n 40) 6–7. For more information on DSA data access requests, see arts. 3, 7 – 13 of the Access to Data Delegated Regulation.

[85] Referring to the need to combine several data access points and transparency mechanisms to research platforms' compliance with the DSA, see Rishabh Kaushal and others, 'Automated Transparency: A Legal and Empirical Analysis of the Digital Services Act Transparency Database' (arXiv, 2024) 14 <http://arxiv.org/abs/2404.02894> accessed 6 April 2024.

[86] Cary Coglianese, 'Regulating New Tech: Problems, Pathways, and People' 5 <https://scholarship.law.upenn.edu/faculty_scholarship/2753>; Madalina Busuioc, Deirdre Curtin and Marco Almada, 'Reclaiming Transparency: Contesting the Logics of Secrecy within the AI Act' (2023) 2 European Law Open 79, 82, 88; Marta Maroni, '"Mediated Transparency": The Digital Services Act and the Legitimisation of Platform Power' in Päivi Leino-Sandberg, Maarten Zbigniew Hillebrandt and Ida Koivisto (eds), *(In)visible European Government: Critical Approaches to Transparency as an Ideal and a Practice* (Routledge, 2023).

[87] Marsh (n 40) 13–14; Julian Jaursch, Jakob Ohme and Ulrike Klinger, 'Enabling Research with Publicly Accessible Platform Data: Early DSA Compliance Issues and Suggestions for Improvement' (Weizenbaum Institute 2024) <https://www.weizenbaum-library.de/handle/id/572> accessed 28 November 2024; Mateus Correia de Carvalho, 'Researcher Access to Platform Data and the DSA: One Step Forward, Three Steps Back' (*Tech Policy Press*, 31 May 2024) <https://techpolicy.press/researcher-access-to-platform-data-and-the-dsa-one-step-forward-three-steps-back> accessed 27 September 2024; Philipp Darius, 'Researcher Data Access Under the DSA: Lessons from TikTok's API Issues During the 2024 European Elections' (*Tech Policy Press*, 24 September 2024) <https://techpolicy.press/-researcher-data-access-under-the-dsa-lessons-from-tiktoks-api-issues-during-the-2024-european-elections> accessed 26 September 2024.

[88] ibid.

Bard.[89] Consequently, the AIA risk management framework applies to any AI systems and models producing systemic risks, but that are not considered as embedded or integrated in a platform's service per the DSA. This is, for example, the case of AI systems and models whose content is diffused or amplified by platforms' recommender systems. Through the AIA, the companies that develop those AI systems (AI providers) and those that are placing them on the market (AI deployers) might be called upon to manage their systemic risks.

43 Having clarified the scope of application of the AIA's systemic risk management scheme, a new question arises: how does the AIA define and purport to manage systemic risks? According to art. 3(65) AIA, an AI systemic risk may solely stem from a specific type of AI model, i.e., a general-purpose AI model (hereinafter 'GPAI'), which is an AI model that can competently perform a wide range of distinct tasks, being typically trained with a large amount of data (art. 3(63) AIA). More specifically, a GPAI can only be a source of systemic risk if it displays either (i) 'high-impact capabilities' - meaning those capabilities that, according to some computational metrics, match or exceed those recorded in the most advanced GPAIs[90]; or, alternatively, capabilities or an impact deemed by the Commission to be equivalent to 'high-impact capabilities'.[91]

44 For a GPAI to present a systemic risk, its high-impact capabilities must negatively affect at least one of a number of protected issues (i.e., public health, safety, public security, fundamental rights or other goods that benefit societies as a whole), with a reach and propagation at a scale that is significant enough to warrant the qualification of 'systemic'.[92] It is the Commission who will ultimately decide, either ex officio or following a qualified alert issued by a scientific panel of independent experts (arts. 52(4), 68, and 90 AIA), whether a given GPAI presents a systemic risk (art. 52 AIA).[93]

45 If a GPAI is deemed to present new systemic risks, then their providers must comply with a set of product safety and risk management obligations listed in art. 55(1) AIA. Of most relevance here are the obligations for GPAI providers to perform model evaluations and testing in order to identify, assess and mitigate emerging systemic risks (art. 55(1)(a), (b) AIA). The AIA offers two main ways to simplify compliance with these obligations: (i) the compliance by the GPAI provider with a European harmonised technical standard (arts. 40 and 55(2) AIA);[94] or (ii) the adherence to codes of practice drawn up at EU level and containing several measures aimed at assessing and managing systemic risks of GPAI models (arts. 55(2) and 56 AIA).[95]

[89] Vermeulen and Lemoine (n 13).

[90] Arts. 3(64) and (67); art. 51(1)(a) AIA). The computational metric privileged in the context of this assessment is the number of floating operation points (FLOP) of an AI system, see art. 51(2) AIA.

[91] Art. 51(1)(b) and Annex XIII AIA. See, Charlie Bullock and others, 'Legal Considerations for Defining "Frontier Model"' (Institute for Law & AI, LawAI Working Paper Series, No 2-2024 2024) 13 <https://law-ai.org/wp-content/uploads/2024/09/Legal-Considerations-for-Defining-Frontier-Model.pdf> accessed 29 January 2025, pointing out that the AIA's systemic risk regime may become underinclusive of certain GPAIs if it overemphasizes computational metrics in this classification. Also criticising this regime for its uncertainty and potential underinclusive nature, Cornelia Kutterer, 'Regulating Foundation Models in the AI Act: From "High" to "Systemic" Risk' (AI-Regulation Papers 2024) 6–7 <https://ai-regulation.com/wp-content/uploads/2024/01/C-Kutterer-Regulating-Foundation-Models-in-the-AI.pdf> accessed 29 January 2025.

[92] Per art. 3(65) AIA, a "'systemic risk' means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain".

[93] Annex XIII to the AIA contains a (non-exhaustive) set of criteria that the Commission shall take into account when designating that a GPAI presents systemic risk, such as indicators related to the design and ability of the AI model (e.g., number of parameters, size of dataset, autonomy to perform new tasks), the number of registered end-users of the GPAI, as well as the reach of the GPAI in the internal market, which shall be presumed when the model has at least 10.000 registered business users in the Union.

[94] European harmonised standards are technical standards developed by private standardisation organisations at the request of the Commission and containing technical specifications on how to comply with the requirements set in EU secondary law. Voluntary compliance with these standards will grant an AI provider a presumption of conformity with the obligations set out in art. 55(1) AIA. See Regulation 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation art. 2(1); Annalisa Volpato, 'The Legal Effects of Harmonised Standards in EU Law: From Hard to Soft Law, and Back?', *The Legal Effects of EU Soft Law* (Edward Elgar Publishing 2023).

[95] Codes of practice are drawn up in a collaborative process coordinated by two new governance bodies created by the AIA to support the Commission in its enforcement actions, i.e., the AI Office and the AI Board, and involving GPAI providers, national competent authorities, civil society organisations, researchers, and other stakeholders (art. 56(3) AIA). The first code of practice on GPAI systemic risk management is already started being drafted, see Nuria Oliver and others, 'First Draft of the General-Purpose AI Code of Practice' (European Commission 2024) <https://shorturl.at/irQTc> accessed 19 November 2024.

46 Although according to the AIA, the DSA systemic risk management framework of platforms' AI systems is to be primarily complemented by the AIA provisions on systemic risk management of GPAIs described up until this point, some other AIA requirements are relevant for platforms as deployers of AI models and systems. Indeed, the AIA systemic risk regime focuses on risks stemming from AI models and systems *as a whole* (i.e., a given GPAI presents, in itself, a systemic risk that should be managed accordingly). In addition, however, the AIA creates two distinct legal regimes for certain *practices* or *uses* of all AI systems (including GPAIs[96]): those that (i) present unacceptable risks considering the EU's values (art. 5 AIA); or that (ii) are used for high-risk purposes (art. 6 AIA). How are these two additional risk regimes relevant for the risk management of platforms' AI systems?

47 Looking first at the prohibited AI practices of art. 5 AIA, these are said to be "particularly harmful and abusive" and should, according to the EU legislator, be prohibited in the EU, since they "contradict Union values of respect for human dignity, freedom, equality, democracy and the rule of law and Union fundamental rights" (recital 28 and art. 5 AIA). It would, therefore, be contradictory to prohibit certain AI practices in the AIA because of a fundamental misalignment with the EU's core values and, at the same time, not extend that prohibition to AI practices of VLOP/SEs in the DSA. Some of these prohibited practices should be considered by platforms when designing and integrating AI systems into their services. Namely, they should consider the prohibition of deployment of manipulative or deceptive subliminal techniques operating beyond a person's consciousness, with the objective or effect of materially distorting a person's or a group of persons' behaviour by appreciably impairing their ability to make an informed decision (art. 5(1)(a) AIA). This links well with certain provisions of the DSA that prohibit the design of platforms' services in a way that materially distorts or impairs individuals' ability to make free and informed decisions (recitals 37, 79, 81, 83 and arts. 25, 34(2) DSA).

48 At the same time, any AI systems and models integrated or diffused in platforms (and not covered by the DSA) might also be classified as high-risk AI systems if they are used for any of the purposes mentioned in art. 6 and Annex III to the AIA. Even if these AIA provisions do not mention the AI systems of online platforms explicitly, the Commission may, theoretically, add them to that list by adopting a delegated act that modifies Annex III (art. 7 AIA). In any event, there might already be leeway in Annex III to classify as high-risk certain AI systems integrated or used in platforms, namely those that are "intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda" (Annex III, point 8(b)).

49 Should an AI system (including a GPAI, per art. 25(1)(c) AIA) be used for high-risk purposes in accordance with art. 6 AIA, then its providers and deployers will be subject to a host of legal requirements destined to ensure the safety of the AI system in question and a related adequate level of fundamental rights protection.[97] Amongst those requirements, some are particularly relevant in the context of risk management. According to Art. 9 AIA, AI providers and deployers should establish a risk management scheme that first, regularly and systematically identifies the (high-level) risks of AI systems and, subsequently, mitigates or eliminates them. This risk management scheme focuses particularly on risks posed by high-risk AI systems to health, safety and fundamental rights (art. 9(2)(a) AIA). It is worth nothing that art. 9 AIA does not require AI providers and deployers to mitigate or eliminate all identified high-risks, but only to do so up to a *reasonable* extent (art. 9(3) AIA) and through specific courses of action: either the better design and development of the high-risk AI systems *or* the provision of adequate technical information. In addition, articles 10 to 15 AIA contain a number of AI safety requirements that can conceivably be implemented as risk mitigation measures, e.g., ensuring the quality of the data sets used for training, validation and testing of AI systems (art. 10 AIA), human oversight requirements (art. 14 AIA), or ensuring an appropriate level of cybersecurity of AI systems (art. 15 AIA).

50 Crucially, however, it is for AI providers and deployers themselves to judge whether high-risk mitigation is sufficient. Although art. 9 AIA prefers to formulate most risk management requirements in the passive voice,[98] it is clear from the logic of the article and related provisions (e.g., recital 46, and arts. 6(4) and 8(1) AIA) that it is for AI providers and

[96] GPAIs are in fact subject to both the systemic risk management framework of articles 51-55 AIA, as well as to the prohibitions and requirements flowing from certain unacceptable or high-risk uses of such models. This is made clear in art. 25(1)(c) AIA, points out the possibility for a GPAI – which can, as seen above, be classified as presenting systemic risk - to be classified as presenting high-risk pursuant to art. 6 AIA.

[97] Similarly to compliance with the obligations imposed on providers of GPAIs with systemic risk, compliance with the requirements imposed on providers and deployers of high-risk AI systems will be presumed through adherence to harmonised standards (art. 40 AIA).

[98] E.g., 'A risk management system shall *be established* (...)'; 'The risks (...) *may be* resonsably *mitigated or eliminated*'; or 'The risk management measures shall be such that the relevant residual risk (...) is *judged* to be acceptable'.

deployers to implement a risk management system and ultimately decide, with considerable flexibility, whether identified risks have been reduced to an acceptable level.[99]

2. The AIA's other DSA-relevant provisions: of deepfakes and sandboxes

51 Aside from the AIA risk regimes that were discussed, other provisions are relevant for VLOP/SEs as they seek to manage the systemic risks of their AI systems pursuant to the DSA.

52 Namely, the Commission's DSA risk mitigation guidelines state that VLOP/SEs should pay particular attention to the "creation (...) and large-scale dissemination of generative AI content" and that the AIA contains particularly relevant obligations of watermarking and labelling of 'deep fakes' and synthetic AI content.[100] This is a clear reference to art. 50 AIA, which requires that, on the one hand, providers of genAI ensure that the outputs of those models are marked as artificially generated and manipulated; and that, on the other hand, deployers of genAI use state-of-the-art technical solutions to disclose that synthetic AI content was artificially generated (recital 120 and art. 50(2) AIA). These requirements are, as stated in recital 120 AIA, particularly relevant for the effective implementation of the DSA when it comes to mitigating systemic risks to democratic processes and civic discourse.

53 One final note should be made to reference AI regulatory sandboxes, which the AIA institutes (arts. 3(55) and 57 AIA) as a controlled framework set up by a supervisory authority where current or prospective AI providers can develop their AI systems with a view to identifying potential risks to fundamental rights, health and safety of future users. Regulatory sandboxes are often referred to as a valuable feature of any risk regulation toolbox, due to their experimental nature and related potential to gauge and anticipate emerging risks of AI systems.[101] VLOP/SEs could conceivably use this AIA-institutionalised framework when developing or adapting their AI systems.

C. What they have in common: sociotechnical and contested systemic risk

54 In the previous section, I have described the AI risk management regimes of the DSA and the AIA applicable to platforms' AI systems. That, however, is not enough to answer the main question guiding this paper: how do the DSA and AIA foresee creating an effective risk regulatory regime applicable to the AI systems of digital platforms? In other words, how do these regulations intend to address the typical challenges of any risk-based approach?

55 As laid out in the introduction, two main typical challenges are posed to effective risk regulation, especially in the field of AI. Firstly, its excessively quantitative and actuarial focus might make platforms and public authorities overlook less quantifiable AI risks whose impact is not reduced nor explained through single instances of harm caused to individuals. A second challenge is that AI risk regulation gives significant discretion to private regulated actors regarding how they identify, measure and mitigate emerging AI risks. If not adequately controlled, platforms and other AI providers or deployers might exercise this discretion in self-serving ways, by overlooking certain emerging AI risks, underestimating their impact, and/or putting insufficient measures in place to adequately mitigate those risks.

56 In this section, I argue that both the DSA and AIA contain similar guiding ideas for how to address the abovementioned challenges. Besides the foreseen complementarity between the two regulations' AI risk management schemes, they have in common three main normative commitments and aspirations as to how AI risks should be managed. In particular, AI risks should be framed as systemic (I.); their identification, assessment and mitigation should be done through methodologies that socially contextualise the impact of those risks (II.) and civil society should be actively involved in the corresponding risk management processes (III.). These three main commonalities between the DSA and AIA's risk management regimes give further credence to the argument, advanced at the beginning of Section B., that one integrated EU AI

[99] See, to this effect, in recital 46: '(...) providers of a product that contains one or more high-risk AI systems (...) should have flexibility with regard to operational decisions on how to ensure compliance of a product that contains one or more AI systems with all applicable requirements of the Union harmonisation legislation in an optimal manner'; and in art. 9(5) AIA: 'The risk management measures (...) shall be such that the relevant residual risk (...) as well as the overall residual risk of the high-risk AI systems is judged to be acceptable'.

[100] European Commission, 2024 (n 9) paras. 25-30.

[101] Sofia Ranchordas and Valeria Vinci, 'Regulatory Sandboxes and Innovation-Friendly Regulation : Between Collaboration and Capture' (2024) <https://papers.ssrn.com/abstract=4696442> accessed 5 March 2024; Kaminski (n 19) 1371.

risk management framework applicable to digital platforms with similar normative foundations can be distilled from the two regulations.

57 In the remainder of this section, I will detail these three main normative commonalities. Before proceeding, it is important to underline that this section's argument should not be interpreted to mean that *only* three commonalities exist between the DSA and AIA. Instead, they are argued to be the normative commitments that most acutely and specifically impact how AI risk management processes should be carried out in EU law. For example, one could also note that both the DSA and AIA place a special emphasis on fundamental rights protection as one of their main aims.[102] This point logically extends to the two regulations' systemic risk provisions.[103] Such an emphasis on fundamental rights requires that this risk management framework be interpreted in light of the EU Charter and the ECHR, as an integrated attempt to protect fundamental rights through risk in the AI context.[104]

I. The emphasis on systemic risk: a sociotechnical frame

58 One first commonality that can be distilled from the description of the DSA and AIA's respective AI risk management regimes is that they frame the risks of digital platforms as 'systemic'. Furthermore, as laid out above, the two regulations foresee the complementary of their systemic risk management regimes.[105] Therefore, there should be some communication between the concept of systemic risk adopted in the AIA and DSA. This is an important insight since the DSA does not define the concept of systemic risk. As pointed out in Section B.I., there is no clear indication in the DSA of when an AI risk should be considered to be systemic. To answer this question, one can look at the corresponding AIA definition (contained in art. 3(65) and Annex XIII AIA). The AIA considers that an AI risk is considered systemic if it has "a significant impact on the internal market due to its reach, (...) with actual or reasonably foreseeable negative effects (...) that can be propagated at scale".

59 This conceptualisation of systemic risk as implying a considerable reach and propagation at scale of the effects of AI systems aligns well with the fact that both the DSA and AIA conceptualise the negative effects of AI systemic risks by reference to forms of collective - rather than individual - harm. In fact, both the DSA[106] and the AIA[107] refer to the societal negative effects of platforms' AI systems on democratic processes such as elections and civic discourse; public health, safety and security; or widespread gender-based violence and negative effects on fundamental rights and mental health. The AIA even mentions "negative effects (...) on society as a whole" (art. 3(65) AIA). This also means that AI risks are not just seen in these regulations as actuarial, quantitative and reduced to individual instances of harm. Therefore, and although the literature rightly points out the possibility that these regulations may be implemented with an exclusive (or, at least, predominant) focus on individual[108] and quantifiable[109] interests, there is potential for the DSA and AIA to also take into account collective, societal or cumulative[110] forms of AI harm that are not explainable nor reducible to singular instances of individualised harm.[111]

60 Despite all the foregoing indications regarding the meaning of 'AI systemic risk', many conceptual questions remain:

- Is an AI risk systemic only if it affects societal systems, structures and collective goods, such as democratic processes,[112] the free access to

[102] De Gregorio and Dunn (n 15) 493-498; Almada and Petit (n 22) 17-18. See also the Commission's explanatory memorandum in the AIA proposal in European Commission, 2021 (n 35) 1-4; recitals 3, 9, 36, 40, 41, 47, 51, 52, 63, 79, 81, 86, 87, 107, 109, 111, 153, 155, and arts. 1, 14(4), 34(1)(b), 35(1) DSA; and recitals 1, 2, 3, 5, 6, 7, 8, 10, 28, 32, 43, 46, 48, 52, 65-66, 93, 96, 118, 139-140, and arts. 1, 3(65); 6(3), (6)-(8); 7(1)(b), 2(e) and (i), 3; 9(2)(a); 10(2)(f), (5); 13(3)(b)(iii); 14(2); 40(3); 41(1)(a) (iii); 57(6); 58(2)(i), (4); 77; 82(1) AIA.

[103] Recitals 79, 81, 86, and arts. 34(1)(b), 35(1), (3) DSA; recital 118 and art. 3(65) AIA.

[104] This fundamental rights-friendly interpretation is mentioned specifically in recitals 153 DSA and 2, 7 AIA. See also European Commission, 2021 (n 35) 4.

[105] See Section B.; and recital 118 AIA; European Commission, 2024 (n 9) paras. 26-30.

[106] See Section B.I.1. and, in particular, *supra* footnote 54.

[107] Art. 3(65) AIA.

[108] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22 Computer Law Review International 97, 99; Rachel Griffin, 'Rethinking Rights in Social Media Governance: Human Rights, Ideology and Inequality' (2023) 2 European Law Open 30, 42–46.

[109] Kaminski (n 19) 1391–1393.

[110] By cumulative, I mean forms of harm that are caused in successive instances over time 'without a single event tripping a threshold of seriousness, leaving it difficult to prove', as defined in Veale and Borgesius (n 108) 99.

[111] Making a similar argument with regard to the DSA, Eder (n 24) 3. For an overview of the different forms of AI harm that AI regulations may address, see Nathalie A Smuha, 'Beyond the Individual: Governing AI's Societal Harm' (2021) 10 Internet Policy Review 4–12.

[112] Barbara Zmušková, 'Progressive Slovakia Becomes Target of AI Misinformation, Tops Polls' (*Euractiv*, 28

information and exchange of ideas in public fora,[113] or the environment?[114]

- Or also if it harms so many individuals at a scale (due to a platform's reach) that makes it systemic as, for example, in the case of widespread potential effects of platforms' AI systems promoting or heightening the risk of generating mental addiction[115] or gender-based violence?[116]

- Or even if AI recommender systems and genAI threaten to harm a few select individuals but with such a big reach that such harm attains a significant level of propagation across societies, as for example in the case of targeted deepfake porn campaigns towards female figures that are prominent opposers of Vladimir Putin's regime?[117]

- Or all of the above?

- Relatedly, how should one conceive and measure systemic AI risks – which, at least in part, allude to negative *collective* effects on social and political structures - when it comes to affected legal goods, such as fundamental rights, that are traditionally conceived as belonging to individuals and protected through *individual* remedies that address specific instances of harm?[118]

- Or, differently, an AI risk is systemic not (or not just) because of its systemic effects on societies and individuals, but (also) because those risks arise from systems, e.g. the digital public spaces created by digital platforms, their mediation through AI systems integrated therein, or the system-level content moderation policies of platforms that are, in turn, implemented by automated systems?[119]

61 A recent report found that researchers working on the DSA systemic risk provisions struggle to answer these and other related questions since they have very different views on whether different specific

September 2023) <https://www.euractiv.com/section/politics/news/progressive-slovakia-becomes-target-of-ai-misinformation-tops-polls/> accessed 23 September 2024; Joy Hyvärinen, 'Hostile Information Campaigns Could Test a Divided Finland' (*Tech Policy Press*, 30 May 2024) <https://techpolicy.press/hostile-information-campaigns-could-test-a-divided-finland> accessed 31 May 2024; Victoria Oldemburgo de Mello, Felix Cheung and Michael Inzlicht, 'Twitter (X) Use Predicts Substantial Changes in Well-Being, Polarization, Sense of Belonging, and Outrage' (2024) 2 Communications Psychology 1.

[113] Laufer and Nissenbaum (n 11) 5–6; Article 19 and others, 'Civil Society Open Letter to Commissioner Breton' (17 October 2023) <https://www.article19.org/wp-content/uploads/2023/10/Civil-society-open-letter-to-Commissioner-Breton.pdf> accessed 9 October 2024.

[114] Rachel Griffin, 'Climate Breakdown as a Systemic Risk in the Digital Services Act' (*Hertie School Centre for Digital Governance*, 7 September 2023) <https://www.hertie-school.org/en/digitalgovernance/news/detail/content/climate-breakdown-as-a-systemic-risk-in-the-digital-services-act> accessed 19 February 2024.

[115] Aksha M Memon and others, 'The Role of Online Social Networking on Deliberate Self-Harm and Suicidality in Adolescents: A Systematized Review of Literature' (2018) 60 Indian journal of psychiatry 384; Amandeep Dhir and others, 'Online Social Media Fatigue and Psychological Wellbeing—A Study of Compulsive Use, Fear of Missing out, Fatigue, Anxiety and Depression' (2018) 40 International Journal of Information Management 141; Ashlee Milton and others, '"I See Me Here": Mental Health Content, Community, and Algorithmic Curation on TikTok', *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery 2023) <https://doi.org/10.1145/3544548.3581489> accessed 15 February 2024.

[116] Silvia Semenzin and Lucia Bainotti, 'The Use of Telegram for Non-Consensual Dissemination of Intimate Images: Gendered Affordances and the Construction of Masculinities' (2020) 6 Social Media + Society 2056305120984453; Thiago Dias Oliva, Dennys Marcelo Antonialli and Alessandra Gomes, 'Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online' (2021) 25 Sexuality & Culture 700; Brennan Suen, Carly Evans and Alex Paterson, 'Right-Leaning Facebook Pages Earned Nearly Two-Thirds of Interactions on Posts about Trans Issues' (*Media Matters for America*, 9 November 2021) <https://www.mediamatters.org/facebook/right-leaning-facebook-pages-earned-nearly-two-thirds-interactions-posts-about-trans> accessed 3 July 2024.

[117] European Commission, 2023 (n 9) 30. See also Gretchen Peters, 'Time to Act on Harmful Deepfakes & Algorithms' (*Tech Policy Press*, 31 October 2024) <https://techpolicy.press/time-to-act-on-harmful-deepfakes-algorithms> accessed 20 November 2024.

[118] For a lengthier discussion of this theoretical issue applied to platform regulation see Griffin, 'Rethinking Rights in Social Media Governance' (n 108) 46–55.

[119] To see similar approximations to this conceptual question, see Sally Broughton Micova and Andrea Calef, 'Elements for Effective Systemic Risk Assessment under the DSA' (Centre on Regulation in Europe (CERRE) 2023) 11–13 <https://cerre.eu/wp-content/uploads/2023/07/CERRE-DSA-Systemic-Risk-Report.pdf> accessed 16 May 2024, discussing the systemic provenance and effects of digital platforms' potential harms and how they contributed to the use of the notion of 'systemic risk' in the DSA; and Griffin, 'Rethinking Rights in Social Media Governance' (n 108) 55, mentioning that the DSA (although with regard to its art. 14[4] and not art. 34) might serve to address cases of 'systemic injustice', stemming from 'system-level enforcement of platforms' content policies'; Pielemeier and Sullivan (n 40).

cases constitute evidence of systemic risk.[120] Beyond researchers, it is highly likely that different platforms, public authorities and civil society organisations will have different understandings of what an AI systemic risk is. And, as demonstrated in Section B., the legislative indications for the definition of this concept are scarce. It appears that, in EU AI systemic risk management processes, the definition of what is to be managed – systemic risks – will be iteratively constructed, on a case-by-case basis. This is consistent with the fact that AI risk regulation, as any form of risk regulation, is process-based: it is not concerned with setting substantive standards beforehand but, differently, is predicated on the fact that those substantive standards will be generated by the successive outcomes of risk management processes.[121] Consequently, and as tautological as it may seem, an AI systemic risk will be whatever is defined (and then managed) as an AI systemic risk in the DSA and AIA's risk management processes.

62 This lack of conceptual clarity might disappoint some. But instead of causing disappointment, the indeterminacy of the concept of 'AI systemic risk' should, I argue, prompt a shift in our analytical focus. Particularly, if the definition of 'AI systemic risk' is to be constructed as different types of AI systemic risks are progressively identified and managed, it is key to analyse *how* the corresponding risk management processes develop and, especially, *who* has more agency in influencing their outcomes and, therefore, in shaping the meaning of AI systemic risks. When presented with the set of conceptual questions listed above, we should, I argue, answer with another set of - preliminary - questions. These are more oriented towards methodological and practical issues, but answering them will necessarily lead us to bigger conceptual clarity on the meaning of AI systemic risks:[122]

- Who has more agency/power – private regulated actors, public supervisory authorities, civil society organisations, researchers, or other stakeholders – in shaping the concept of AI systemic risk as the DSA and AIA are implemented?
- What are the ideas of what AI systemic risks are that gain more currency in the early regulatory dialogue?
- Based on which information and evidence do different actors across the DSA institutional ecosystem conclude for the (in)existence of a systemic risk? Are all actors given the same possibility to access high quality and up-to-date evidence to assess AI systemic risks?[123]
- Which frameworks and methodologies are used by different actors to identify and measure systemic risks in concrete cases?
- What (political)[124] priorities are set by different actors regarding systemic risk management? In other words, on what specific types of systemic risk will these actors concentrate their resources for risk assessment and management?

63 In this sense, it is worth noting that both risk regulation and, more specifically, the notion of 'systemic risk' have 'baggage'. As Kaminski points out, risk regulation in general has a certain policy baggage: the typical tools, tactics, and troubles of risk regulation as implemented in other fields are transposed into AI regulation by the policymaking decision to frame and regulate AI harms as risks.[125] Among several elements of such policy baggage are the difficulty of risk regulation to capture and manage unquantifiable harms, as well as its typical technocratic and "techno-correctionist" nature, which means that "it largely tries to fix problems with existing technologies rather than considering whether it would be better to put regulatory energy elsewhere – including not to use a technology at all".[126] As such, when risk regulation is used to address technological problems that entail policy and political decisions, it can obfuscate the latter

[120] Marsh (n 40) 5–12.

[121] *Supra* footnote 45.

[122] Here, I take a slightly different stance than Marsh (n 40) 1, who, when reporting on researchers perceptions on systemic risk assessment in the DSA, argued that the "more pressing problems" when researching systemic risks under the DSA are "practical rather than conceptual". In my view, more practical questions are indeed very important but, crucially, because they influence and inform one's answer to the conceptual questions regarding the definition and assessment of systemic risks under the DSA. Conceptual questions are, ultimately, still more pressing; but they are, to a large extent, pre-determined by practical and material considerations.

[123] Early reports of DSA access to information suggest that researchers/civil society have significant difficulties in accessing information both from platforms and public authorities. See, e.g. ibid 14; Darius (n 87).

[124] Josephine Adekola, *Power and Risk in Policymaking: Understanding Public Health Debates* (Springer International Publishing 2020) 13–19; Griffin, 'What Do We Talk about When We Talk about Risk?' (n 15).

[125] Kaminski (n 19) 1389–1403.

[126] ibid 1390.

and "shield them from democratic accountability".[127]

64 The concept of 'systemic risk' arguably also has a distinct baggage. This concept has most extensively been used to measure risks of widespread instability in the financial sector.[128] It is within this field that the literature on systemic risk is most developed. This has already led some to test the application of that systemic risk framework to the DSA context.[129] The adequacy of the transplant of financial systemic risk frameworks to the DSA can be questioned for many reasons. Those frameworks are, equally, predominantly quantitative and highly technical,[130] which may lead to the same troubles signalled by Kaminski regarding risk regulation in general. If these systemic risk frameworks are transplanted into the management of AI systemic risks in EU law, they may thus turn such management into a predominantly technical exercise that fails to fully engage with the social meaning of platforms' AI systems and, therefore, to address less quantifiable AI harms.[131]

65 Conversely, both the DSA and AIA call for AI systemic risks to be framed in sociotechnical terms. Indeed, both regulations mention that risk management processes must consider the impact of AI technology on public values, and political and societal processes.[132] Moreover, they stress the need for AI risks to be assessed and managed depending on the specific social contexts where platforms' AI systems operate and with which they interact.[133] This can only be achieved if AI systemic risk management is framed in sociotechnical terms. This ultimately means conceiving AI risks as stemming not just from AI systems as technological artifacts; but, instead, from the (dynamic) interactions between AI systems and society.[134] It requires, that we understand AI technologies and AI-generated content - as well as the digital platforms integrating or spreading them - as part of broader social systems, i.e., configurations where they shape and are shaped by existing social practices (including values, norms, institutions, relationships, multiple different actors, and other technologies).[135] With this lens, one cannot escape the fact that AI systems mediate several aspects of social life and, in so doing, catalyse social and cultural change.[136] Therefore, AI systemic risk management should not be reduced to technological considerations, framed in solely technical terms and measured quantitatively. On the contrary, it should also capture the social, political, cultural – and thus less quantifiable – meaning and impact of AI technologies.[137]

[127] ibid 1397; see also Griffin, 'What Do We Talk about When We Talk about Risk?' (n 15).

[128] See, e.g., Paweł Smaga, 'The Concept of Systemic Risk' (The London School of Economics and Political Science 2014) Systemic Risk Centre Special Paper, No 5; Robert Engle, Eric Jondeau and Michael Rockinger, 'Systemic Risk in Europe' (2015) 19 Review of Finance 145.

[129] Broughton Micova and Calef (n 119) 9.

[130] See, e.g., Engle, Jondeau and Rockinger (n 128) 148–156.

[131] For a discussion of other limitations of transplanting financial systemic risk frameworks into EU AI systemic risk management (in this case regarding DSA systemic risk management), see 'Implementing Risk Assessments under the Digital Services Act, Discussion Summary of the Workshop "Implementing Risk Assessments under the Digital Services Act"' (Global Network Initiative, Digital Trust & Safety Partnership and Brainbox 2023) 5 <https://dtspartnership.org/wp-content/uploads/2023/06/Discussion-summary-%E2%80%93-GNI-and-DTSP-workshops-on-implementing-risk-assessments-under-the-DSA-June-2023.pdf#page=12> accessed 1 July 2024; Alice Palmieri, Konrad Kollnig and Aurelia Tamò-Larrieux, 'Systemic Risks of Dominant Online Platforms: A Scoping Review' (Social Science Research Network, 2024) 8–9 <https://papers.ssrn.com/abstract=5002743> accessed 12 December 2024.

[132] E.g. recitals 6, 27, 61, 110, and art. 3(65) AIA; and art. 34(1)(c)(d) DSA.

[133] E.g. recital 20 AIA ("AI literacy should equip providers, deployers (...) with the necessary notions to make informed decisions regarding AI systems. Those notions may vary with regard to the relevant context and can include understanding (...), in the case of affected persons (...) how decisions taken with the assistance of AI will have an impact on them"); or recital 90 DSA ("Providers of very large online platforms and of very large online search engines should ensure that their approach to risk assessment and mitigation is based on the best available information and scientific insights and that they test their assumptions with the groups most impacted by the risks and the measures they take.") and European Commission, 2023 (n 9) 13-15. See more in Section C.II. below.

[134] Merel Noorman and Tsjalling Swierstra, 'Democratizing AI from a Sociotechnical Perspective' [2023] Minds and Machines 4–5 <https://link.springer.com/10.1007/s11023-023-09651-z> accessed 8 February 2024; Brian J Chen and Jacob Metcalf, 'Explainer: A Sociotechnical Approach to AI Policy' (Data & Society 2024) 2–5 <https://datasociety.net/library/a-sociotechnical-approach-to-ai-policy/> accessed 3 October 2024; Brian Chen, 'Why AI Policy Needs a Sociotechnical Perspective' (*Tech Policy Press*, 29 May 2024) <https://techpolicy.press/why-ai-policy-needs-a-sociotechnical-perspective> accessed 29 May 2024.

[135] Noorman and Swierstra (n 134) 4.

[136] Julie E Cohen, 'Configuring the Networked Citizen' in Austin Sarat, Lawrence Douglas and Martha Merrill Umphrey (eds), *Imagining New Legalities: Privacy and Its Possibilities in the 21st Century* (Stanford, CA: Stanford University Press, 2012) 129–130.

[137] A similar argument is made with relation to DSA systemic risk management in Meßmer and Degeling (n 40) 15.

66 All in all, the typically technocratic and quantitative nature of risk regulation and (predominant) financial understandings of systemic risk may, if applied to EU AI systemic risk management, leave outside of the DSA and AIA's frame several systemic forms of harm that these regulations want to address. It may render invisible more intangible forms of AI harm, obfuscate the political decisions necessarily made in risk management, and neglect the sociotechnical meaning of AI technologies. This would ultimately go against a second normative commitment of the EU's integrated AI risk management framework: that risk assessment and mitigation methodologies should be contextual.

II. Methodologically contextual systemic risk management

67 The DSA and AIA's sociotechnical framing has methodological implications. In particular, the methodologies used in both risk assessment and mitigation must be contextual. This means that any decision on whether and how to assess or mitigate a certain AI risk must consider the meaning and impact of AI technologies on the *social contexts* where those technologies are employed and where their effects are felt. As shown in Section B., both the DSA and AIA require that.[138]

68 The two regulations under analysis do not provide a full answer on how to contextualize risk management. However, drawing from several indications gathered in Section B., one can gain different insights from the DSA and AIA risk management frameworks which may be mutually translatable between them.

69 Firstly, risk assessment and mitigation must consider the societal or sociotechnical contexts where VLOP/SEs operate and, therefore, where the effects of their AI systems are felt.[139] References to 'societal context' should be conceived broadly, so that they encompass effects on society as a whole and broader collective goods,[140] particular situations of societal vulnerability,[141] cultural specificities such as regional and linguistic differences between impacted communities (art. 34(2) DSA and art. 13(1)(a)(i) DRA), as well as the political context of certain communities at given moments in time, such as in the case of a concrete election[142] or coordinated disinformation campaign.[143]

70 Such societal context should influence the choice of risk assessment and mitigation methodologies (art. 9(4)(a) DRA). It should also influence the specific contouring of selected methodologies, in terms of scope, processes of consultation of impacted individuals and groups, and data sampling. Regarding scope, the acute societal impact of platforms' AI systems on a particular issue or community might dictate that issue-specific (as opposed to general) risk management processes be carried out, e.g. for election periods[144] or for child harm online.[145] Still relating to scope, if the AI risks of platforms are not specific to an isolated VLOP/SE but are rather caused by many platforms, then risk assessments must be longitudinal and consider the compounded negative effects of platforms' AI systems on a given societal good.[146]

[138] Recital 79, 90, and art. 34 (2) DSA; European Commission, 2023 (n 9) 10, 13, 15, 63; European Commission, 2024 (n 9) paras. 11-13. In the AIA, this is mainly noticeable regarding the identification and management of high-risk AI systems, see recital 64, 93, and arts. 3(12) and 9(5) AIA. As for systemic risk management in the AIA, although the regulation is somewhat silent regarding risk management methodologies, one can observe the emphasis on its sociotechnical framing and contextual methodologies by looking into the draft of the forthcoming code of practice on systemic risk management of GPAIs: Oliver and others (n 95) 4, 19, 30. A similar focus on context-based risk assessment and mitigation can be found in the AI risk assessment methodology being developed in the Council of Europe, with which the Commission seeks to align the AIA, see Luca Bertuzzi, 'EU Commission Seeks Alignment of AI Treaty's Risk Methodology with AI Act' (*MLex*, 8 November 2024) <https://shorturl.at/N4lWh>. Therein, Bertuzzi reports: "The European Commission wants to ensure that the methodology for risk and impact assessment for AI systems being developed in the Council of Europe is aligned with the EU's AI Act while remaining non-binding. (...) The methodology is based on four building blocks: a context-based risk assessment to collect and map the relevant information, a stakeholder engagement process to contextualize potential harm and risk mitigation measures (...)".

[139] European Commission, 2023 (n 9) 15-16, 24-25; recitals 20, 24 and art. 9(4)(a) DRA; Oliver and others (n 95) 19.

[140] It is useful here to look at the indicated impacted goods of the DSA and AIA risk management framework in art. 34(1)(c) and (d) DSA and art. 3(65) AIA.

[141] Oliver and others (n 95) 19.

[142] European Commission, 2024 (n 9), paras. 31, 36, and 43.

[143] European Commission, 2023 (n 9), 24-25.

[144] *Supra* footnotes 52 and 142.

[145] 'Call for Evidence for an Initiative: Digital Services Act - Guidelines to Enforce the Protection Online' (*European Commission*, 25 September 2024) 2 <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14352-Protection-of-minors-guidelines_en> accessed 25 September 2024; 'Commission Opens Proceedings against TikTok under the DSA' (n 80).

[146] European Commission, 2023 (n 9), 8, 11, 13, 32, 46, 48, 63, 69; European Commission, 2024 (n 9), para. 16(h)(ii); Marsh (n 40) 11–12.

71 Furthermore, the assumptions that platforms, public authorities and auditors make regarding the existence, assessment, and how best to mitigate emerging AI risks must be tested with the groups (and, if applicable, their representative organisations) impacted by AI systems of platforms (recital 90, DSA; art. 13(1)(a)(v) DRA; recital 116 and art. 56(4) AIA[147]). This requires the consultation and involvement of impacted individuals and communities in risk management methodologies, something that the European Commission has already started doing.[148] Finally, and also to achieve an accurate portrayal of the population affected by AI risks, the samples of data to be used in risk assessments and in auditing risk mitigation measures should be representative and, in particular, appropriately depict the concerns of especially affected groups (with particular regard given to minor, vulnerable groups and minorities).[149]

72 The above are just a few non-exhaustive indications found in law and related policy recommendations regarding the selection and contouring of risk management methodologies. The two regulations do not prescribe a single adequate methodology for AI risk management; nor do they answer the question of how to ultimately calculate and assess the risks and impacts of platforms' AI systems. The latter remains an open question to be answered as iterative risk management procedures are developed by private regulated actors and scrutinised by public authorities.[150]

73 This section sought, however, to distil from these methodological indications a common, principle-level, emphasis placed by both the DSA and AIA on the need to socially contextualise AI risk management. Such methodological principle, as well as the sociotechnical frame of systemic risks depicted in Section C.I., are better accommodated by so-called 'social sciences approaches' or 'sociocultural theories' of risk.[151] These perspectives of risk[152] were developed in criticism of the limitations of dominant technical and probabilistic assessments of risk, which are carried out in abstraction from social contexts. Therefore, sociocultural theories of risk sustain that risk assessment and mitigation decisions should, at least in part, consider the *subjective* perceptions of individuals and groups regarding different sources of risk and their potential negative impacts.[153] In that sense, it is arguable that AI risk management methodologies, however they may be concretely tailored, should ensure that individuals and communities are able to articulate their perceptions of AI risks. This links to a third normative commitment of the integrated AI systemic risk management framework under analysis: that risk governance should be participated.

III. Participated systemic risk governance: in comes civil society

74 The DSA and AIA heavily rely on self-regulation by the providers and deployers of AI systems to assess and mitigate relevant emerging risks. As shown in Section B., digital platforms (as AI deployers) and AI providers are the primary decision-makers when it comes to assessing and mitigating emerging AI risks. This means they have the discretion to (i) determine what systemic risks are posed by AI systems in each concrete moment; (ii) which methodologies are used to identify and measure those risks; and (iii) whether and how identified AI risks are mitigated.

75 This discretion afforded to regulated tech companies entails a risk of their lack of accountability. Particularly, those companies may be able to entrench and privilege their own interests in how AI systemic risks are managed. This is supported by existing literature on previous experiences of empowering regulated tech companies to implement and concretise legislative requirements imposed on them.[154]

[147] The AIA prescribes that codes of practice are drawn up with the input of, amongst others, "affected persons". One of the codes of practice to be drawn up in the context of the AIA is the one whereby procedures and measures for systemic risk assessment and management will be agreed upon by several AI providers and deployers.

[148] See, e.g., European Commission, 'Call for Evidence for an Initiative: Digital Services Act - Guidelines to Enforce the Protection Online' (n 145) 4; European Commission, 'AI Act: Have Your Say on Trustworthy General-Purpose AI' (30 July 2024) <https://digital-strategy.ec.europa.eu/en/consultations/ai-act-have-your-say-trustworthy-general-purpose-ai> accessed 19 November 2024, mentioning 'rightsholders'.

[149] Arts. 12 and 13(1)(a)(v) DRA.

[150] On this note, the publication of systemic risk management reports and audits under the DSA is already underway, see Hohfeld (n 66). For an early commentary on published risk assessments, see Sally Broughton Micova, 'Evaluating Systemic Risk Management under the DSA' (*CERRE*, 6 December 2024) <https://cerre.eu/news/evaluating-systemic-risk-management-under-the-dsa/> accessed 6 December 2024.

[151] Ortwin Renn, *Risk Governance: Coping with Uncertainty in a Complex World* (Earthscan 2008) 22–45; Haines (n 20) 184–185.

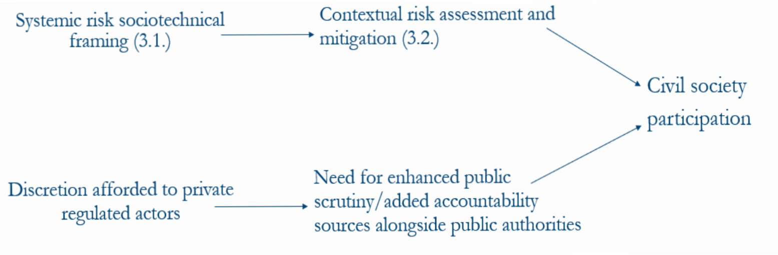
[152] For an overview of the different social sciences approaches to assessing and managing risk, see Renn (n 151) 13–45.

[153] ibid 40–42.

[154] *Supra* footnotes 24 and 79; Kaminski and Malgieri (n 34) 128–129, 133–134, 140; Griffin, 'Rethinking Rights in Social Media Governance' (n 108) 43–51.

76 Applying these insights to the implementation of the EU's AI risk management framework, there is a distinct risk that regulated AI companies gain the dominant power to influence how the meaning of 'AI systemic risk' is shaped from the early stages of DSA and AIA implementation. Indeed, being very broad, the concept of AI systemic risk is open to different interpretations. Equally, systemic risk mitigation can also be done through different measures. As primary decision-makers in the EU's AI risk management provisions, may thus be able to decide how AI systemic risks are defined and mitigated *without* there being appropriate public accountability structures with adequate informational capacity.

77 In order to counter these accountability gaps favouring regulated tech companies, both the EU legislator and several scholars have highlighted the role that civil society actors can have in enhancing the public scrutiny over systemic risk management processes.[155] Civil society is here understood as encompassing not just civil society organisations, but also digital and non-digital NGOs, academic researchers, research institutes, investigative journalists, and fact-checkers.[156] There are two main stated rationales in the DSA and AIA for civil society involvement in risk management processes (Graph 1). First, if systemic risk should be framed in sociotechnical terms and its assessment and mitigation must be contextual, then civil society participation allows for the articulation of (at least some) competing visions of how AI systemic risks should be defined and mitigated. Such risk management decisions should be informed by the experiences and concerns of affected individuals and communities.[157] Second, civil society participation can enhance public scrutiny over platforms' AI risk management choices, thus complementing and feeding into the regulatory supervision of competent European and national public authorities.[158]


Systemic risk sociotechnical framing (3.1.)
Contextual risk assessment and mitigation (3.2.)
Civil society participation
Discretion afforded to private regulated actors
Need for enhanced public scrutiny/added accountability sources alongside public authorities

Graph 1: the rationales for civil society participation in the DSA and AIA

[155] Recitals 40, 90, 92, 95-98, 137 DSA; and recitals 20, 27, 65, 74, 111, 116, 121, 139, 148, 150, 165 and arts. 56, 67, 95(2)(d) and (3) AIA. See also Martin Husovec, 'Will the DSA work?: On money and effort' (Verfassungsblog, 9 November 2022) <https://verfassungsblog.de/dsa-money-effort/> accessed 3 October 2023; Eder (n 24); European Commission, 2024 (n 9) paras. 12, 18, 31-36.

[156] Marsh (n 40) 4; Suzanne Vergnolle, 'Putting Collective Intelligence to the Enforcement of the Digital Services Act: Report on Possible Collaborations between the European Commission and Civil Society Organisations' [2023] SSRN Electronic Journal 12 <https://www.ssrn.com/abstract=4435885> accessed 11 January 2024; Margot E Kaminski and Gianclaudio Malgieri, 'Impacted Stakeholder Participation in AI and Data Governance' (2024) 43–46 <https://papers.ssrn.com/abstract=4836460> accessed 18 September 2024. Although Vergnolle mentions industry groups as part of her operative definition of civil society organisations (which is perfectly conceivable since such groups are indeed called to participate in the implementation of the DSA and AIA, mainly by the Commission), I have left them out of this paper's definition of civil society, since they represent regulated actors, i.e., the tech industry companies on whom legislative requirements are imposed and whose compliance with such requirements public authorities and other civil society actors seek to scrutinise. To include industry representative groups in the definition of civil society in a paper directed at mapping how civil society participation can hold regulated companies to account would, for that reason, be illogical.

[157] Recitals 90, 140 DSA; art. 13(1)(v) DRA; recitals 27, 93, 96 and arts. 56(4) AIA; Oliver and others (n 95) 15.

[158] Recitals 40 and 90 DSA; Recital 1 DRA; Recital 20 AIA; Oliver and others (n 95) 23.

78 The enhancement of public scrutiny enabled by civil society participation can be understood in both quantitative and qualitative terms. Quantitatively, civil society will logically provide more instances of control, resources and data in addition to those of public authorities. Qualitatively, civil society participation may be a conduit for increasing the expertise needed to oversee regulated actors' compliance with their risk management obligations under EU law. In this sense, the DSA and AIA explicitly seek two types of expertise when it comes to AI risk management processes. First, there is a need for technical expertise on the technological capabilities of AI technologies and their impact.[159] In addition to technical expertise, the two regulations also look for first-hand or mediated lived knowledge of the impact of AI systems on those individuals and communities that are particularly affected by them.[160] Kaminski and Malgieri have designated this form of knowledge as 'lived expertise',[161] building on prior scholarly work that argued for the articulation of the lived experiences of affected individuals and communities in participatory schemes of AI governance.[162] The concept of lived expertise is not specific to AI governance. Indeed, prior work in areas such as criminal justice[163] or medical research[164] has analysed in those terms the idea of gathering and using the 'lived' lay knowledge of individuals - as experts on the effects of, e.g., certain laws, policies, institutional practices, social violence, mental health or physiological conditions – in processes of law-making, legal enforcement, institutional reform or highly technical research.[165] In addition, a recent turn in EU legal scholarship has called for the investigation of 'lived experiences' of individuals in order to better understand "the significance, challenges and opportunities" of the implementation of EU law.[166]

79 It must be acknowledged that 'expertise' in EU AI risk management may still be interpreted narrowly, so as to only include technical expertise. That much latitude is offered to the public authorities and regulated actors responsible for setting up participated procedures of AI risk management. Nevertheless, this paper argues that lived expertise is crucial for effectively achieving the normative commitments highlighted in sections C.I. and C.II., namely that systemic risk management processes be contextual and fully grasp the sociotechnical meaning of AI systems . For this to happen, civil society participation must also be about articulating the concerns and lived experiences of individuals and communities regarding the impact of AI systems. Their perceptions of AI systemic risks – even if not concretised in technical jargon – should influence, I argue, how regulators, researchers and platforms understand AI systemic risks and, in turn, shape how those risks are assessed and mitigated.

80 To sum up, both the DSA and AIA call for participated AI systemic risk management, as a way to contextualise those processes and inform public regulatory scrutiny with different forms of knowledge on the impact of AI systems. Although this possibility for civil society participation in risk management processes is explicitly endorsed in the DSA and AIA, the corresponding procedures are unclear. The next section uncovers and systematises them.

[159] E.g., European Commission, 2024 (n 9), para. 18; recital 96 DSA; recitals 111, 151 and art. 68(2) AIA; Husovec (n 155).

[160] E.g., art. 12(2)(f) and 13(1)(v) DRA; recital 20 and art. 56(4) AIA; Oliver and others (n 95) 15. In European Commission, 2024 (n 9) para. 35, we can notice an appeal to VLOP/SEs to engage with not just academics and civil society organisations but also with "representatives of various communities" in order to identify systemic risks that need mitigation in the context of electoral processes and civic discourse. One can imagine that the communities mentioned here will be those that suffer some negative effects that may then contribute to the identification of emerging systemic risks.

[161] Kaminski and Malgieri (n 156) 55.

[162] Ngozi Okidegbe, 'The Democratizing Potential of Algorithms?' (2022) 53 Connecticut Law Review 739, 762–765, 776. Okidegbe's work relates to the use of algorithmic technologies in pre-trial criminal procedures. In this context, she calls 'communal knowledge' to individuals' lived experience of the impact of algorithmic technologies used to determine whether they would be subject to pre-trial incarceration.

[163] Benjamin Levin, 'Criminal Justice Expertise' (2022) 90 Fordham Law Review 2777. Kaminski and Malgieri take Levin's work as inspiration for their idea of 'lived expertise' being feature in AI governance.

[164] Evelyne Baillergeau and Jan Willem Duyvendak, 'Experiential Knowledge as a Resource for Coping with Uncertainty: Evidence and Examples from the Netherlands' (2016) 18 Health, Risk & Society 407; Eva Marie Castro and others, 'Patients' Experiential Knowledge and Expertise in Health Care: A Hybrid Concept Analysis' (2019) 17 Social Theory & Health 307.

[165] Baillergeau and Duyvendak (n 164) 408–410; Levin (n 163) 2821, 2828.

[166] Floris de Witte, 'Here Be Dragons: Legal Geography and EU Law' (2022) 1 European Law Open 113, 116; Loïc Azoulai, 'Reconnecting EU Legal Studies to European Societies' [2024] Verfassungsblog <https://verfassungsblog.de/reconnecting-eu-legal-studies-to-european-societies/> accessed 27 March 2024.

D. The cracks in the law: mapping the loci of civil society participation in EU AI risk management

81 In the legal regime just described, there is no clear and systematised understanding of the modalities of civil society involvement in this regulatory framework. This section fills that gap, by mapping out all possible formal and informal avenues for civil society participation and involvement in EU AI risk governance. One key point is that civil society participation should not be understood here as only encompassing formal ways of public participation in the implementation of the law.[167] While it may include those mechanisms, to fully capture how civil society may attempt to influence platform AI risk regulation, this paper adds more informal avenues of civil society *involvement*. Indeed, legal mobilisation literature has pointed out that civil society actors may strategically opt to influence legal implementation and adjudication through informal means, i.e., those not explicitly recognised in the law as modes of public participation.[168]

82 All the possibilities for civil society participation in the EU's AI risk governance structure – so-called 'loci of participation' - have been mapped in Table 1. It should be added that many of the mapped loci are not designed to enable civil society to intervene *specifically* in the management of platforms' AI risks. Indeed, many DSA-related loci can be used for intervening in the management of risks stemming from other features of digital platforms beyond their AI systems. Similarly, many of the AIA-related loci may be used to influence the management of risks of non-platform-related AI systems.

83 Furthermore, it is expected that this mapping exercise evolves over time, as new stakeholder participation and involvement initiatives surface in this field.

[167] Deirdre Curtin, 'Transparency and Political Participation in EU Governance: A Role for Civil Society?' (1999) 3 Cultural Values 445; Deirdre Curtin and Joana Mendes, 'Transparence et participation: des principes démocratiques pour l'administration de l'union européenne' (2011) 137–138 Revue française d'administration publique 101; Joana Mendes, *Participation in EU Rule-Making: A Rights-Based Approach* (Oxford University Press 2011) <https://academic.oup.com/book/11861> accessed 21 November 2023.

[168] Muir, Dawson and Claes (n 32); Conant and others (n 32).

Table 1: The loci of participation of the integrated EU AI risk management framework applicable to digital platforms

Locus of participation	Legal basis	Civil society actor(s)	Institutional type	Rationale	Expected sought or used expertise
Audits of platforms' risk assessment and mitigation action by vetted researchers (within the context of research supported by privileged access to information in accordance with DSA)	Recitals 96-98 and art. 40(4)(8) DSA; recital 1 DRA	DSA's vetted researchers, who must be affiliated to a research organisation devoted primarily to scientific research (art. 40(8)(11) DSA)	Informal	**Contestation:** conduct research based on privileged access to information that contributes to the assessment and contestation of platforms' risk assessment and mitigation choices. Research output will be made publicly available and put at the service of public authorities and the public at large.	Mostly technical expertise (see emphasis on scientific research), but maybe lived expertise, depending on the specific research project of the researcher in question
Supporting Digital Service Coordinators in data access processes pursuant to art. 40 DSA	Recital 23 and art. 14 Access to Data Delegated Regulation	Expert individuals or organisations with relevant expertise on specific elements of data access process[169]	Formal (organic)	**Enhanced public scrutiny:** support and facilitate the exercise by DSCs of their decisional function regarding the determination of whether and how much data should be shared by VLOP/SEs with vetted researchers pursuant to art. 40 DSA.	Technical expertise
Commission audits with the involvement of individual experts	Art. 69(3) and (5), and 72(2) DSA; Para. 53, DSA risk mitigation guidelines; recital 3 and art. 3(5)-(7) Implementing Regulation 2023/1021	Individual experts (can be vetted researchers) invited by the Commission to help support the platform audits it carries as part of its monitoring powers[170]	Formal (organic)	**Enhanced public scrutiny (and maybe explicit contestation):** support the performance of Commission-led audits to assess platforms' risk assessment and mitigation choices. If the Commission finds shortcomings and pursues corrective action, contestation will be involved.	Technical expertise

[169] Some exemples given in recital 23 of the Access to Data Delegated Regulation are 'the determination of the access modalities, including appropriate interfaces, the formulation of the reasoned request [for data access] and any amendment requests [to the researcher's reasoned request] by the data provider'.

[170] See in 'Commission Sends Preliminary Findings to X for Breach of DSA' (*European Commission*, 12 July 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761>: "Based on an in-depth investigation that included, among others, the analysis of internal company documents, *interviews with_experts* [N.B. emphasis added by author], as well as cooperation with national Digital Services Coordinators (...)". Similarly, but mentioning "third parties" and not "experts", see 'Commission Opens Formal Proceedings against Temu under DSA' (n 81).

Submission of evidence to Commission and/or national regulators	Recital 141 and art. 51(1)(a), 67(1), 68(1), 72 DSA; art. 79(7), 90(3)(c) AIA (although not specifically, they signal the Commission's normal openness to receiving evidence from interested parties)	Individual experts and all interested/impacted stakeholders; civil society organisations	Informal	**Enhanced public scrutiny:** provide evidence to public supervisory authorities about emerging AI risks, their societal impact, and how platforms and other AI deployers and providers contribute to, assess and manage those risks.	Technical and/or lived expertise (type of expertise might be limited by the willingness of the Commission or the national regulator to receive certain types of evidence)
Participation in public consultations[171] and calls for evidence[172] for the development of guidelines or elaboration of risk assessments[173]	Recital 103 and art. 35(3), 39(3), 63(1)(e) DSA; DSA risk mitigation guidelines; Art. 96 AIA *juncto* arts. 3(2)(c) and 4(1)(b) AI Office Decision	Individual experts and all interested/impacted stakeholders; civil society organisations	Formal (procedural)	**Inclusiveness/legitimacy-building + enhanced public scrutiny:** Commission seeks to hear experts and impacted individuals and communities in the preparation of guidelines. This is aimed at combatting the opaqueness of guideline development and ensuring stakeholder representation; and have stakeholders provide additional information and give their opinions on what are relevant risks, what methodologies or metrics for risk assessment should be considered, and what risk mitigation best practices should be contained in the guidelines.	Lived[174] and technical expertise

[171] See, e.g., European Commission, 2024 (n 9); European Commission, 'Multi-Stakeholder Consultation for Commission Guidelines on the Application of the Definition of an AI System and the Prohibited AI Practices Established in the AI Act' (13 November 2024) <https://digital-strategy.ec.europa.eu/en/news/commission-launches-consultation-ai-act-prohibitions-and-ai-system-definition>.

[172] See, e.g., European Commission, 'Call for Evidence for an Initiative: Digital Services Act - Guidelines to Enforce the Protection Online' (n 145).

[173] Vergnolle (n 156) 44.

[174] European Commission, 'Call for Evidence for an Initiative: Digital Services Act - Guidelines to Enforce the Protection Online' (n 145) 3–4.

Collaborative development of codes of conduct and codes of practice[175]	Recitals 98, 103, 107 and art. 45(2) DSA; recital 27, 165 and arts. 56(3)(4) and 95(2)(d) and (3) AIA; arts. 3(2)(i) and 4(1)(b) AI Office Decision	Individual experts and all interested/impacted stakeholders; civil society organisations[176]	Formal (procedural)	**Inclusiveness/legitimacy-building**: hear experts and impacted individuals and communities and have them contribute to the drawing up of codes of conduct.	Lived and technical expertise
Other outreach initiatives directed at civil society and researchers (such as hackathons, stress tests or other crowdsourced events)[177]	No legal basis and unclear form, but that possibility has been mentioned by the Commission[178] in the context of the DSA and, theoretically, nothing excludes that it puts together these initiatives also in the context of the AIA	Individual researchers and civil society organisations	Formal (procedural)	**Inclusiveness/legitimacy-building + enhanced public scrutiny:** this will very much depend on the objective and format of every outreach initiative but, in general, it is assumed that this pursues Commission's double objective of obtaining, centralising and analysing additional data on a regulatory matter of interest. This also increases the representative credentials of the Commission's enforcement action.	Lived and technical expertise (depending on the specific outreach initiative)

[175] E.g., European Commission, 'First Draft of the General-Purpose AI Code of Practice Published, Written by Independent Experts' (2024) <https://digital-strategy.ec.europa.eu/en/library/first-draft-general-purpose-ai-code-practice-published-written-independent-experts> accessed 19 November 2024.

[176] In cases like that of the General-Purpose AI Code of Practice referenced in the footnote above, the involvement of stakeholders is done in accordance with a layered process, where there are multiple opportunities (with different levels for civil society access) to offer inputs to the development of a code of practice. In this case, some independent experts chosen by the Commission elaborated a first draft with contributions from general-purpose AI providers which then was submitted to an open multi-stakeholder consultation during two months. In parallel, 1000 stakeholders (civil society organisations, researchers, business groups, and others – there is no clarity regarding all types of represented actors) were selected by the Commission based on an open call for applications. The selected 1000 stakeholders will meet with the drafters of the code of practice in 3 iterative rounds, with the code being amended based on stakeholder input. For this, see European Commission, 'AI Act: Participate in the Drawing-up of the First General-Purpose AI Code of Practice' (30 July 2024) <https://digital-strategy.ec.europa.eu/en/news/ai-act-participate-drawing-first-general-purpose-ai-code-practice> accessed 19 November 2024; European Commission, 'Meet the Chairs Leading the Development of the First General-Purpose AI Code of Practice' (30 September 2024) <https://digital-strategy.ec.europa.eu/en/news/meet-chairs-leading-development-first-general-purpose-ai-code-practice>; 'The Kick-off Plenary for the General-Purpose AI Code of Practice Took Place Online' (30 September 2024) <https://digital-strategy.ec.europa.eu/en/news/kick-plenary-general-purpose-ai-code-practice-took-place-online> accessed 19 November 2024.

[177] Vergnolle (n 156) 47; The European Board for Digital Services, 'Report on the European Elections: Digital Services Act and Code of Practice on Disinformation' (2024), p. 5 <https://digital-strategy.ec.europa.eu/en/library/european-board-digital-services-publishes-post-election-report-eu-elections>. For some examples of DSA and AIA-relevant outreach events to researchers, see European Commission, 'Info Webinar for Researchers: DSA Art 40 Delegated Act' (*EUSurvey*, 2024) <https://ec.europa.eu/eusurvey/runner/DataAccessInfoWebinar> accessed 18 November 2024; European Commission, 'Call for Evaluators: Participate in the European AI Office Workshop on General-Purpose AI Models and Systemic Risks' (25 November 2024) <https://digital-strategy.ec.europa.eu/en/news/call-evaluators-participate-european-ai-office-workshop-general-purpose-ai-models-and-systemic> accessed 12 December 2024.

[178] European Commission, 'Call for Evidence for an Initiative: Digital Services Act - Guidelines to Enforce the Protection Online' (n 145) 3.

Involvement in platform audits at the invitation of VLOPs and VLOSEs	Art. 37 DSA *juncto* arts. 12 and 13(1)(a)(v) DRA	Impacted individuals and communities, especially those deemed as most vulnerable	Regulatee-promoted	**Inclusiveness/legitimacy-building + Participatory design:** to test platforms' assumptions regarding the effectiveness of their risk management choices with impacted stakeholders.	Lived and technical expertise
Participation in Commission-led processes for updating delegated acts that update risk lists of the AIA	Recital 173 and Art. 112(11) AIA; arts. 3(2)(a) and 4(1)(b) AI Office Decision	Individual experts and all interested/impacted stakeholders; civil society organisations	Formal (procedural)	**Inclusiveness/legitimacy-building:** Commission hears experts and impacted individuals and communities in the preparation of updating the delegated acts that determine which AI systems present high-risk, unacceptable risk, and systemic risk; combat opaqueness of delegated act adoption.	Lived and technical expertise
Participation in hearings of EP that may inform potential objections to how the Commission has exercised its delegated power to update the risk lists of the AIA	Linked to arts. 6 and 7 AIA. No explicit basis in AIA, but linked to parliamentary practice of the European Parliament[179]	Individual experts and impacted stakeholders who are invited by the EP or the Council to be heard on a specific AI risk; civil society organisations	Formal (procedural)	**Enhanced public scrutiny:** provide information on the risks of specific AI systems whose risk status under the AIA was changed through a Commission delegated act. This additional information informs the scrutiny exercised by the Parliament over the Commission's decision to adopt such a delegated act.	Lived and/or technical expertise (depending on Parliament's request)
Involvement in red teaming exercises and other adversarial testing	Para. 27, d), DSA risk mitigation guidelines; recital 60q and art. 55(1)(a) AIA; Commission press release on election Stress Test[180]	Individual experts invited by VLOPs and VLOSEs as well as GPAI providers	Regulatee-promoted	**Enhanced public scrutiny:** have independent experts test genAI and other GPAIs for bias and other risk sources by seeking to game/exploit their design and other vulnerabilities.	Technical expertise
Ad hoc cooperation projects, expert consultations, and constructive dialogues between platforms and civil society	Paras. 18, 23, 31-36, DSA risk mitigation guidelines	Individual experts, research community and all relevant stakeholders	Regulatee-promoted	**Enhanced public scrutiny:** obtain extensive feedback and additional insights on risk management policies and actions.	Technical expertise

[179] See, e.g., Amandine Crespy and Louisa Parks, 'The European Parliament and Civil Society' in Olivier Costa (ed), *The European Parliament in Times of EU Crisis: Dynamics and Transformations* (Springer International Publishing 2019); Laura Landorff, 'Who Gets a Seat at the Table? Civil Society Incumbents and Challengers in the European Parliament's Consultations' in Håkan Johansson and Anna Meeuwisse (eds), *Civil Society Elites: Exploring the Composition, Reproduction, Integration, and Contestation of Civil Society Actors at the Top* (Springer International Publishing 2024).

[180] This is a very recent example of collaborative (i.e., Commission-promoted) testing of how VLOPs and VLOSEs mitigate specific risks according to the DSA. See 'Commission stress tests platforms' election readiness under the Digital Services Act' (*European Commission*, 24 April 2024), available at: <https://shorturl.at/cdmT8>. Another example of red teaming exercises potentially relevant for AIA enforcement can be found here: Will Douglas Heaven, 'How OpenAI Stress-Tests Its Large Language Models' (*MIT Technology Review*, 21 November 2024) <https://shorturl.at/dmqOp> accessed 10 January 2025.

Cooperation between platforms and independent fact-checking organisations[181]	Paras. 12-14, 16(c), 36, 51, DSA risk mitigation guidelines	Independent fact-checking organisations (e.g. the European Digital Media Observatory[182]) and journalists	Regulatee-promoted	**Enhanced public scrutiny:** capacity-building for adopting risk mitigation measures applied by platforms to manage systemic risks to electoral processes and civic discourse, namely by helping to flag false/deceptive AI-promoted and/or generated content.	Technical expertise
Issuance of qualified alerts of systemic risks of GPAIs	Arts. 51(1)(b), 68 and 90 AIA	Appointed/invited independent experts of scientific panel created by Commission in governance structure of AIA	Formal (organic)	**Enhanced public scrutiny (maybe through explicit contestation):** to provide a qualified alert[183] to the AI Office[184] flagging that a GPAI presents a systemic risk that needs to be managed at Union level (this will entail contestation if the GPAI provider has stated that the model does not present a systemic risk); based on this qualified alert, the Commission will designate the GPAI as presenting systemic risk, triggering a series of risk management obligations for the GPAI provider.	Technical expertise
Membership of scientific panel or advisory forum created in the governance framework of AIA	Recitals 148, 150 and 151 and arts. 67 and 68 AIA	Appointed/invited individual experts (for scientific panel); and also, for advisory forum, of civil society organisations and other interested/ impacted stakeholders with recognised expertise in the field of AI	Formal (organic)	**Enhanced public scrutiny:** to support the Commission (including its AI Office) and the AI Board in their implementation tasks under the AIA.	Mainly technical, but possibly also lived expertise for the advisory forum[185]

[181] See, e.g., The European Board for Digital Services (n 177) 6.

[182] ibid.

[183] Relying on privileged access to information on GPAIs based on art. 91(3) AIA.

[184] The AI Office is an internal division of the Commission entrusted with overseeing advancements in AI development, as well as the enforcement and monitoring of the AIA. See art. 55b AIA and AI Office Decision for more details.

[185] This will ultimately depend on the interpretation by institutional actors of the concept of 'expertise' in art. 58a AIA.

Membership of DSA expert groups at European or national level[186]	Art. 64 DSA	Individual experts or members of civil society organisations with expertise in platform regulation (and related AI matters) which are invited to join expert groups set up by the Commission or by the DSA national regulators of each Member State (Digital Services Coordinators)	Formal (organic)	**Enhanced public scrutiny:** to support the Commission and Digital Services Coordinators in their supervision tasks under the DSA, including overseeing risk management processes.	Technical expertise
Invitation to attend meetings or to be consulted by the European Board for Digital Services (EBDS)	Art. 62(5) and (6) DSA	Individual experts or interested stakeholders who are invited to attend/ observe the meetings of EBDS, or that are consulted by it at its own initiative	Formal (organic)	**Enhanced public scrutiny:** to support the EBDS in its meetings and the fulfilment of its advisory and coordination tasks.	Technical expertise
Participation (and possible contestation) in processes of development of harmonised standards + implementing acts for Commission to adopt common specifications *in lieu* of harmonised standards	Recital 121 and art. 40(2) AIA; art. 3(2)(d) and 4(1)(b) AI Office Decision; art. 5 of Regulation 1025/2012	All interested/impacted stakeholders; civil society organisations; researchers	Formal (organic or procedural, depending on the modalities of stakeholder inclusion)	**Enhanced public scrutiny (maybe through explicit contestation) + Inclusiveness/legitimacy-building:** Standardisation organisations (for the development of technical standards) and Commission (for development of request for the production of standards or development of common specifications[187]) hears experts and other interested/ impacted stakeholders (including through advisory forum) in the preparation of European harmonised standards and other related documents; ensure inclusiveness/legitimacy of standardisation processes.[188]	Technical and lived expertise

[186] This is a concrete proposal of civil society involvement made by Suzanne Vergnolle to concretise the mandate of the Commission and national regulators of creating the necessary expertise and capabilities to oversee DSA compliance. See Vergnolle (n 156) 23–43, 50–51.

[187] Per art. 41 AIA, the Commission will adopt common specifications in case harmonised standards have not been, are deemed insufficiently protective of fundamental rights concerns, or otherwise do not comply with the corresponding Commission's request.

[188] For general theory on civil society participation in technical standardization processes, see, inter alia, Annalisa Volpato and Mariolina Eliantonio, 'The Participation of Civil Society in ETSI from the Perspective of Throughput Legitimacy' (2024) Innovation: The European Journal of Social Science Research 1.

Participation in regulatory sandboxes	Arts. 53(16)(17) and 58(2)(f) AIA; arts. 3(2)(e) and 4(1)(b) AI Office Decision	Individual researchers and all interested/impacted stakeholders	Formal (organic or procedural, depending on the modalities of stakeholder inclusion)	**Technological participatory design:** involvement in the testing and, accordingly, in the design and/or structural adaptation of AI models/systems under development in the sandbox.	Technical expertise
Collaboration and dialogue between members of civil society, public authorities, and/or private regulated actors in conferences, workshops or other roundtable events[189]	None (but relevant for both DSA and AIA)	Individual researchers and members of civil society organisations/NGOs (conceivably also journalists and fact-checkers, although less likely in academic settings)	Informal	**Enhanced public scrutiny:** members of public authorities seek additional information and insights over their area of regulation. Researchers and members of other civil society organisations seek to both gain new insights from their colleagues and members of public institutions about their research interests, but also to shape regulatory dialogue through the dissemination of their research findings.	Technical expertise (and, possibly, second-hand dissemination of lived expertise)
Innovation-fostering initiatives promoted by Commission and/or private sector[190]	None (but relevant for both DSA and AIA)	Research community	Informal	**Enhanced public scrutiny+ Technological participatory design:** have researchers contribute to regulatory dialogue with new ideas for how to design (i) technical solutions for the adaptation of AI models and systems to legal requirements; or (ii) ways for regulated actors to navigate compliance with legal requirements.	Technical expertise

[189] Vergnolle (n 156) 18, 21; Marsh (n 40) 4. For concrete examples, see 'The DSA and Platform Regulation Conference 2024' (*DSA Observatory*, 11 December 2023) <https://dsa-observatory.eu/the-dsa-and-platform-regulation-conference-2024/> accessed 5 November 2024; Pielemeier and Sullivan (n 40); European Commission, 'Commission Gathers Good Practices to Combat Online Harm for Minors' (7 October 2024) <https://digital-strategy.ec.europa.eu/en/news/commission-gathers-good-practices-combat-online-harm-minors> accessed 18 December 2024; John Albert, 'DSA Risk Assessment Reports: A Guide to the First Rollout and What's next' (9 December 2024) <https://dsa-observatory.eu/2024/12/09/dsa-risk-assessment-reports-are-in-a-guide-to-the-first-rollout-and-whats-next/> accessed 13 December 2024.

[190] Vergnolle (n 156) 18; See, e.g., 'Call for Participation to the Innovation Challenge "AI Act Compass: Navigating Requirements for High-Risk AI Systems"' (*Legality Attentive Data Scientists (LeADS)*, 1 August 2024) <https://www.legalityattentivedatascientists.eu/2024/08/01/innovation-challenge-call-for-participation/> accessed 5 November 2024; 'EU Boosts European AI Developers' (*European Commission*, 10 September 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4621> accessed 9 October 2024.

Publication of policy reports,[191] academic articles/books,[192] press releases[193]	None (but relevant for both DSA and AIA)	Civil society organisations or individual researchers	Informal	**Enhanced public scrutiny + explicit contestation:** contributing to or influencing the regulatory dialogue on the implementation of the DSA/AIA by providing new evidence or perspectives regarding emerging AI risks and how those are being/should be assessed and mitigated. These publications can be highly persuasive and be considered or taken up as evidence by both public authorities (Commission[194] or national regulators) or by platforms.	Lived and technical expertise (depending on publication's content)
Online activism and journalistic work	None (but relevant for both DSA and AIA)	Interested/impacted individuals, NGOs and journalists	Informal	**Explicit contestation (with possible enhancement of public scrutiny:** contributing to or influencing the regulatory dialogue on the implementation of the DSA/AIA by providing new evidence or perspectives regarding emerging AI risks and/or potential non-compliance with risk management obligations. Whereas journalistic work will have less of a contestatory tone, activism posts will often entail explicit contestation and thus articulate competing visions of overlooked AI risks and/or non-compliance with the law. These publications can be used by supervisory or parliamentary authorities, thus broadening their informational resources.[195]	Lived and technical expertise (often not formulated in technical terms, but rather highlighting the lived impact of AI technologies and social media algorithms)

[191] Meßmer and Degeling (n 40); Broughton Micova and Calef (n 119).

[192] E.g., Claudio Novelli and others, 'How to Evaluate the Risks of Artificial Intelligence: A Proportionality-Based, Risk Model for the AI Act' (2023) <https://papers.ssrn.com/abstract=4464783> accessed 9 November 2023.

[193] E.g., 'European Digital Rights and Others, An EU Artificial Intelligence Act for Fundamental Rights: A Civil Society Statement' (30 November 2021) <https://edri.org/wp-content/uploads/2021/12/Political-statement-on-AI-Act.pdf>; Article 19 and others, 'Civil Society Open Letter to Commissioner Breton' (17 October 2023) <https://www.article19.org/wp-content/uploads/2023/10/Civil-society-open-letter-to-Commissioner-Breton.pdf>; Access Now, ARTICLE 19, and Electronic Frontier Foundation (EFF), 'Commissioner Breton: Stop Politicising the Digital Services Act' (*Access Now*, 19 August 2024) <https://www.accessnow.org/press-release/commissioner-breton-stop-politicising-the-digital-services-act/>.

[194] Francesco Duina, 'Is Academic Research Useful to EU Officials? The Logic of Institutional Openness in the Commission' (2022) 29 Journal of European Public Policy 1493.

[195] See e.g. Samira Rafaela, 'Parliamentary Question, The Practice of Shadow-Banning Content on Social Media Platforms, E-003111/2023' (European Parliament 2023) <https://www.europarl.europa.eu/doceo/document/E-9-2023-003111_EN.html> where a member of the European Parliament made a question to the Commission based on, amongst others, investigative journalistic work; Sydney Bauer, 'Elon Musk Has Made Anti-Trans Hatred One of Twitter's Core Features' (*The Nation*, 23 June 2023) <https://www.thenation.com/article/society/elon-musk-transphobia-twitter/>.

Complaints to national authorities or Commission	Arts. 53 and 86 DSA; art. 85 AIA for administrative complaints. Specifically for the DSA there are other ways of presenting complaints and flagging illegal content moderation practices through out-of-court dispute settlement procedure (art. 21 DSA) or submitting notices to platforms by way of trusted flagger status (art. 22 DSA)	Interested/impacted individuals or civil society organisations/NGOs representing them	Informal (even though it involves the leveraging of formal procedures, this is not seen formally in the law as a mode of public participation)	**Explicit contestation + Defence function:** offering a competing vision of how private regulated actors have identified or mitigated a potential AI systemic risk. Complaints can be a powerful source of information for regulators,[196] and may escalate all the way up to judicial litigation leading to the establishment of substantive standards on a certain regulatory matter.[197]	Lived and technical expertise
Strategic challenge of Commission's transparency policy regarding its enforcement actions[198]	Recital 13 and art. 8(1) and (3) of Regulation 1049/2001(relevant for both DSA and AIA)	Interested/impacted individuals, NGOs, researchers, and journalists	Informal (even though it involves the leveraging of formal procedures, this is not seen formally in the law as a mode of public participation)	**Explicit contestation + defence function:** to challenge the Commission's information policies and seek bigger transparency regarding how they monitor compliance with DSA and AIA risk management obligations. This often takes place in the form of a challenge to the European Ombudsman of a Commission's individual decision to deny access to information to a certain individual.	Technical expertise

[196] Vergnolle (n 156) 21, 45–46.

[197] For an example of such escalation in a field of EU digital regulation, i.e., the General Data Protection Regulation, see *Schrems I*, Court of Justice of the European Union, C-362/14, Judgment of the Court (Grand Chamber) of 6 October 2015, ECLI:EU:C:2015:650, paras. 26-36; *Schrems II*, Court of Justice of the European Union, Judgment of the Court (Grand Chamber) of 16 July 2020, ECLI:EU:C:2020:559, paras. 50-68; James Jacoby, 'The Facebook Dilemma - Interview with Max Schrems, a Privacy Advocate' (*FRONTLINE*, 2018) https://www.pbs.org/wgbh/frontline/interview/max-schrems/; Giovanni De Gregorio, 'The Rise of Digital Constitutionalism in the European Union' (2021) 19 International Journal of Constitutional Law 41, 54–55.

[198] Curtin (n 167) 460, suggesting that an assertive approach challenging EU institutions' information policies may be viewed through a participatory lens.

Engaging in and/or supporting strategic litigation	None (but relevant for both DSA and AIA)	Affected individuals, communities and/or legal persons; NGOs representing or supporting them[199]	Informal (even though it involves the leveraging of formal procedures, this is not seen formally in the law as a mode of public participation)	**Explicit contestation + defence function:** offering a competing vision of how a private regulated actor have identified or mitigated a potential AI systemic risk; and seeking to obtain from courts the establishment of substantive standards that protect the litigant's interests (and those of individuals in a similar position). Strategic litigation may be done at the national level and escalate all the way up to the CJEU.[200]	Lived and technical expertise

84 The mapped loci of participation come in many shapes and sizes. To begin with, they empower different civil society actors and provide a space, in varying degrees, for the articulation of different types of expertise (lived and/or technical), as was already discussed in Section C.III. Furthermore, institutionally speaking, they might be more[201] (i) formal, (ii) informal, or (iii) regulatee-promoted. They will be:

- formal, if they are explicitly mentioned in law as modes of public participation in the implementation of these regulations. Drawing from Mendes, formal participation may be organic, if participating actors are included in the institutional structures where participation takes place; or procedural, if participants remain outside institutional structures and are determined based on the

[199] In Germany, the NGO Gesellschaft für Freiheitsrechte (Society for Civil Rights, GFF) is already supporting individual actors in DSA strategic litigation at the national level Jürgen Bering Vezzoso Simonetta, 'Meta's Fundamental Rights Blunder - And a Happy German Antitrust Fix' (*Tech Policy Press*, 6 August 2024) <https://techpolicy.press/metas-fundamental-digital-rights-blunder-and-a-german-antitrust-fix> accessed 14 January 2025.

[200] Alberto Alemanno, 'Beyond EU Law Heroes: Unleashing Strategic Litigation as a Form of Participation in the Union's Democratic Life' (2025) <https://shorturl.at/nw891> accessed 8 November 2024.

[201] This is a taxonomy offering a heuristic model to interpret the different institutional setups of loci of participation. As with any other taxonomy, it has limits as it reduces the observed complexity in this field. It is, however, important to note that the distinction between formal, informal, and regulatee-promoted loci of participation is one of degree: the institutional structures of different loci of participation might present more or less features of each type. Therefore, the institutional type of each locus of participation indicated in Table 1 is the predominant one in how each locus is structured. I adopt the same taxonomical approach as Simon Halliday, 'After Hegemony: The Varieties of Legal Consciousness Research' (2019) 28 Social & Legal Studies 859, 861. He states: "The sketch of these [categories] should be interpreted lightly. I am not suggesting, for example, that there is no overlap or dialogue between them. Rather, they are presented in the manner of Weberian ideal types - 'exaggerated or one-sided depictions that emphasise particular aspects of what is obviously a richer and more complicated reality' (...). The sketch is thus intended merely as an analytical device, (...)".

subject-matter of the procedure or process where they intervene.[202] Participation in public consultations or calls for evidence launched by the Commission are examples of procedural participation, whereas the invitation of experts to be part of the AIA advisory forum or DSA Commission audits of VLOP/SEs constitute examples of organic participation;

- informal, if they are not explicitly mentioned in law as loci of civil society participation but may nevertheless be used to attempt to influence AI risk management processes – for example, online activism or the publication of policy reports. Informal loci of participation may also entail the leveraging of other public procedures or legal provisions not primarily designed to enable civil society participation. Examples of this are the presentation of complaints by interested individuals or organisations to regulators about potential non-compliance with risk management provisions,[203] or the use of access to information provisions by researchers to conduct their own audits of how platforms have (or have not) identified and mitigated emerging systemic AI risks;[204]
- regulatee-promoted, if participation occurs within an institutional framework set up by private regulated actors, e.g., the participation of researchers in red teaming exercises organised by platforms or other AI providers and deployers; or the cooperation between VLOP/SEs and fact-checking organisations in the context of mitigating systemic risks of AI-generated or algorithmically-spread disinformation.

85 At the same time, the mapped loci of participation pursue different underlying rationales. This is of great importance since, as explained by Kaminski and Malgieri, the theoretical explanations behind civil society participation "lead to calls for different kinds of interventions by civil society".[205] By that same token, the underlying rationale pursued, in theory, by a certain locus of participation will necessarily shape what civil society may achieve therein. Drawing from the work of Kaminski and Malgieri – who have sought to disentangle different theoretical explanations for stakeholder participation in AI governance[206] – and from other legal and political science literature on how third parties intervene in complex regulatory arrangements, I have developed a taxonomy of five rationales for civil society participation in the EU's AI risk governance framework[207]: (i) inclusiveness and legitimacy-building; (ii) technological participatory design; (iii) enhanced public scrutiny; (iv) explicit contestation; and (v) defence function. The following paragraphs describe each rationale in detail.

86 Firstly, when participation aims to promote inclusiveness and legitimacy-building, that means ensuring the representation of all relevant stakeholders in procedures of legal implementation. This rationale derives from theories of democratic representation.[208] From the perspective of the represented persons, participation is seen as a tool to make legal implementation processes less opaque, actively involve civil society actoirs, and thus reconcile the bureaucratic domination of public authorities over those processes with democratic values.[209] In this sense, participation does not mainly concern the substantive results of legal implementation processes (contrarily to the rationales below), but rather how inclusive and open those *processes* are.[210] However, democratic rationales of participation also allow for claims regarding the legitimacy and goodness of law-making and legal implementation. In particular, participation in the democratic sense may be used for legitimacy-building purposes, inasmuch as the involvement of civil society allows institutions to claim their own legitimacy, as well as the legitimacy and goodness of the output of the procedures where civil society actors were involved.[211]

[202] For further elaboration on this distinction, which is beyond the scope of this paper, see Mendes (n 167) 30–31.

[203] See, e.g., 'Commission Sends Request for Information to LinkedIn on Potentially Targeted Advertising Based on Sensitive Data under Digital Services Act' (14 March 2024) <https://digital-strategy.ec.europa.eu/en/news/commission-sends-request-information-linkedin-potentially-targeted-advertising-based-sensitive-data> accessed 25 September 2024. Here, the Comission states that "This enforcement action is based on a complaint submitted to the Commission by civil society organisations".

[204] See Section B.I.3.(c).

[205] Kaminski and Malgieri (n 156) 22.

[206] ibid 22–42.

[207] The same considerations made in footnote 201 apply to this taxonomy as well. Furthermore, this taxonomy constitutes a tentative exercise that is expected to evolve with an empirical inquiry of the purposes of civil society participation in the DSA and AIA's risk management frameworks.

[208] Curtin (n 167) 455–457; Kaminski and Malgieri (n 156) 22–25.

[209] Curtin (n 167) 445–446, 461; Gloria Golmohammadi, 'Realizing the Principle of Participatory Democracy in the EU: The Role of Law-Making Consultation' (Stockholm University 2023) 88–89 <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-220289> accessed 21 November 2023.

[210] See, to this effect, Kaminski and Malgieri (n 156) 22–24.

[211] Mendes (n 167) 91, 94, 126, 129; Danai Petropoulou Ionescu, 'Habemus Legitimacy? The European Commission Opens Public Consultation for a Guidance Document' (2021) 12 European Journal of Risk Regulation 861.

87 The second identified rationale is technological participatory design. In this case, civil society participation is part of a distinct methodological approach to computational design.[212] It does not aim to ensure stakeholder representation (contrary to the inclusiveness and legitimacy-building rationale), nor to assess and potentially contest choices that regulatees have already made. Differently, it aims to integrate stakeholder interests into technological design. Importantly, it does not seek to address and represent all civil society perspectives, but only those that may be integrated into technological design in a resource-efficient way.[213]

88 Thirdly, civil society participation may serve to enhance public scrutiny over regulated actors, thus adding to public regulatory capacity. Civil society actors may do so in two, distinct ways: they may facilitate the exercise of public authorities' functions; or, alternatively, act as surrogate regulators. Starting with the former case, civil society participation may be a source of new factual and/or technical information for administrative authorities, thereby facilitating the exercise of their supervisory and decisional functions on any given regulatory area.[214] In addition, civil society actors may also act as surrogate regulators, providing an added level of scrutiny over both (i) regulated actors about how they comply with the law and (ii) public authorities for how they enforce it. This function of civil society participation is highlighted by theories of tripartism. According to them, public accountability is not ensured through a top-down relationship between States and regulated actors but, instead, in a tripartite scheme where civil society participates in regulatory enforcement next to the State and regulated actors.[215] In general, enhanced public scrutiny may take a lot of forms, such as collaboration and dialogue with public authorities and regulated actors, research (potentially disseminated), and assessment.[216]

89 Fourthly, civil society participation may serve a defence function. In this sense, the intervention of persons in administrative procedures is aimed at allowing them to defend their subjective rights or legally relevant interests potentially affected by administrative decisions. In this way, participation enables administrative authorities to account for the interests of persons potentially affected by administrative action, serving as an ex-ante complement to judicial review.[217]

90 Finally, as a fifth rationale for civil society participation, civil society actors may intend to *explicitly* contest how a specific piece of law is being implemented and, relatedly, challenge and politicise the dominant regulatory arrangements of each given time, often portrayed as a form of technical or apolitical consensus between divergent interests.[218] Contestation is here defined as the use of a locus of participation by civil society actors to articulate competing visions[219] of (i) what are AI systemic risks, (ii) what concrete risks emerge over time; and (iii) alternatives ways of assessing and mitigating those risks. In this sense, contestation is not just concerned with articulating the systemic impacts of platforms' AI systems. It also contains an ambition to change the risk management choices made by private actors and overseen by public authorities. In this sense, contestation may have several (potentially simultaneous) objects. This is to say, that it may seek to challenge (i) concrete compliance decisions made by public authorities or private regulated actors through administrative or judicial means;[220] (ii) general institutional policies that impact regulatory enforcement by public authorities or compliance by private actors;[221] or even the regulatory agenda, meaning the regulatory issues that gain the attention of the members of the public and government officials, thereby becoming priorities in the policy and enforcement debate.[222]

[212] Kaminski and Malgieri (n 156) 32–34; Ned Cooper and Alexandra Zafiroglu, 'From Fitting Participation to Forging Relationships: The Art of Participatory ML', *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery 2024) <https://dl.acm.org/doi/10.1145/3613904.3642775> accessed 30 January 2025.

[213] Kaminski and Malgieri (n 156) 32–34.

[214] Curtin (n 167) 459; Mendes (n 167) 32.

[215] Ian Ayres and John Braithwaite, 'Tripartism: Regulatory Capture and Empowerment' (1991) 16 Law & Social Inquiry 435, 439, 441–445; Darren Sinclair and Neil Gunningham, 'Smart Regulation' [2017] Regulatory theory: Foundations and applications 133, 133.

[216] Sinclair and Gunningham (n 215) 135–138.

[217] Mendes (n 167) 33.

[218] Curtin (n 167) 467; Chantal Mouffe, *The Democratic Paradox* (Verso 2000); Eugen Octav Popa, Vincent Blok and Renate Wesselink, 'An Agonistic Approach to Technological Conflict' (2021) 34 Philosophy & Technology 717; Daniel E Walters, 'The Administrative Agon: A Democratic Theory for a Conflictual Regulatory State' (2022) 132 Yale Law Journal 1, especially 48-57.

[219] Competing visions are here defined as those counter-hegemonic approaches to how a particular regulatory arrangement is shaped at any given time. See Paulina Tambakaki, 'The Tasks of Agonism and Agonism to the Task: Introducing "Chantal Mouffe: Agonism and the Politics of Passion"' (2014) 20 Parallax 1, 7–10.

[220] See, in Table 1, the possibility to lodge administrative complaints with national supervisory authorities or the Commission, as well as strategic litigation.

[221] Curtin (n 167) 460; Walters (n 218) 56.

[222] Contestation by civil society in this last case is aimed at 'agenda-setting', see Thomas A Birkland, 'Agenda Setting in Public Policy' in Frank Fischer and Gerald Miller (eds), *Handbook of public policy analysis: theory, politics, and methods*

91 A few final remarks are needed to fully explain how contestation as a rationale for civil society participation may be present in the DSA and AIA. Specifically, I argue that contestation-driven civil society participation may prove particularly crucial in the context of AI systemic risk management. If, as seen in Sections 2. and C.I., the definition and mitigation of AI systemic risks are open to different interpretations, and if regulated actors will concretise them in time, then their determinations of what AI systemic risks are and how they should be managed are contingent. And, if they are contingent, they are also open to contestation and concomitant change.[223] In this sense, contingency and contestation of regulatory arrangements are intimately tied: in Mouffe's words, a definitive and depoliticized rational consensus around a certain regulatory arrangement cannot exist. Instead, consensus will always exist as a temporary result of a provisional hegemony that stabilises power and social relations in a particular way. This is done necessarily to the exclusion of some, who may choose to contest such temporary regulatory arrangements in order to change them.[224]

92 Applying this to the DSA and AIA, if the two regulations explicitly count on civil society actors to enhance public scrutiny in this field and help contextualise risk assessment and mitigation, then contestation of risk management choices is a possibility for them. Therefore, the different loci of participation mapped in Table 1 may be theoretically framed as spaces of contestation,[225] albeit in varying degrees. Some loci are specifically geared towards *explicit* contestation by civil society, in that they allow or intend individuals and organisations to explicitly articulate their concerns about (and alternative proposals for) AI systemic risk management. Those loci are the ones identified in Table 1 as predominantly pursuing a contestatory rationale. This is the case of, for example, adversarial audits of VLOP/SEs' risk management reports based on vetted researcher access to information under the DSA: contestation is expected in this locus of participation and, therefore, it is practised explicitly.

93 Regulatory scrutiny through explicit contestation by civil society is, however, not the predominant rationale of all mapped loci of participation. In fact, it is not certain for numerous loci whether there will be a possibility for explicit contestation, as can be seen in Table 1. But in all loci, there is, I argue, space for implied contestation to occur.[226] This would be the case if civil society intervenes in a locus of participation not primarily designed to allow for explicit contestation but, nonetheless, the specific form of such an intervention either (i) implicitly builds upon a competing vision of risk management or (ii) engages in contestation despite the main rationale and expected form of participation of that locus. In simpler words, contestation is implied if it is not expected in a locus of participation but nonetheless practiced. This would be the case if, for example, during the collaborative development of an AIA code of conduct, intervening researchers or civil society organisations use their presence to propose alternatives to risk assessment or mitigation solutions advocated by regulated actors or the Commission. There would equally be implied contestation if, in an AI regulatory sandbox or innovation-fostering initiative set up by the Commission, intervening civil society would propose solutions of AI and/or platform design that build upon underrepresented gained knowledge of the concerns, perspectives, and lived experiences of individuals and communities regarding platforms' AI systems.

E. Conclusion: AI systemic risk will be what we want it to

94 Exegesis of legal texts through the canons learned in the continental legal tradition can only take us so far. It is increasingly common in legislation to see certain key concepts being so broadly defined that they may encompass many different, often conflicting meanings. Choosing one of those meanings from a set of possibilities is a value-laden choice. One can opt to use legal interpretation techniques to settle on one meaning for a certain broad legal concept, and then fictionalise that that was that concept's innate meaning all along. That is the way of legal practice and how, after all, most legal researchers in Europe learned to reason in law school. An alternative option would be to accept that certain vague legal concepts do not have such an innate meaning and, on the contrary, will be shaped by regulatory practice and dialogue. I must confess (as unorthodox as this candidness might be) that, when I started researching how EU law regulates digital platforms' AI systems, I still very much had in mind the traditional way of continental legal thought. However, after understanding that the EU opted for a risk-based regulatory approach and while seeking to answer the main research question guiding this paper - how do the DSA and AIA foresee creating an effective risk regulatory regime applicable to the AI

(1st edition, Routledge 2017) 63–65.

[223] Mouffe (n 218) 97–98, 100, 104; Crawford (n 3) 82–83.

[224] Mouffe (n 218) 104, 113, 126. In p. 126, Mouffe posits that contestation could be achieved through the promotion of civil society associations.

[225] To clarify, this theoretical framing constitutes the formulation of a hypothesis regarding the nature of civil society interventions in EU AI risk management which, necessarily, begs empirical questioning.

[226] The same remarks of the above footnote apply here.

systems of digital platforms? – I have quite clearly encountered the limits of that continental approach. And all due to the concept of 'AI systemic risk'.

95 As shown in Section B., the EU's risk regulation of platforms' AI systems revolves around the concept of systemic risk, primarily through the DSA systemic risk management scheme, which is complemented by numerous relevant AIA provisions. Despite several indications of the two regulations, AI systemic risks are not fully defined in the law. As is typical of risk-based approaches, that conceptual determination will be iteratively achieved by the concrete compounded outcomes of successive concrete risk management processes. Simply put, what is considered an AI systemic risk in each instance of AI risk management will eventually flesh out the meaning of this concept. Similarly, the strategies most commonly adopted to manage identified risks will be considered best practices of AI systemic risk mitigation. Most importantly, these will all be contingent choices, which may change.

96 Acknowledging the contingency of AI systemic risk definition and mitigation should inform both the regulatory implementation and research agenda of this EU's AI risk management framework. Specifically, there are three concrete implications of this acknowledgement which may turn into possible trajectories of future regulatory and scholarly dialogue.

97 First, significant conceptual focus should be put on the *who* - and not just the *what* - of AI systemic risk management. There is, of course, space for attempts to conceptualise the meaning of AI systemic risks and find adequate indicators and measures for such determination. But, in light of all the above, there must be significant empirical inquiry into how the meaning of 'AI systemic risk' is constructed in EU law. Namely, this means empirically questioning which actors have more agency in the field of AI systemic risk management, who influences systemic risk management choices the most, and whose ideas of what are AI systemic risks gain more currency in the developing regulatory dialogue.[227] This is especially true for the early stages of implementation of the DSA and AIA, where the meaning of this concept is still particularly undefined and is thus more malleable. In this sense, both enforcers and researchers should pay special attention to (i) the risk management methodologies used by different actors; (ii) the priorities and interests of those setting the regulatory agenda and thus focusing on certain types of systemic risks as opposed to others;[228] (iii) and the frames and models used to represent and capture the impact of platforms' AI systems in early regulatory dialogue.

98 Second, if systemic risk management is process-based, one should turn to the law for guidance on how those processes *should* go. By distilling from the law its normative aspirations for risk management processes, one gains not only an important benchmark for their internal assessment and critique but also a transparent enunciation of legal ambitions that can then be critiqued from external, non-legal viewpoints. That is what Section C. sought to accomplish. It identified three common normative aspirations cutting across the DSA and AIA risk management provisions.[229] The DSA and AIA frame the risks of platform-related AI systems as 'systemic'. They do so in socio-technical terms, by requiring that risk assessment and mitigation not be focused just on the technical traits of those AI systems as technological artifacts that cause isolated instances of harm; but rather on the more structural and collective impact that AI systems may have in their interactions with societal systems (C.I.). Consequently, risk management should be concretised through methodologies that socially contextualise the risks of platforms' AI systems and thus take risk perceptions by individuals and communities as a measure for their identification and subsequent management (C.II.). Institutionally speaking this requires that risk governance be participated and, specifically, the DSA and the AIA count with civil society involvement as a conduit for contextualising risk management and enhancing the scrutiny over platforms' risk management choices (C.III.). This also entails that if the perspectives of civil society on risk assessment and mitigation do not align with those of private regulated actors and platforms, civil society actors should have space to contest how AI systemic risks are assessed and mitigated in light of their technical and lived expertise.

99 Third and finally, if AI systemic risk management is to be contested and participated, then a fruitful focus of research and regulatory action is to map and consolidate an understanding of how such contestation and participation in AI systemic risk governance may occur. Section D. aimed to take a first step in this direction by looking at the law and mapping all existing and very different possibilities for civil society participation and involvement in the EU's AI risk management framework. Then, it more specifically identified and disaggregated the

[227] This empirical objective is framed in Bordeusian terms, see Yves Dezalay and Mikael Rask Madsen, 'The Force of Law and Lawyers: Pierre Bourdieu and the Reflexive Sociology of Law' (2012) 8 Annual review of law and social science 433.

[228] Similarly see Griffin, 'What Do We Talk about When We Talk about Risk?' (n 15).

[229] These are by no means the only ones, and future research could uncover more.

rationales of the identified loci of participation. Crucially, however, this was a theoretical exercise based on a reading of the existing law and civil society practice. It is, therefore, a 'best-case scenario' mapping of all the possible cracks in the law that different civil society actors can exploit to influence the legal implementation of platforms' AI risk management. Only a very optimistic person would expect all these loci to allow, in practice, for meaningful civil society interventions. Whether these can, in fact, become meaningful depends much more on practice. And there are many reasons why civil society participation could go wrong. Civil society actors could just be performatively involved in risk management processes, thereby legitimising private actors' risk management choices without having much ability to influence these outcomes.[230] Furthermore, different civil society actors have starkly disparate material and technical resources, available information, and access to participation fora, which may lead to a limitation of the types of concerns and proposals raised through civil society participation.[231] Finally, there is a possibility that the lived experiences of impacted individuals and communities are either not sufficiently represented by participating organisations or considered by private and public actors responsible for AI systemic risk management.[232]

100 All these open questions will eventually dictate how the meaning of AI systemic risks will be shaped in EU law. Above all, they beg a broad empirical research agenda, one that involves scholars of different perspectives in the task of scrutinising how AI systemic risks are identified and managed in EU law. Such a research agenda should both evolve and inform policymaking and regulatory implementation in this field. Scholarly and regulatory dialogue should be mindful of this: the EU legislator gave the concept of AI systemic risk enough latitude for it to be many things. Its definition and corresponding management are not purely technical matters; rather, they require legal, political and even ethical[233] choices to be made. These choices should not be the exclusive purview of those with more informational capacity and technological understanding of AI systems. In that sense, AI systemic risks can be what we, as a society, want to. Whether we will be given the space to articulate our concerns, perspectives and experiences and thereby shape risk management will dictate, to a large extent, the future of platform and AI regulation in EU law.

[230] Michele Gilman, 'Beyond Window Dressing: Public Participation for Marginalized Communities in the Datafied Society' (2022) 91 Fordham Law Review 503, 529–532; Kaminski and Malgieri (n 156) 39, 50.

[231] Griffin, 'Rethinking Rights in Social Media Governance' (n 108) 71–73; Marsh (n 40) 13–14; Karolina Iwánska and others, 'Towards an AI Act That Serves People and Society: Strategic Actions for Civil Society and Funders on the Enforcement of the EU AI Act' (European Center for Not-for-Profit Law 2024) 51–53 <https://europeanaifund.org/wp-content/uploads/2024/09/240827_FINAL_AI_ACT_Enforcement.pdf> accessed 25 September 2024. General reports on effective civil society participation highlighted the need to ensure independent public funding and adequate staff training of civil society organisations, see e.g. Vanja Skoric, 'Standards and Good Practices for Public Funding of Civil Society Organisations' (European Center for Not-for-Profit Law Stichting 2020) 14–15, 60–63 <https://ecnl.org/sites/default/files/2020-09/TUSEV%20Public%20Funding%20Report_Final.pdf?utm_source=chatgpt.com> accessed 29 January 2025.

[232] Gilman (n 230) 529; Kaminski and Malgieri (n 156) 16.

[233] Mittelstadt and others (n 3); 'Ethics Guidelines for Trustworthy AI' (Independent High-Level Expert Group On Artificial Intelligence - European Commission 2018) 9–13.

The Cyber Resilience Act and Open-Source Software: A Fine Balancing Act

by **Mattis van 't Schip** *

Abstract: Open-source software, a type of software that can be publicly accessed, shared, and modified, is an integral part of modern digital infrastructure. Many products, from personal computers to internet-connected devices, run on open-source systems (e.g., Linux). Developers may work voluntarily or for limited compensation on such software. The character of this work, however, does not reduce the impact of cybersecurity incidents within these environments. Proprietary software, meaning software with restrictive license models, regularly implements open-source software: a vulnerability in the open-source software thus directly affects proprietary software too. Recent large-scale vulnerabilities (e.g., Log4j) highlighted this dual nature of open-source software: developers work on projects based on personal passion or ideologies, while the software is often equally as critical as software created and maintained by larger technology enterprises.

The Cyber Resilience Act, the recently proposed European cybersecurity legislation for products, aims to offer a legal response to cybersecurity problems in modern software and hardware. This paper addresses the role of open-source software cybersecurity in the Cyber Resilience Act with specific attention to the difficulties of reconciling cybersecurity responsibilities and open-source products. I show that the Cyber Resilience Act does achieve a balance between regulation for open-source software and advancing cybersecurity, but only through a narrowly applicable and, at times, complex legislative approach.

Keywords: Open-Source Software; Cybersecurity; Cyber Resilience Act

Recommended citation: Mattis van 't Schip, The Cyber Resilience Act and Open-Source Software: A Fine Balancing Act, 16 (2025) JIPITEC 73 para 1.

A. Introduction

1 Behind the facade of giant technology enterprises exists an ecosystem of 'open-source software'. The source code of this type of software is publicly accessible and developers write the code under licenses that allow for use, redistribution, modification, and sharing by third parties. 'Open source' does not merely mean public access to source code. The Open Source Initiative (OSI), a body responsible for the generally accepted definition of 'open source', indicates that the concept holds certain additional criteria.[1] For instance, open-source software licenses should not discriminate based on intended use.[2]

2 Many types of open-source software support today's largest software packages: Linux, an open-source operating system, powers many modern ICT products, from desktop computers to Internet of Things devices; millions of websites rely on Apache, an open-source web server. Open-source software is thus an important cornerstone of the modern digital infrastructure.[3]

3 The advantages of open-source software align with recent regulatory efforts in the EU that aim to curtail the market power of the major digital enterprises. For instance, the Digital Services Act regulates online

* Ph.D. Candidate at the Interdisciplinary Research Hub on Digitalization and Society (iHub), Radboud University. This research is funded through the NWO INTERSCT project [NWA.1160.18.301].

[1] Open Source Initiative, 'The Open Source Definition' (22 March 2007) <https://opensource.org/osd> accessed 19 January 2024.

[2] <https://opensource.org/licenses/>.

[3] Chinmayi Sharma, 'Tragedy of the Digital Commons' (2023) 101 North Carolina Law Review 1129.

platforms (and especially the "very large" online platforms, i.e., the major social media platforms), while the Digital Markets Act imposes responsibilities on "gatekeepers" (e.g., Microsoft, Meta).[4] Open-source software can serve as a transparent, public alternative to these dominant platforms.

4 Like other types of software, open-source software comes with cybersecurity risks.[5] For example, Log4j, a piece of open-source software for logging purposes, suffered a critical vulnerability which allowed hackers to remotely access systems.[6] Some experts held that the vulnerability affected virtually every digital service globally.[7] The Log4j vulnerability was critical because open-source software is often incorporated in larger proprietary software packages; the vulnerability in Log4j thus directly affected numerous other products.[8]

5 In September 2022, the European Commission introduced a new legislative proposal for the cybersecurity of software and hardware products, the Cyber Resilience Act.[9] At the end of 2024, the Act came in effect.[10] The Act applies when manufacturers and/or software developers place software or hardware products on the market of the European Union "in the course of a commercial activity".[11] If they place these products on the market, software developers must implement certain cybersecurity requirements in their product and, in certain cases, follow strict assessment procedures.

6 Although this requirement potentially excludes open-source software, the 'commercial activity' condition offers few assurances, as evident from the legislative discussions surrounding its interpretation.[12] The commerciality of open-source software projects can range from monetising other services on the open-source software platform (e.g., Android) to occasional donations from end users (e.g., hobby projects).[13] The Commission proposal merely mentioned these examples, but did not offer a further clarification of what "supplying in the course of a commercial activity" entails.

7 The text adopted by the Parliament, instead, includes a rather comprehensive set of Recitals, which cover many open-source software development and financing methods. The compromise text therefore exempts nearly every known type of open-source software development from the scope of the Cyber Resilience Act. This exemption helps developers, who do not have to comply with legal burdens for software that they provide openly to the public.

[4] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [2022] OJ L277/1 (Digital Services Act); Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 [2022] OJ L265/1 (Digital Markets Act).

[5] Jaap-Henk Hoepman and Bart Jacobs, 'Increased Security through Open Source' (2007) 50 Communications of the ACM 79.

[6] For an extensive overview, see Raphael Hiesgen and others, 'The Log4j Incident: A Comprehensive Measurement Study of a Critical Vulnerability' [2024] IEEE Transactions on Network and Service Management 1.

[7] Sean Lyngaas, 'US Warns Hundreds of Millions of Devices at Risk from Newly Revealed Software Vulnerability' (*CNN*, 13 December 2021) <https://www.cnn.com/2021/12/13/politics/us-warning-software-vulnerability/index.html> accessed 19 January 2024; Ars Technica spoke of 'arguably the most severe vulnerability ever', see Dan Goodin, 'As Log4Shell Wreaks Havoc, Payroll Service Reports Ransomware Attack' (*Ars Technica*, 13 December 2021) <https://arstechnica.com/information-technology/2021/12/as-log4shell-wreaks-havoc-payroll-service-reports-ransomware-attack/> accessed 19 January 2024; Similarly, see the Guardian Associated Press, 'Recently Uncovered Software Flaw "Most Critical Vulnerability of the Last Decade"' *The Guardian* (11 December 2021) <https://www.theguardian.com/technology/2021/dec/10/software-flaw-most-critical-vulnerability-log-4-shell> accessed 19 January 2024.

[8] Sharma (n 4) 1131–1133.

[9] Proposal for a Regulation of the European Parliament and of the Council on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020 COM(2022) 454 final [Cyber Resilience Act].

[10] Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) No 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act) [2024] OJ L (to be published).

[11] Art 3(22) CRA.

[12] See the calls for support from the open source community when the Cyber Resilience Act proposal was published in, inter alia, Maarten Aertsen, 'Open-Source Software vs. the Proposed Cyber Resilience Act' (*The NLnet Labs Blog*, 14 November 2022) <https://blog.nlnetlabs.nl/open-source-software-vs-the-cyber-resilience-act/> accessed 20 December 2023; Deb Nicholson, 'Python Software Foundation News: The EU's Proposed CRA Law May Have Unintended Consequences for the Python Ecosystem' (*Python Software Foundation News*, 11 April 2023) <https://pyfound.blogspot.com/2023/04/the-eus-proposed-cra-law-may-have.html> accessed 20 December 2023; Simon Phipps, 'What Is the Cyber Resilience Act and Why It's Dangerous for Open Source' (*Voices of Open Source*, 24 January 2023) <https://blog.opensource.org/what-is-the-cyber-resilience-act-and-why-its-important-for-open-source/> accessed 20 December 2023.

[13] David A Wheeler, 'F/LOSS Is Commercial Software' [2009] Open Source Business Resource <http://timreview.ca/article/229>.

At the same time, these broad exemptions could undermine the overall aim of the Cyber Resilience Act to improve the state of cybersecurity for software and hardware.

8 This paper analyses the difficulties of reconciling open-source software development with cybersecurity risk management responsibilities. The research question is: To what extent does the Cyber Resilience Act impose responsibilities on open-source software developers that achieve a balance between stimulating open-source software development and, simultaneously, mitigating cybersecurity problems within open-source software?

9 This paper proceeds as follows: Section B summarises the history and meaning of open-source software and its cybersecurity implications. Section C discusses the Cyber Resilience Act, with specific attention to the definition of supplying a product 'in the course of a commercial activity' for open-source software products. Section D highlights, using several examples, how difficult an assessment of 'supplying in the course of a commercial activity' is under the current legal terminology in the Recitals. Section E then looks at specific rules pointed at open-source software within the Cyber Resilience Act, such as the special regulatory regime for 'open-source software stewards'. Based on this legal framework, Section F questions whether the Cyber Resilience Act now achieves a balance between encouraging open-source software development and mitigating cybersecurity problems. Based on this balance, Section G looks at the future of open-source software under EU law. Section H concludes.

B. Open-Source Software

10 Open-source software originates from an academic environment. At MIT, Richard Stallman intended to design a free operating system that opposed the barriers developing against sharing software in the 1980's.[14] To support the GNU project, Stallman established the Free Software Foundation (FSF). The FSF focused on free access and usability of software ('a matter of liberty'[15]) instead of 'free of charge' software.[16] A decade later, the quickly growing community surrounding 'free software' moved towards a new label: 'open source'. The 'free' label was unattractive to many companies, which prevented larger enterprises from becoming involved in the development of 'free' software.[17] Therefore, under the Open Source Initiative, the community created a definition for 'open-source' software next to 'free' software.[18]

11 Open-source software is a type of software with source code that is publicly accessible. The use of open-source software comes with some requirements, which different developers have formalised in specific licenses.[19] Some developers, for instance, specify that users accept that they receive the software 'as-is', so that the developers cannot be held liable for damages caused by the software.[20] At the same time, the licenses also formalise that the developers cannot discriminate based on the envisioned use of the software: any type of user (e.g., large technology companies, hobby developers) can freely access and use the code how they desire (e.g., modification, sharing).[21]

12 This Section analyses open-source software and its unique characteristics in comparison to its counterpart, proprietary/closed-source software. In addition, the Section highlights the cybersecurity characteristics of both software development methods.

I. The Development and Ideologies of Open-Source Software

13 Open-source software is published on a diverse set of platforms by equally diverse developers. Developers participate to different degrees (e.g., occasional code change to full-time work), receive different types of remuneration (e.g., full salary, donations), and contribute based on diverse motivations (e.g., passion, peer recognition). This diversity laid the groundwork for 'open source' as a community of people involved with all types of projects that aim at providing open access to information and knowledge, such as open-source software and open access science.

14 The counterparts to open-source software exists in two forms: proprietary software (restrictive

[14] Richard Stallman, 'Initial Announcement' (*GNU*, 27 September 1983) <https://www.gnu.org/gnu/initial-announcement.html> accessed 19 January 2024.

[15] 'What Is Free Software? - GNU Project - Free Software Foundation' <https://www.gnu.org/philosophy/free-sw.html.en> accessed 13 January 2025.

[16] Moreno Muffatto, *Open Source: A Multidisciplinary Approach* (Imperial College Press 2006) 7.

[17] 'History of the OSI' (*Open Source Initiative*, 19 September 2006) <https://opensource.org/history/> accessed 19 January 2024.

[18] Open Source Initiative (n 2); Muffatto (n 17) 14.

[19] P McCoy Smith, 'Copyright, Contract, and Licensing in Open Source' in Amanda Brock (ed), *Open Source Law, Policy and Practice* (2nd edn, Oxford University Press 2022).

[20] See, as an example, the 1-clause BSD license: <https://opensource.org/license/bsd-1-clause>.

[21] Open Source Initiative (n 2).

licensing) and closed-source software (restricted access to source code). Proprietary software works with licenses that severely restrict the user in their use of the software (e.g., no modifying the source code). Proprietary software can thus also be open-source software, as software with publicly accessible source code but a restrictive license.[22] In addition, proprietary software exists as closed-source software, where the source code is not available *and* the license restricts the user. Open-source or closed-source software is thus a choice during the development phase of a software package, while proprietary software refers to the distribution phase.

15 The dichotomy between open-source and proprietary/closed-source software can be illustrated through the Linux and Microsoft Windows operating systems: Linux is an open-source operating system, with many different versions existing today, because the license allows modification of the code (e.g., Linux Mint, Ubuntu, Arch Linux).[23] Microsoft develops the proprietary and closed-source Windows operating system; its source code is not publicly available and its license restricts any modification to the Windows source code. Microsoft thus solely develops and controls the different Windows versions.

16 Open-source software exists in many forms. Linux is a prominent example because, as a popular operating system, it has millions of users. However, open-source software also exists on a smaller scale, for example as a small web app that maybe a hundred people may use. When the software license complies with the open source definition of the Open Source Initiative (OSI),[24] the open source community considers it open-source software.[25]

17 There is no singular form of organization behind open-source software development.[26] Since open-source software is usually – but certainly not always – free for users, open-source software developers often rely on smaller financial resources to build their software. Open-source developers often have other intrinsic and extrinsic motives. Intrinsic motives rely on "the tendency to seek out novelty and challenges" (e.g., improving knowledge of a certain programming language), while extrinsic motives focus on the outcome of certain conduct (e.g., improving reputation among peers in the development community).[27] Some developers therefore band together under a non-commercial entity and offer technical support to their largest users for a fee, while other developers work on projects completely voluntarily or based on small donations from end users.

18 In connection to the structure of different open-source software, the users of the software differ considerably, as anyone can access the software's source code. Major technology enterprises frequently use open-source software as a foundation on which they build their proprietary software packages; individuals might instead use open-source software because of its lower cost or as an alternative to the monopoly power of large technology enterprises.[28]

19 In line with these different structures and users of software, I identify three types of open-source software projects: 1) a standalone open-source project (e.g., a developer publishing some personal code); 2) open-source software incorporated into other proprietary and/or open-source software (e.g., Log4j);[29] 3) commercialized open-source software (e.g., where the organisation requires a fee for usage).[30] The difference between a standalone project (1) and an integrated project (2) largely relies on the use case of the software package, since some packages do not offer standalone functionalities.[31] Section C illustrates the meaning of this categorization within the legal framework of the Cyber Resilience Act.

20 Open-source developers often publish the source code of their software on online repositories (e.g., GitHub, SourceForge, personal websites). Other developers can access the code there, and download it for further use, or review the code and offer

[22] For some examples, see <https://en.wikipedia.org/wiki/List_of_proprietary_source-available_software>.

[23] For an extensive list of Linux distributions, see <https://en.wikipedia.org/wiki/List_of_Linux_distributions>.

[24] <https://opensource.org/licenses/>.

[25] This does not mean that there are no open-source software licenses outside the OSI's list, but merely that the OSI has not (yet) classified them as compliant with the open source definition.

[26] Muffatto (n 17) ch 3.

[27] Jürgen Bitzer, Wolfram Schrettl and Philipp JH Schröder, 'Intrinsic Motivation in Open Source Software Development' (2007) 35 Journal of Comparative Economics 160; Muffatto (n 17) 58–62.

[28] Muffatto (n 17) 62–64.

[29] For instance, on Microsoft's evolving stance towards open-source software, see Benjamin J Birkinbine, *Incorporating the Digital Commons: Corporate Involvement in Free and Open Source Software* (University of Westminster Press 2020) 49–72.

[30] RedHat is the most prolific example of such projects, see also ibid 73–88; Although Red Hat recently changed its company policies, to the dismay of the open source community, Kevin Purdy, 'Red Hat's New Source Code Policy and the Intense Pushback, Explained' (*Ars Technica*, 30 June 2023) <https://arstechnica.com/information-technology/2023/06/red-hats-new-source-code-policy-and-the-intense-pushback-explained/> accessed 13 December 2023.

[31] The Cyber Resilience Act also speaks of certain types of open-source software 'intended for integration by other manufacturers'. Recital 18 CRA.

feedback.[32]

21 Open-source software is part of the broader 'open source' movement, which is based on certain philosophical (e.g., about information and knowledge) or pragmatic beliefs (e.g., free alternatives for users) about the need for open-source software.[33] These beliefs explain the altruistic nature of open source and relate back to the Free Software Foundation: many developers offer their software to the public because they are part of a wider community movement which aims to keep knowledge, in a broad sense, publicly accessible and shareable.[34]

II. Open-Source Software and Cybersecurity

22 Open-source software represents a deliberate choice for transparency: the source code of the software is accessible and the developers are transparent about its inner workings. An alternative to such transparency is 'security through obscurity'.[35] This dichotomy between 'transparency' and 'obscurity' forms the foundation for many security-related discussions about open-source software.[36]

23 By hiding the inner workings of the software, closed-source software does not show its internal processes; attackers cannot view the source code to discover exploitable vulnerabilities.[37] In contrast, advocates for open-source software development believe transparency allows open-source software to be more secure.[38] In the following, I illustrate the security dynamics of open-source and closed-source software in two phases: 1) during the development of the software and 2) after publication of the software.

1. Development of Software

24 Proponents often use the transparent nature of open-source software as an argument that open-source software is more secure; if developers can peer review source code, they can identify and patch vulnerabilities and similar problems quickly.[39]

25 Raymond coined this view of security of open source code as 'Linus' Law': "Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone."[40] Thus, an open-source software package is more secure if – and only if – many developers view and co-operate on the source code as the project benefits from their diverse views.[41]

26 The prevention of backdoors is an example of the benefits of the 'many eyeballs' system. If attackers change the source code of open-source software to allow themselves backdoor access to the system, or if the backdoor existed from the start, other developers can easily notice such changes and prevent the attackers from exploiting the backdoor.[42] This is not the case for closed-source systems, where such backdoors are not immediately visible to others.

27 An opposing view to the 'many eyeballs' principle of Linus' Law is the view of 'too many cooks in the kitchen'.[43] In the latter view, the security of open-source software diminishes because too many developers are working on the software simultaneously and in fragmented ways.[44] A single developer may decide to contribute solely to their preferred elements of the project, without

[32] GitHub had more than 400 million contributions to open-source projects in 2022. See <https://github.blog/news-insights/research/octoverse-2022-10-years-of-tracking-open-source/>.

[33] Ian Walden, 'Open Source as Philosophy, Methodology, and Commerce: Using Law with Attitude' in Amanda Brock (ed), *Open Source Law, Policy and Practice* (2nd edn, Oxford University Press 2022).

[34] Charlotte Hess and Elinor Ostrom (eds), *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007).

[35] Hoepman and Jacobs (n 6).

[36] Charles-H Schulz, 'Open Source Software and Security: Practices, Governance, History, and Perceptions' in Amanda Brock (ed), *Open Source Law, Policy and Practice* (2nd edn, Oxford University Press 2022); Christian Payne, 'On the Security of Open Source Software' (2002) 12 Information Systems Journal 61.

[37] Ross Anderson, 'Open and Closed Systems Are Equivalent (That Is, in an Ideal World)' in Joseph Feller and others (eds), *Perspectives on free and open source software* (MIT Press 2005).

[38] Eric Raymond, 'The Cathedral and the Bazaar' (1999) 12 Knowledge, Technology & Policy 23.

[39] ibid.

[40] 'Linus' refers to the founder of the Linux operating system, Linus Torvalds. Raymond also more informally coins Linus' Law as "Given enough eyeballs, all bugs are shallow", see ibid 29.

[41] Raymond (n 38).

[42] Payne (n 36) 66–67.

[43] Andrew Meneely and Laurie Williams, 'Secure Open Source Collaboration: An Empirical Study of Linus' Law', *Proceedings of the 16th ACM conference on Computer and communications security* (ACM 2009) 453; Ann Barcomb and others, 'Managing Episodic Volunteers in Free/Libre/Open Source Software Communities' (2022) 48 IEEE Transactions on Software Engineering 260.

[44] Martin Pinzger, Nachiappan Nagappan and Brendan Murphy, 'Can Developer-Module Networks Predict Failures?', *Proceedings of the 16th ACM SIGSOFT International Symposium on Foundations of software engineering* (ACM 2008).

contributing to the overall project goals. This 'unfocused contribution' forms a security risk.[45] Unfocused contributions disrupt the concept of Linus' Law in large-scale open-source projects, as the additional 'eyeballs' do not necessarily improve the project.[46] Therefore, the idea that open-source software is more secure simply because a diverse set of developers can access the source code is not clearly proven.

2. Post-Release Vulnerabilities

28 Linus' Law mainly relates to the development phase of open-source software projects. However, security problems can also develop in the post-release phase, after publication of the software or a new version release.

29 In a comprehensive study, Schryen found that there was no statistical significance in terms of the severity of vulnerabilities between open-source and closed-source software equivalents.[47] He also found that the type of patching behaviour, in terms of speed and type of vulnerabilities, differed significantly between different open-source and closed-source vendors. This difference existed across open-source and closed-source vendors: the mode of open-source or closed-source development seemed, therefore, not to influence patching behaviour.[48]

30 Ransbotham analyses how threat actors exploit vulnerabilities differently between open-source and closed-source projects based on two years of log data from intrusion detection systems.[49] He holds that vulnerabilities of open-source software projects have a generally greater risk of exploitation and receive more exploitation attempts. These differences can be partially attributed to the difference in transparency between open- and closed-source software. If a vulnerability is discovered internally in a closed-source environment, the developers have some additional time to work on fixing the vulnerability before they make the changes public. In open-source projects, changes in the source code – and thus possible vulnerabilities – are immediately publicly accessible.[50]

31 In general, there are thus small differences between open-source and closed-source software security, both in the development and post-release phase. Vulnerabilities exist in and impact both types of software.

C. The Cyber Resilience Act and Open-Source Software Cybersecurity

32 European law did not consider cybersecurity rules for open-source software until 2022. This lack of regulation changed when the European Commission proposed the 'Cyber Resilience Act', which contained specific rules for open-source software cybersecurity.[51] The Cyber Resilience Act was adopted at the end of November 2024 and comes into effect on 10 December 2024.[52]

I. The Cyber Resilience Act in Short

33 The Cyber Resilience Act imposes 1) *cybersecurity requirements* on 2) *manufacturers* of 3) *products with digital elements* that they 4) *place on the Union's market* in the course of a 5) *commercial activity*.[53] Below, I briefly review these elements in light of the applicability of the Act to open-source software.[54]

1. Cybersecurity Requirements

34 The cybersecurity requirements for products with digital elements form the focal point of the Cyber Resilience Act. These requirements include security throughout the lifecycle of the product (security-by-design), releasing the product without known exploitable vulnerabilities, and protection of the integrity and authenticity of data.[55] Next to these requirements, the Act contains traditional product requirements (e.g., providing documentation) and security-specific duties (e.g., providing security

[45] Meneely and Williams (n 43) 456.

[46] Meneely and Williams (n 43); Andrew Meneely and Laurie Williams, 'Strengthening the Empirical Analysis of the Relationship between Linus' Law and Software Security', *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (ACM 2010).

[47] Guido Schryen, 'Is Open Source Security a Myth?' (2011) 54 Communications of the ACM 130, 136–137.

[48] ibid 139.

[49] Sam Ransbotham, 'An Empirical Analysis of Exploitation Attempts Based on Vulnerabilities in Open Source Software' [2010] Workshop on the Economics of Information Security 1.

[50] ibid 5.

[51] Cyber Resilience Act proposal (n 10).

[52] Cyber Resilience Act (n 11).

[53] Art 1 CRA.

[54] See also Liane Colonna, 'The End of Open Source? Regulating Open Source under the Cyber Resilience Act and the New Product Liability Directive' (2025) 56 Computer Law & Security Review 106105.

[55] Annex I Part 1 CRA.

updates).[56]

35 There are two main methods for developers to show their compliance in the proposal: 1) performing a self-assessment; and 2) receiving a third-party audit.[57] In general, the choice for a specific route depends on the type of product. The Cyber Resilience Act categorises products with certain privileges in networks or computer systems (e.g., password managers, operating systems) as 'important' products.[58] Important products must, if they cannot follow certain European technical standards, perform a third-party audit to prove their compliance with the Act's requirements.[59] Open-source software is exempted from a third-party audit, even if they are considered 'important products', as long as they provide technical documentation to the public.[60]

36 The provision and the supporting Recital do not indicate a reason for this exemption. However, many open-source software packages have certain elevated privileges and would therefore be important products (e.g., Log4j). In that context, the Parliament and Council most likely wanted to prevent a 'chilling effect' on open-source software development in the face of possibly costly third-party audits.

37 A further category exists for 'critical' products with digital elements, with even stricter conformity requirements.[61] The Act currently lists three critical products: hardware devices with security boxes; smart meter gateways; and smartcards.[62]

2. Manufacturers

38 A manufacturer is a "natural or legal person who develops or manufactures products with digital elements".[63] Both traditional hardware manufacturers and software developers are 'manufacturers' under the Cyber Resilience Act. In case manufacturers do not strictly produce the product themselves, but place their trademark on products produced by another actor, they remain the manufacturer of the final product.[64]

39 As highlighted above, not all open-source software forms a standalone package. Some of the most prominent open-source software packages derive their popularity from integration by proprietary software developers. Google, for instance, uses numerous pieces of open-source software, such as databases,[65] for their own software packages (e.g., Google Maps). Google, in this example, creates and markets their end product and is thus the manufacturer for the end product under the Cyber Resilience Act.[66] The proprietary developers must thus also ensure that they securely integrate the open-source database system – the open-source developer is not responsible for compliance in this case.[67] I delve into this separation further in Section E.II.

40 The Cyber Resilience Act includes a set of rules for importers and distributors too. These rules ensure that manufacturers cannot evade compliance by letting importers and distributors bring the product to the Union market.[68] An importer brings products with digital elements to the Union market of "a natural or legal person established outside the Union."[69] A distributor is an actor that is not a manufacturer or importer, but who still places the product on the market.[70] Importers and distributors have separate responsibilities to ensure that the products they place on the Union market comply with the requirements of the Cyber Resilience Act.[71]

3. Product with Digital Elements

41 The provisions of the Cyber Resilience Act apply to 'products with digital elements', meaning "any software or hardware product".[72] Open-source software is thus a 'product with digital elements' if: 1) the open-source project develops *software* or *hardware*; and 2) that software or hardware is a *product* under the Cyber Resilience Act.

[56] Art 13 CRA.

[57] Art 32(1) CRA.

[58] Art 7(1) CRA & Annex III CRA. The Commission proposal used the term 'critical' products, which is now an even more critical class above important products.

[59] Art 32(2) CRA.

[60] Art 32(5) & Recital 91 CRA.

[61] Art 8 & Art 32(4) CRA.

[62] Annex IV CRA.

[63] Art 3(13) CRA.

[64] Art 3(13) CRA.

[65] For instance, Google moved their database systems to the open-source MariaDB, see Jack Clark, 'Google Swaps out MySQL, Moves to MariaDB' *The Register* (12 September 2013) <https://www.theregister.com/2013/09/12/google_mariadb_mysql_migration/> accessed 14 August 2024.

[66] Art 3(13) & Art 13(5) CRA.

[67] Izquierdo Grau analyses this division between standalone open-source and integrated open-source in the context of the recent Product Liability Directive proposal, see Guillem Izquierdo Grau, 'An Appraisal of the Proposal for a Directive on Liability for Defective Products' (2023) 12 Journal of European Consumer and Market Law 198.

[68] Art 19 & 20 CRA.

[69] Art 3(16) CRA.

[70] Art 3(17) CRA.

[71] Art 19(2) & 20(2) CRA.

[72] Art 3(1) CRA.

42 The Cyber Resilience Act defines open-source software as "software the source code of which is openly shared and [...] made available under a free and open-source license."[73] From this definition, however, it is not immediately clear that open-source software is also a software *product.*

43 The Cyber Resilience Act itself does not define what a 'product' is. The EU's Blue Guide, the Commission's interpretation guide for product rules, offers some additional guidance for definitions related to European product legislation.[74] The Guide defines a product in relation to its placing on the market: "Union harmonisation legislation applies to products which are intended to be placed (and/or put into service) on the market."[75] This element of 'placing onto the market' is thus an important qualifier for open-source software as a software *product* under the Cyber Resilience Act.

4. Placing on the Market

44 The Cyber Resilience Act defines that a product is placed on the market when it is "made available" on the Union market, meaning "the supply of a product [...] for distribution or use [in the Union] *in the course of a commercial activity*, whether in return for payment or free of charge."[76] These definitions highlight that open-source software can thus be offered on the market – and therefore be a product under the Cyber Resilience Act – even if the software is offered for free.

45 Additionally, open-source software is "placed on the market" in the sense of the Cyber Resilience Act if the developer supplies the product "in the course of a commercial activity". Although this is an additional requirement, its abstract character caused much discussion after the Commission's proposal.[77]

5. Commercial Activity

46 The provisions of the Cyber Resilience Act do not clearly define 'supplying a product in the course of a commercial activity'. Recital 18 of the Cyber Resilience Act states that "only free and open-source software made available on the market, and therefore supplied for distribution or use in the course of a commercial activity should be covered by this Regulation." Although the Recitals are not legal provisions, they offer an interpretation of what 'supplying in the course of a commercial activity' means in the context of open-source software.[78]

47 The Recitals note several examples of open-source software supplied in the course of a commercial activity. Open-source software is supplied in the course of a commercial activity if the developer 1) charges a price for a product; 2) charges a price for technical support services that does not serve the recuperation of actual costs; 3) provides a software platform where the manufacturer monetises other services; or 4) if the software requires as a condition for use the processing of personal data, unless for certain legitimate purposes (e.g., security).[79] The legislators seemingly had particular open-source projects in mind when drafting these examples. For instance, the provision of a software platform where the manufacturer monetises other services can relate to Android: the core of Google's mobile operating system is open source, but Google integrates the Google Play Store, Google Drive, and other similar services into Android when providing the platform to smart phones.[80]

48 This list is not exhaustive, as the Recital notes that supply within the course of a commercial activity "might be characterised" by the options mentioned above.[81] Other activities and conditions can also bring the open-source software project in the context of a commercial activity, placing additional emphasis on the question when an activity is 'commercial' under the Cyber Resilience Act.

49 Many hobby developers add donation options to their open-source software (e.g., Patreon, PayPal). Developers often make such donation requests to cover the project's maintenance costs (e.g., website

[73] Art 3(48) CRA.

[74] Commission notice – The 'Blue Guide' on the implementation of EU product rules [2022] OJ C247/1.

[75] Blue Guide (n 75), 17.

[76] Art 3(22). Emphasis mine.

[77] Aertsen (n 13); Webmink In Draft, 'Fixing The CRA For Open Source' (*Webmink In Draft*, 20 February 2023) <https://the.webm.ink/fixing-the-cra-for-open-source> accessed 21 February 2023; Nicholson (n 13).

[78] See Llio Humphreys and others, 'Mapping Recitals to Normative Provisions in EU Legislation to Assist Legal Interpretation', *JURIX* (2015) 42–44 and cases cited therein.

[79] Recital 15 CRA.

[80] Ron Amadeo, 'Google's Iron Grip on Android: Controlling Open Source by Any Means Necessary' (*Ars Technica*, 21 July 2018) <https://arstechnica.com/gadgets/2018/07/googles-iron-grip-on-android-controlling-open-source-by-any-means-necessary/> accessed 19 January 2024.

[81] Recital 15 CRA.

costs).[82] At the same time, research shows that, in certain large-scale open source projects, code contributions by companies can be ten times larger than contributions by volunteers.[83] Such large-scale contributions might lead to the conclusion that the entire open-source project falls into a 'commercial activity', as commercial parties maintain nearly the entire project. A strict dichotomy between open-source software and commerciality does not exist.[84] There are diverse ways in which an open-source project can obtain financial and/or organisational support.[85]

50 The Commission proposal lacked insight into these diverse methods of commerciality, as the text only gave examples of open-source software supplied during a commercial activity.[86] The Council and Parliament, in response, significantly expanded the Recitals, especially regarding open-source software. As a result, the legislators exempted many types of open-source software from the scope of the Act. For example, the amended Recitals state that asking for donations does not constitute supply in the course of a commercial activities, as long as the developers do not seek to gain profits from those donations.[87] Furthermore, the Recitals state that an open-source project is not supplied in the course of a commercial activity merely due to development support from commercial entities.[88] In sum, the role of open-source software within the Cyber Resilience Act largely depends on whether the software is supplied in the course of a commercial activity.

D. Assessing the commerciality of a project

51 The commerciality of open-source software largely determines whether the software falls under the scope of the Cyber Resilience Act. Therefore, the exact meaning of 'supplying in the course of a commercial activity' merits further examination.

52 Most activities are commercial if developers use them to earn a profit, i.e. the income from these actions exceed maintenance costs. For example, the Cyber Resilience Act lists charging a price for the software or for technical support, when this exceeds maintenance costs, as indicative of supplying the software in the course of a commercial activity.[89]

53 In contrast, certain projects are not supplied during a commercial activity. Again, developers of such projects mostly do not earn income that exceeds their maintenance costs, such as receiving small donations.

[82] Cassandra Overney and others, 'How to Not Get Rich: An Empirical Study of Donations in Open Source', *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering* (ACM 2020).

[83] Yuxia Zhang and others, 'Companies' Participation in OSS Development–An Empirical Study of OpenStack' (2021) 47 IEEE Transactions on Software Engineering 2242, 2249.

[84] Wheeler (n 14).

[85] ibid.

[86] Aertsen (n 13).

[87] Recital 15 CRA.

[88] Recital 18 CRA.

[89] Recital 15 CRA.

Table 1: The scope of the Cyber Resilience Act for open-source software[90]

Indicative of a supplying the software *in* the course of a commercial activity	• An intention to monetise beyond the recuperation of actual costs • Charging a price for the product • Charging a price for technical support • Personal data processing as a condition for use of the software (except for certain justified purposes) • Accepting donations exceeding the costs of developing and maintaining the software, without the intention to make a profit.
Indicative of a supplying the software *outside* the course of a commercial activity	• Monetisation only to recuperate costs of maintenance, instead of making a profit (e.g., by public administration entities) • Supply of software intended to be integrated by other manufacturers, without monetisation of original software • Products which receive financial support or developmental support from manufacturers • The mere presence of regular releases • Development by non-profit organisations, if they use their earnings after cost for non-profit objectives • Contributions to open-source software when not involved in project leadership/ownership • Mere distribution on repositories
Special regulatory regime	Open-source software stewards, legal persons who "provide support on a sustained basis" for the development of open-source software and play a "main role in ensuring the viability" of open-source software

[90] Recital 16-20 CRA.

54 Table 1 shows how the Recitals include and exempt numerous open-source software projects from the scope of the Cyber Resilience Act. Based on this overview, a few questions remain.

55 The list of commercial activities in the Recitals is non-exhaustive; the Recital states that a commercial activity "*might be* characterised" by the options mentioned.[91] In the future, courts may thus amend the list and determine that other activities are also commercial.

56 An assessment of other activities, however, is difficult, as the Recitals further state that "the mere circumstances under which the product has been developed, or how the development has been financed, should [...] not be taken into account" when assessing the commercial nature of the software.[92] This limitation seems to directly contradict the Recitals themselves. As shown in Table 1, the Recitals explicitly exempt certain types of development (e.g., development by commercial entities) and financial models (e.g., receiving donations) from the scope of 'supplying a product in the course of a commercial activity'. A court can thus seemingly not assess the commerciality of a project as the Recitals currently do.

57 Additionally, the Recitals contain an unclear role for the *intention* of gaining a profit. In the context of donations, the Recitals state that accepting donations "exceeding the costs [of] design, development and provision of a product" means that the software is supplied in the course of a commercial activity.[93] In contrast, when developers accept donations "without the *intention* of making a profit", they do not supply the product in the course of a commercial

[91] Recital 15 CRA. Emphasis mine.

[92] Recital 18 CRA.

[93] Recital 15 CRA.

activity.[94] It is unclear from the Recitals when intent is measured: at the start of the project or when the developers introduce certain financing methods. A developer may not intend to make a profit initially, but, as the project grows, may consider it reasonable. Likewise, the developer may not intend to make a profit, but may receive such large donations from enthusiastic users that they completely exceed all maintenance costs. Such situations, which are not clearly determined in the Recitals, remain complex.

58 The Commission may still resolve some of the Recital's complexities. Pursuant to Article 26 of the Act, the Commission may publish guidance to support the application of the Cyber Resilience Act. The scope of the Act for free and open-source software is of particular importance when they provide such guidance.[95]

59 In sum, the Recitals, in general, indicate clearly when open-source software is commercial. If developers publish open-source software for which consumers pay a commercial price or other consideration (e.g., personal data), they supply the software in the course of a commercial activity. If, in contrast, developers merely maintain or support open-source software, they do not supply the software during a commercial activity. Simultaneously, when moving beyond a general assessment, the Recitals do contain certain conflicting statements. These statements might hinder clear answers to future questions surrounding the position of open-source software under the Cyber Resilience Act.

E. Specific provisions for open-source software within the Cyber Resilience Act

60 The Cyber Resilience Act does not only regulate open-source software developers to improve the cybersecurity of open-source software. The Act also prescribes specific rules for 'open-source software stewards', proprietary software developers, and other parties with the aim of improving the overall cybersecurity of the open-source software ecosystem.

I. Open-source software stewards

61 In the open-source software community, there are certain organisations that support the development of open-source software as part of their overall mission statement. In some cases, these organisations also develop core open-source software. An example of such an organisation is the Python Software Foundation, which aims to advance the Python programming language and its community. The foundation organises conferences, offers grants to developers, and "produces the core Python distribution".[96] Python is a core programming language for software worldwide; it ranks second, after JavaScript, in a recent study from Github on the open-source software hosted on their platform.[97] The Python Software Foundation thus offers core support to the open source community, both through development and support.

62 The Cyber Resilience Act addresses organisations such as the Python Software Foundation as 'open-source software stewards'.[98] A steward is a legal person that provides systematic support for the development of open-source software, which is *intended for commercial activities*, as part of their overall objectives.[99] Importantly, the definition states that a steward is *not* a manufacturer.

63 Open-source software stewards receive a special position within the supervision scheme of the Cyber Resilience Act. Stewards are subject to a "light-touch and tailor-made regulatory regime".[100] The idea behind this scheme seems to be that open-source software stewards are vital to the continuation of the open-source ecosystem; the legislators believe they have a "main role in ensuring the viability of [open-source software]".[101]

64 Open-source software stewards have several obligations.[102] First, stewards must put in place cybersecurity policies for secure development of open-source software and vulnerability handling by the developers of that software.[103] The Python Foundation, for instance, has a vulnerability handling system where users can contact the 'Python Security Response Team' for support.[104] Stewards cannot be fined for non-compliance with these obligations,[105] but they can be required to take certain corrective

[94] Recital 15 CRA.

[95] Art 26(2)(a) CRA.

[96] <https://www.python.org/psf/mission/>.

[97] Kyle Daigle and GitHub Staff, 'Octoverse: The State of Open Source and Rise of AI in 2023' (*The GitHub Blog*, 8 November 2023) <https://github.blog/news-insights/research/the-state-of-open-source-and-ai/> accessed 13 August 2024.

[98] Recital 19 CRA: 'open-source software stewards include certain foundations[.]'

[99] Art 3(14) CRA.

[100] Recital 19 CRA.

[101] Recital 19 CRA.

[102] Art 24 CRA.

[103] Art 24(1) CRA.

[104] <https://www.python.org/dev/security/>.

[105] Art 64(10)(b) CRA.

actions.[106] This exemption also means that stewards cannot affix a CE-mark to their product.[107]

65 Stewards must also co-operate with market surveillance authorities to mitigate vulnerabilities in open-source software packages.[108] Market surveillance authorities are responsible for taking corrective measures when developers do not comply with the rules of the Act. This co-operation seems to be the essence of the steward role: providing communication between the open-source community and authorities in cases such as Log4j. In that line, it is logical that open-source software stewards provide support for software *with commercial intent*, meaning integration into proprietary products or services.[109] Through commercial integration, these software packages – and their vulnerabilities – have considerable influence on the global software ecosystem.

66 Finally, there are obligations for stewards that are also involved with development of open-source software.[110] They must also comply with certain notification obligations for developers, particularly the notification of actively exploited vulnerabilities.[111] However, as open-source software stewards are *not* manufacturers per the definition in Article 3(14), they do not have equal obligations to traditional manufacturers. Article 24(3) only lists notification obligations.

67 It is imaginable that a strict delineation between open-source software stewards and manufacturers is not feasible in practice. Stewards, such as the Python Foundation, also develop software. Are such stewards then manufacturers for that software independently – assuming the software is supplied in the course of a commercial activity – or are they stewards – and thus *not* manufacturers – for both providing support and developing products? As described above, in the former they must comply with the Act's many obligations for manufacturers, while in the latter they only carry the notification obligations of Article 24(3).

68 As with the Recitals above, the Commission may provide some answers to the role of open-source software stewards when it publishes guidance on the application of the Cyber Resilience Act.[112] Moreover, regulators could eventually solve such conflicts through the 'tailor-made' regulatory regime for open-source software stewards.

II. Proprietary manufacturers using open-source software

69 The Cyber Resilience Act applies to manufacturers of software and hardware products. This scope means that proprietary manufacturers are also responsible for improving open-source software cybersecurity, through several ways.

70 First, the Cyber Resilience Act inherently applies the broad applicability of the Act means that proprietary software – and proprietary software developers – must adhere to certain cybersecurity requirements. Since open-source software is virtually always part of proprietary software, the requirements for the proprietary software package inherently involve the underlying open-source software.

71 This connection between the cybersecurity of the proprietary package and the open-source software is made explicit in the Act. The Cyber Resilience Act requires manufacturers to exercise due diligence when integrating third-party components, including open-source components, into their own product.[113] This obligation seems to stem from cases such as the Log4j vulnerability, in which a vulnerability in an open-source component puts the entire (proprietary) software package at risk.

72 When exercising this due diligence, manufacturers may discover certain vulnerabilities. If a manufacturer identifies a vulnerability within an open-source component of their own software, they must, under the Act, report it to the open-source developers.[114] The manufacturers must also remediate the vulnerability according to the vulnerability handling requirements of the Act.[115] If, as part of this remedy, the manufacturers modify the code or hardware to address the vulnerability, they must also share this code with the open-source developer.

73 Other parties may help identify and remediate vulnerabilities in open-source software through voluntary security attestation programmes.[116] The Commission can set-up such a programme through delegated acts. These programmes strive to improve the overall cybersecurity of open-source software which is exempted from the scope of the Cyber Resilience Act.[117] The exact content of a security attestation programme, i.e. if the Commission

[106] Art 52(3) CRA.

[107] Recital 19 CRA.

[108] Art 24(2) CRA.

[109] Recital 19 CRA.

[110] Art 24(3) CRA.

[111] Art 24(3) & 14(1) CRA.

[112] Art 26(2)(a) CRA.

[113] Art 13(5) CRA.

[114] Art 13(6) CRA.

[115] Art 13(6) & Annex I Part 2 CRA.

[116] Art 25 CRA mentions 'developers or users' of open-source software and 'other third parties'. See also Recital 21.

[117] Recital 21 CRA speaks of open-source software 'not subject to the essential requirements' of the Act.

provides financial or organisational support, is not clear from the provisions.

74 The due diligence obligation and the voluntary security attestation programmes help to expand the parties which support the cybersecurity of open-source software packages.

F. Cybersecurity and open-source software: a problem solved?

75 There is a fine balance between enhancing open-source software cybersecurity and regulating the open=source ecosystem which may rely on ad-hoc and voluntary work. The Cyber Resilience Act shows how delicate this balance is, with its many exemptions and categorisations of open-source software, to ensure that only software supplied within the course of a commercial activity is regulated. The question is then whether these considerations achieve a balance between mitigating cybersecurity risks of open-source software and introducing feasible legal obligations for the sector.

76 A project like Log4j, for instance, does not fall under the scope of the Cyber Resilience Act. The project does not charge a price for the software nor conducts any activities explicitly listed as commercial in the Cyber Resilience Act. The project is merely supported by certain donators and commercial entities, which are both explicitly exempted as commercial activities.[118] Most likely, Log4j itself would, therefore, not fall within the scope of the Cyber Resilience Act. The only cybersecurity obligations related to Log4j exist for entities who integrate Log4j into their own proprietary software.

77 On a general level, the Cyber Resilience Act is a step in the right direction for cybersecurity, regardless of the rules imposed on open-source software. Many cybersecurity requirements introduced by the Act were not present in existing legislation.[119] The Act thus, at minimum, might improve the cybersecurity of proprietary software, even if it would not cover open-source software.

78 In the specific context of open-source software, the Act aims to balance between improving cybersecurity of open-source software while not discouraging open-source software development. Broadly speaking, the Act only covers ‘commercial’ open-source software. Many types of open-source software are non-commercial, as evident by the Recitals, which means that most open-source software is not regulated by the Cyber Resilience Act. The balance seems, thus, to fall in favour of alleviating regulatory pressure on open-source software developers, instead of (fully) improving open-source software cybersecurity. However, the cybersecurity side is also supported by the responsibilities imposed on integrators of open-source software and the voluntary security attestation programmes.

79 In sum, the Cyber Resilience Act aims to make open-source software more secure than it is currently, without imposing responsibilities on developers that may discourage further open-source software development. The legislation certainly emphasizes not discouraging the development, but responsibilities on both developers and users of open-source software will likely help improve its cybersecurity.

G. The future of open-source software under EU law

80 The Cyber Resilience Act is the first piece of legislation that aims to strike a balance between responsibilities for open-source software and supporting its ecosystem.[120] This means that the legislative choices made in the Act will have consequences for the future of open-source software under EU law. However, the Cyber Resilience Act includes many of its considerations for open-source software in the Recitals. This legislative choice has two consequences: 1) there is no clear embedded legal framework for open-source software in the Cyber Resilience Act, due to the applicability of the Recitals and 2) many of the considerations are specific to the current landscape of open-source software and therefore overly restrictive when considering future developments.

81 Recitals only have legal power insofar as the Court of Justice of the European Union and supervisory authorities use them to interpret the provisions of the Cyber Resilience Act. In 1998, the Court held that “the preamble to a Community act has no binding legal force and cannot be relied on as a ground for derogating from the actual provisions of the act in question.”[121] Recitals, therefore, can be useful for interpretation of ambiguous legal provisions (e.g., supplying in the course of a commercial activity)

[118] Based on the assumption that the donations do not exceed the project’s maintenance costs. For further information, see <https://logging.apache.org/log4j/2.x/support.html>.

[119] Pier Giorgio Chiara, ‘The Cyber Resilience Act: The EU Commission’s Proposal for a Horizontal Regulation on Cybersecurity for Products with Digital Elements: An Introduction’ [2022] International Cybersecurity Law Review.

[120] Colonna (n 54).

[121] Case C-162/97 *Nilsson and others* ECLI:EU:C:1998:554, [1998] ECR I-7477, para 54.

but are not separate legal provisions on which the Court will rely.

82 In addition, the Recitals are very specific and pinpoint different commercial modes within the current landscape of open-source software. Future developments may fall outside the scope of the current Recitals. For instance, a developer could place advertisements in their software, based on user consent to see them. These advertisements allow the developer to continue working full-time on the project and similar projects. Would this choice constitute an "intention to monetise",[122] which places the project inside the course of a commercial activity? Or is this just a circumstance under which "the development has been financed",[123] although the developer also uses the money to work on other projects? European consumer law tackles this problem for 'information society services' by stating that they are "provided for remuneration".[124] 'Remuneration' is a broad concept which involves advertisement income, but also the request for personal data by the service, as in the Cyber Resilience Act.[125] In comparison, the Cyber Resilience Act's notion of a commercial activity then seems overly restrictive, while a concept such as 'for remuneration' more easily adapts to future developments.

83 It seems that the Cyber Resilience Act's approach of placing virtually all considerations for open-source software in the Recitals might make the Act particularly vulnerable to future developments. This focus on the existing landscape, combined with the difficult method for assessing commerciality as described in Section D, may impair the applicability of the Cyber Resilience Act in the future. An embedded legal framework for open-source products in product legislation, which could also adapt to future developments, remains missing.[126]

H. Conclusion

84 This paper analysed the position of open-source software in the Cyber Resilience Act. The paper answered the following question: To what extent does the Cyber Resilience Act impose responsibilities on open-source software developers that achieve a balance between stimulating open-source software development and, simultaneously, mitigating cybersecurity problems within open-source software?

85 Open-source software stems from a unique development culture aimed at distributing knowledge freely. Simultaneously, the software is crucial for the modern digital infrastructure. As with any software, there are certain cybersecurity risks inherent in open-source software. The European Union aims to mitigate some of those risks through the Cyber Resilience Act.

86 The Cyber Resilience Act aims to regulate cybersecurity risks without discouraging open-source software development. The Act achieves this balance by covering only open-source software 'supplied in the course of a commercial activity'. The Act also introduces several other mechanisms to support the cybersecurity of open-source software. First, the Act prescribes a special regulatory regime to open-source software stewards, legal persons who support and advance the open-source software ecosystem. Second, proprietary manufacturers may only integrate open-source software components in a diligent manner. Therefore, they must also fix vulnerabilities discovered in open-source components and share such fixes with the developers of the component. Through voluntary security attestation programmes, the Act also supports other parties interested in advancing open-source software cybersecurity.

87 At the same time, the Recitals contain complex legal terminology. The Recitals mention many modes of financing and development of open-source software and if those modes are 'supplying in the course of a commercial activity'. However, the Recitals also note that an assessment of a project based merely on financing or development modes is not sufficient. It is currently unclear how this situation should be resolved in practice when an open-source project is neither an explicitly included nor excluded commercial activity.

88 The Cyber Resilience Act, however, does certainly advance cybersecurity of open-source software compared to the current regulatory landscape. Through rules for proprietary integration, proprietary software developers are also responsible for the cybersecurity of open-source software. Such rules mean that, even when a project is exempted

[122] Recital 15 CRA.

[123] Recital 18 CRA.

[124] Art 1(b) Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules of Information Society services.

[125] Recital 18 Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

[126] See also Colonna on the role of open-source software in the new Product Liability Directive and the AI Act, Colonna (n 54).

from the CRA's scope, it will receive cybersecurity support through the Act's obligations on other parties.

89 The future position of open-source software under EU law remains somewhat unclear after the Cyber Resilience Act, especially since so many of its considerations for open-source software occur in the Recitals. In sum, the Cyber Resilience Act achieves a balance between encouraging open-source software development and mitigating cybersecurity risks within open-source software, but some key challenges remain for the future.

Technical Challenges of Rightsholders' Opt-out From Gen AI Training after Robert Kneschke v. LAION [1]

by **Stepanka Havlikova** *

Abstract: This paper explores the evolving legal landscape surrounding generative AI model training on publicly available - often copyrighted - data, spotlighting the challenges in the wake of recent decision of German Court in Robert Kneschke v. LAION. On top of already explored implementation of copyright reservations by machine-to-machine and human-to-machine communication, this paper explores potential gaps and technical challenges stemming from the text and data mining exception including technical issues surrounding Robots.txt as well as data memorisation and regurgitation of verbatim snippets in AI outputs.

The Robert Kneschke v. LAION case exemplifies how non-profit organizations may leverage the TDM exceptions and offers insights that could influence commercial development of Gen AI. While the TDM exceptions may seem workable in theory, implementing them in practice presents a variety of practical challenges. Practical implications, such as requirements for "machine-readable" opt-out options for rightsholders considering current technological landscape, may ultimately reduce the practical benefits of these exceptions. Dataset creation and AI model training in practices occurs via chain of parties from copyright holders, licensors or publishers, non-profit organisations populating datasets to commercial AI developers which may bring additional interpretational issues and gaps when applying exception for research purposes or searching for validly applied opt-out. This paper discusses legal requirements and interpretation introduced by Robert Kneschke v. LAION and presents practical and technical implications stemming from the TDM exceptions and suggests possible outcomes thereof.

Keywords: Artificial Intelligence, Web Scraping, Text and Data Mining, Machine-readable, Copyright

Recommended citation: Stepanka Havlikova, Technical Challenges of Rightsholders' Opt-out From Gen AI Training after Robert Kneschke v. LAION, 16 (2025) JIPITEC XX para 1

[1] LG Hamburg, Urteil vom 27. September 2024 – 310 O 227/23 (Robert Kneschke v. LAION).

[*] PhD Candidate at the Institute of Law and Technology at Masaryk University and a Senior Associate at Dentons Law Firm. I thank Pavel Koukal, Jacopo Ciani Sciolla, Massimo Durante, Alessandro Cogo and Péter Mezei, for their feedback and helpful suggestions either on various drafts of this paper or ideas presented therein. This article is the result of the project of the Grant Agency of the Czech Republic [Copyrighted Works and the Requirement of Sufficient Precision and Objectivity (GA22-22517S)].

A. Introduction

1 During the preceding months we can see a significant rise of lawsuits in the United States based on copyright infringement[3] in connection with generative artificial intelligence[4] and scraping of large amounts of publicly available information to train artificial intelligence.[5] As the Economist recently pointed out in its article addressing copyright and artificial intelligence, "*it is the oceans of copyrighted data the bots have siphoned up while being trained to create humanlike content*" while "*often, it is alleged, AI models plunder the databases without permissions*".[6] Lemley and Casey noted that this may well be one of the most important legal questions of the coming century: *Will copyright law allow robots to learn?*[7] It may be only question of time whether and when similar cases are initiated in the EU, especially in connection with the Representative Action Directive[8] currently being implemented across the EU[9] while at the same time heavily supporting AI development and launch across the EU.[10] Considering the broad interpretation of the concept of reproduction[11] (for copyright) and extraction[12] (for database rights) under EU law, scraping publicly available copyright (or database) protected content may indeed constitute copyright or database right infringements,[13] unless rightsholders grant their authorisation or statutory exception applies.[14]

2 When considering potential development of similar cases under EU law, recently adopted set of two

[3] For example the Author's Guild claims that OpenAI's and Microsoft's AI models were "*trained," .. by reproducing a massive corpus of copyrighted material, including, upon information and belief, tens or hundreds of thousands of fiction and nonfiction books*" and that ¨*the only way that Defendants' models could be trained to generate text output that resembles human expression is to copy and analyze a large, diverse corpus of text written by humans*". With this argumentation the plaintiffs are requesting the defendants namely to cease using the infringing content and to provide financial compensation for past infringements. Brown, T.T., et al., Language Models are Few-Shot Learners. Available at: https://arxiv.org/pdf/2005.14165 [Accessed on 31.12.2024].

[4] Hereinafter also abbreviated to Gen AI.

[5] Cases filed before U.S. District Courts in 2023 against various global AI tools suppliers: Getty Images, Inc. v. Stability AI Ltd, U.S. District Court for the District of Delaware; Sarah Andersen v. Stability AI Ltd, U.S. District Court for the Northern District of California; Authors Guild v. Open AI, U.S. District Court for the Southern District of New York; Chabon v. OpenAI Inc., U.S. District Court for the Northern District of California; Richard Kadrey v. Meta Platforms, Inc., in the U.S. District Court for the Northern District of California; Sarah Silverman v. OpenAI, Inc., U.S. District Court for the Northern District of California.

[6] 'A battle royal is brewing over copyright and AI', The Economist [online], 2023. Available at: https://www.economist.com/business/2023/03/15/a-battle-royal-is-brewing-over-copyright-and-ai [Accessed on 31.12.2024].

[7] Lemley, M.A. and Casey, B., 2020. Fair Learning. Available at SSRN: https://ssrn.com/abstract=3528447 or http://dx.doi.org/10.2139/ssrn.3528447 [Accessed on 31.12.2024].

[8] Directive (EU) 2020/1828 of the European Parliament and of the Council of 25 November 2020 on representative actions for the protection of the collective interests of consumers and repealing Directive 2009/22/EC (the "Representative Actions Directive").

[9] In accordance with deadline for implementation by 25 June 2023.

[10] EU's long-term digital strategies identify the uptake of artificial intelligence as one of the objectives of the Digital Decade Policy Programme 2030. Artificial intelligence was named as one of the technologies (along with cloud computing and big data) which at least 75 % of Union enterprises should take up by 2030 (as part of the digital transformation of businesses which forms one of the digital targets in the Union); See Art. 4 (1) (3) Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030, Available at: https://eur-lex.europa.eu/eli/dec/2022/2481/oj [Accessed on 31.12.2024]. The 2021 Coordinated Plan on Artificial Intelligence explicitly highlighted that "*availability of high-quality data, among other things, in respect of diversity, nondiscrimination, and the possibility to use, combine and re-use data from various sources in a GDPR compliant way are essential prerequisites and a precondition for the development and deployment of certain AI systems*". See the 2021 Coordinated Plan on Artificial Intelligence; Available at: https://digital-strategy.ec.europa.eu/en/policies/plan-ai

[11] *Infopaq International A/S v. Danske Dagblades Forening*, Judgment of the Court of Justice dated 16.07.2009 in case C-5/08.

[12] *Innoweb BV v. Wegener ICT Media BV*, Wegener Mediaventions BV. Judgment of the Court of Justice dated 19.12.2013 in case C-202/12. *CV-Online Latvia SIA v Melons SIA*. Judgment of the Court of Justice dated 3.6.2021 in case C-762/19.

[13] Canellopoulou-Bottis, M., Papadopoulos, M., Zampakolas, C., and Ganatsiou, P., 2019. 'Text and Data Mining in Directive 2019/790/EU Enhancing Web-Harvesting and Web-Archiving in Libraries and Archives', Open Journal of Philosophy, p. 378.

[14] R. Ducato and A. Strowel, 'Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to "Machine Legibility"', CRIDES Working Paper Series (2018) 10.13140/RG.2.2.15392.84482. Available at SSRN: https://ssrn.com/abstract=3278901 or http://dx.doi.org/10.2139/ssrn.3278901 [Accessed on 31.12.2024]. Okediji, R., 2017. Copyright Law in an Age of Limitations and Exceptions. Cambridge: Cambridge University Press. ISBN 978131645090.

exceptions from copyright and database protection[15] for purposes of so-called "*text and data mining*"[16] introduced by the CDSM Directive[17] could emerge as pivotal when aiming to justify use of publicly available data to train artificial intelligence.[18] Existing case law addressing web scraping from various perspectives could also play significant role highlighting that scraping may lead to additional legal consequences such as unfair competition or free riding.[19]

3 Both TDM Exceptions are associated with legal uncertainties whereas some questions have been addressed by the recent decision of the German court in *Robert Kneschke v. LAION.*[20] In Robert Kneschke v. LAION German Hamburg Regional Court recently ruled on a lawsuit filed by German Photographer Robert Kneschke against the nonprofit organisation LAION which created a dataset consisting of image-text pairs subsequently used to train AI which included Kneschke's photos. The case against LAION was dismissed on the grounds of the scientific research TDM exception. Surprisingly, despite the fact the case was in fact dismissed based on TDM exception under Art. 3 CDSM Directive, significant part of the *obiter dictum* was dedicated to the court's view on TDM exception under Art. 4 CDSM Directive.

B. Applying TDM Exception on Gen AI Training

4 TDM Exceptions introduced by the CDSM Directive allow reproductions and extractions of protected content to carry out text and data mining defined as an "*automated analytical technique aimed at analysing text and data in digital form in order to generate information*".[21] Although scholars tend to agree TDM exceptions may serve as a suitable legal basis to justify use of data for generative AI training,[22] there are debates[23] to which extent did the development of artificial intelligence form a ratio behind enacting the TDM exceptions.[24]

[15] And press publisher rights.

[16] Text and data mining (further referred to as "TDM" and Text and Data Mining Exception under Art. 4 of the CDSM Directive also referred to as "TDM Exception").

[17] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (hereinafter referred to as the "CDSM Directive").

[18] CDSM Directive introduces two exceptions or limitations allowing (i) text and data mining for the purpose of scientific research under Art. 3 CDSM Directive and (ii) text and data mining for other purposes unless reserved by rightsholders under Art. 4 of the CDSM Directive. Art. 3 of the CDSM Directive introduces an exception from reproduction rights under copyright protections, extraction rights under sui generis database protections and press publisher rights for reproductions and extractions of lawfully accessible works and other subject matters for the purposes of text and data mining for research purposes. Art. 4 of the CDSM Directive introduces an exception from reproduction rights under copyright protections, extraction rights under sui generis database protections and press publisher rights for reproductions and extractions of lawfully accessible works and other subject matters for the purposes of text and data mining, if such rights have not been expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.

[19] See for example Pagallo U., Ciani Sciolla J., Anatomy of web data scraping: ethics, standards, and the troubles of the law. European Journal of Privacy Law & Technologies, (2023) 2 p. 1 - 19, available at: https://doi.org/10.57230/EJPLT232PS. [Accessed on 31.12.2024]. Due its limited extent, these consequences are excluded from the scope of this paper.

[20] LG Hamburg, Urteil vom 27. September 2024 – 310 O 227/23 (Robert Kneschke v. LAION).

[21] See footnote 18

[22] Mezei, Péter, A saviour or a dead end? Reservation of rights in the age of generative AI (January 15, 2024). European Intellectual Property Review, 2024, 46(7), p. 461-469. Available at SSRN: https://ssrn.com/abstract=4695119 or http://dx.doi.org/10.2139/ssrn.469511. [Accessed on 31.12.2024]. Novelli, Claudio and Casolari, Federico and Hacker, Philipp and Spedicato, Giorgio and Floridi, Luciano, Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity (January 14, 2024). Available at SSRN: https://ssrn.com/abstract=4694565 or http://dx.doi.org/10.2139/ssrn.4694565. [Accessed on 31.12.2024].Rosati, E., Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790, Oxford University Press, Oxford 2021. ISBN: 9780198858591. P. 72. Dusollier, Séverine, 'The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition' (2020) 57 Common Market Law Review 979, 984. Hanjo Hamann, Artificial Intelligence and the Law of Machine-Readability: A Review of Human-to-Machine Communication Protocols and their (In)Compatibility with Art. 4(3) of the Copyright DSM Directive, 15 (2024) JIPITEC 102 para 1.

[23] EU accused of leaving 'devastating' copyright loophole in AI Act', The Guardian [online], 2025. Available at: https://www.theguardian.com/technology/2025/feb/19/eu-accused-of-leaving-devastating-copyright-loophole-in-ai-act [Accessed on 20 March 2025].

[24] TDM exception introduced under Art. 4 CDSM Directive was not part of the Commission Proposal of the CDSM Directive which aimed to introduce solely exception for text and data mining for purposes of scientific research with no text and data mining exception for other purposes. TDM Exception - currently under Art. 4 – was subsequently proposed during the legislative procedure by the Committee on

5 Interestingly, the Commission Proposal of the CDSM Directive aimed to introduce solely the TDM Exception for purposes of scientific research.[25] Non-research TDM exception[26] was subsequently proposed during the legislative procedure by the Committee on Legal Affairs (JURI) and supported by the Parliament and the Council. For example, with the argumentation that "*this type of permitted use was not conceived for artificial intelligence*" the initial Polish legislative proposal for implementing the CDSM Directive included a controversial provision explicitly excluding the creation of generative AI models from the scope of the exceptions – which however did not stand and the final adopted law departed from this proposal and instead closely aligned with the original text of the CDSM Directive.[27] Although sometimes used as an argument against the applicability of the TDM Exception on AI training, such an interpretation was rejected by many scholars[28] as well as German court in *Robert Kneschke v. LAION*.[29] Lastly, the AI Act explicitly references the TDM exception in the context of training general-purpose AI models, underscoring that the exception might indeed be applicable when using protected content for AI training.[30]

6 The TDM Exception under Art. 3 CDSM Directive is limited to research organisations and cultural heritage institutions to carry out text and data mining for the purposes of scientific research and thus cannot be relied on by commercial companies scraping data to develop Gen AI (the interplay between Art. 3 and Art. 4 CDSM Directive will be further debated below). On the contrary, TDM exception under Art. 4 CDSM Directive is not limited by research purposes by research organisations – however applies only insofar such rights have not been "*expressly reserved by their rightsholders in an appropriate manner, such as machine-readable means*

Legal Affairs (JURI) and supported by the Parliament and the Council. See Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market, COM/2016/0593 final - 2016/0280 (COD). Rosati, E., Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790, Oxford University Press, Oxford 2021. ISBN: 9780198858591. P. 65. Mezei, Péter, A saviour or a dead end? Reservation of rights in the age of generative AI (January 15, 2024). European Intellectual Property Review, 2024, 46(7), p. 461-469. Available at SSRN: https://ssrn.com/abstract=4695119 or http://dx.doi.org/10.2139/ssrn.469511. [Accessed on 31.12.2024]. Report on the Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market (COM (2016)0593 – C8-0383/2016 – 2016/0280(COD)) (Rapporteur: MEP Axel Voss), Amendment 65. Dusollier, S., 'The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition' (2020) 57 Common Market Law Review 979, 984. Jan Bernd Nordemann and Jonathan Pukas, 'Copyright Exceptions for AI Training Data – Will There Be an International Level Playing Field?' (2022) 17 Journal of Intellectual Property Law & Practice 973, 974. Hajo Hamann, 'Artificial Intelligence and the Law of Machine-Readability: A Review of Human-to-Machine Communication Protocols and Their (In)compatibility with Art. 4(3) of the Copyright DSM Directive' (2024) 15(2) JIPITEC 102, 105–106.

[25] Currently Art. 3 CDSM Directive.

[26] Currently Art. 4 CDSM Directive.

[27] Draft implementation law published by polish Government for consultation. Available at: https://legislacja.rcl.gov.pl/projekt/12382002. [Accessed on 31.12.2024].

[28] Mezei, Péter, A saviour or a dead end? Reservation of rights in the age of generative AI (January 15, 2024). European Intellectual Property Review, 2024, 46(7), p. 461-469. Available at SSRN: https://ssrn.com/abstract=4695119 or http://dx.doi.org/10.2139/ssrn.469511. [Accessed on 31.12.2024]. Novelli, Claudio and Casolari, Federico and Hacker, Philipp and Spedicato, Giorgio and Floridi, Luciano, Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity (January 14, 2024). Available at SSRN: https://ssrn.com/abstract=4694565 or http://dx.doi.org/10.2139/ssrn.4694565. [Accessed on 31.12.2024]. Rosati, E., Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790, Oxford University Press, Oxford 2021. ISBN: 9780198858591. P. 72. Dusollier, Séverine, 'The 2019 Directive on Copyright in the Digital Single Market: Some Progress, a Few Bad Choices, and an Overall Failed Ambition' (2020) 57 Common Market Law Review 979, 984. Hanjo Hamann, Artificial Intelligence and the Law of Machine-Readability: A Review of Human-to-Machine Communication Protocols and their (In)Compatibility with Art. 4(3) of the Copyright DSM Directive, 15 (2024) JIPITEC 102 para 1.

[29] LG Hamburg, Urteil vom 27. September 2024 – 310 O 227/23 (Robert Kneschke v. LAION).

[30] Recital 105 of the AI Act confirms that the use of literary and artistic works for AI training purposes has copyright relevance and involves acts of text and data mining that require the authorisation of rightholders: "[a]*ny use of copyright protected content requires the authorisation of the rightholder concerned unless relevant copyright exceptions and limitations apply*" and subsequently refers to the TDM exception and notes that "*Where the rights to opt out has been expressly reserved in an appropriate manner, providers of general-purpose AI models need to obtain an authorisation from rightsholders if they want to carry out text and data mining over such works*". Also Mezei, Péter, The Multi-layered Regulation of Rights Reservation (Opt-out) Under EU Copyright Law and the AI Act -For the Benefit of Whom? (v1.0) (December 19, 2024). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5064018 [Accessed on 30.12.2024].

in the case of content made publicly available online".[31]

C. Practical Challenges Associated with Machine-Readable Opt-Out

7 The TDM exception under Art. 4 CDSM Directive faced criticism for its impracticality, particularly due to the rightsholders' opt-out mechanism. As Hugenholtz aptly observed, *the TDM provisions of the CDSM Directive secure considerably less freedom to text and data mine than they initially appear to do. The opt-out clause of Art. 4, in particular, leaves for-profit miners in the EU at the mercy of the content owners."*[32] However, the lack of standardization, ambiguity in how to properly implement the reservation, and technical challenges in decoding these measures introduce further complications including the question who sets the standards and what the level of "*machine-readability*" is expected from reservations. A critical question remains: who will bear the burden: rightsholders, AI companies, or end users?

I. Is "Machine-Readability" a Strict Requirement to Validly Opt-Out?

8 First question arises in connection with interpretation of the "*machine-readable*" requirement which is cited in connection with content made publicly available online. It is worth noting that some scholars are of the view that the machine-readability is not a strict requirement on how the reservation must be made but rather an example of how the reservation could be made – meaning that even non-machine-readable reservation could have legal effect if expressed by appropriate means.[33] This extensive interpretation could lead to the conclusion that *any* reservation expressed by rightsholders is valid if "*appropriate*". However, the absence of "*machine-readable*" form could undermine the sole purpose of the TDM exception of *allowing the automated computational analysis of information*[34] and text and data mining as an "*automated analytical technique aimed at analysing text and data in digital form*".[35] Some countries have not expressly implemented the machine-readability requirement in their national legislation and implemented solely "*appropriate means*" requirement – such as in Italy.[36] On the other hand, countries such as Germany, Austria, Slovakia or the Czech Republic make it clear that machine-readability forms a requirement making the opt-out ineffective if these conditions are not met.[37]

9 In the author's view, machine-readability should in fact be considered as a mandatory legal requirement to form a legally effective reservation from the TDM Exception.[38] This follows also from recitals of the CDSM Directive which states that "*In the case of content that has been made publicly available online, it should* <u>*only*</u> *be considered appropriate to reserve those rights by the use of machine-readable means, […]*" (emphasis added).[39] As a result, even the absence of explicit machine-readability requirement can be overcome by interpretation of the "*appropriate means*" requirement in light with the CDSM Directive.[40]

II. Interpretation of *"Expressly"* Reserved in *"Machine-Readable"* Form

10 The question remains how such "*machine-readable*" means shall be interpreted as CDSM Directive does not provide any legal definition thereof. According to Recital 18 of the CDSM Directive, such machine-readable means may include "*metadata and terms and conditions of a website or a service*".[41] Accordingly, machine-readable means could include for example technical restrictions and disallow commands[42] but

[31] Defined as „*automated analytical technique aimed at analysing text and data in digital form in order to generate information*"

[32] Hugenholtz, B. The New Copyright Directive: Text and Data Mining (Articles 3 and 4) [online]. Kluwer Copyright Blog. 2019. Available at: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/ [Accessed on 31.12.2024].

[33] Discussion held during International Conference Techno-legal challenges of data Scraping hosted at the the University of Turin, Department of Law in November 2023.

[34] Recitals 8 – 11 of the CDSM Directive.

[35] Art. 2 (2) CDSM Directive.

[36] Such as Italy. Section 70 of the Italian Copyright Act.

[37] Löbling, L., Handschigl, Ch. Hofman, K., Schwedhelm, J. Navigating the Legal Landscape: Technical Implementation of Copyright Reservations for Text and Data Mining in the Era of AI Language Models. 14 (2023) JIPITEC 499 para 14.

[38] The arguments for such interpretation are as follows. The beginning of the sentence starting with „such as" relates rather to the designation of „content made publicly available online" which requires as „appropriate means" the „machine-readable means". There may be other types of content not made publicly available online where the „appropriate means" are not specified by the CDSM Directive.

[39] Recital 18 CDSM Directive.

[40] *Costa v. ENEL*, Judgment of the Court of Justice in case 6/64.

[41] As the Recital 18 of the CDSM Directive states: For that has been made publicly available online, it should only be considered appropriate to reserve those rights by the use of machine-readable means, including metadata and the terms and conditions of a website or a service.

[42] Strowel, A., Ducato, R. Artificial Intelligence and Text

also reservations made via a website's terms of use provided they are in a machine-readable format.

11 By analogy, the Open-Data Directive defines machine-readable format of documents as "*a file format structured so that software applications can easily identify, recognise and extract specific data, including individual statements of fact, and their internal structure*". Nevertheless, the ratio behind Open-Data Directive significantly differs from the ration of Art. 4 CDSM Directive and thus it may not be suitable as *analogia legis*. As follows from Recital 35 of the Open Data Directive, "*A document should be considered to be in a machine-readable format if it is in a file format that is structured in such a way that software applications can easily identify, recognise and extract specific data from it. Data encoded in files that are structured in a machine-readable format should be considered to be machine-readable data."* While the Open-Data Directive aims to ensure access and reuse of public-sector information, the CDSM Directive aims to strike a balance between the interests of users of text and data mining (to be able to conduct automated analysis of data) and the interests of rights holders (to protect their rights). As a result, the requirement on machine-readability set forth by the Open-Data Directive is set as low as possible to ensure the easiest possible access of the public to the relevant information. However, setting the same benchmark for "*machine-readability*" under the CDSM Directive would mean shifting the balance significantly to the benefit of the users utilizing text and data mining. As a result, the definition of "*machine-readability*" under the Open-Data Directive cannot be relied on when interpreting the CDSM Directive.

12 Aim of the CDSM Directive is to allow the text and data mining which is defined as *"automated analytical technique [...]"* with the intention of making possible "*the processing of large amounts of information with a view to gaining new knowledge and discovering new trends*" and to "*analyse large amounts of data*".[43] German explanatory memorandum to Act amending the German Copyright Act (implementing the CDSM Directive) provides some guidance by emphasising that machine-readable reservation must enable automated processes because "*[...] the purpose of the regulation is to ensure that automated processes, which are typical criteria of text and data mining, can actually be automated in the case of content accessible online*".[44] Interestingly, the German Explanatory Memorandum mentions that the reservation can be included in the imprint of a given website (*Impressum*) or in its terms and conditions, *provided that it is machine-readable*.[45] On the contrary, Czech Explanatory Memorandum explained that the reservation may be easily implemented through standard metadata (e.g. by structuring the metadata to a format which automated tools are able to read) but noted that general statements on websites on in content terms of use are not a suitable mean to express the reservation.[46]

13 German court in *Robert Kneschke v. LAION* noted that while the term "*machine readability*" must be interpreted in light of the legislative intent underlying it — to enable automated queries by web crawlers — it should be understood in the sense of "*machine understandability*" whereas such question should always be answered based on the technical developments prevailing at the relevant time of use of the work. With reference to state-of-the-art technologies requirement stemming from the AI Act - which applies on providers of general-purpose AI models if intended to utilize TDM Exception - the court noted that "*these "state-of-the-art technologies" undoubtedly include, in particular, AI applications capable of comprehending text written in natural language".* The court further explained that CDSM Directive does not demand that a reservation needs to be declared "*in the simplest way possible*," but rather "*in an appropriate manner*" which suggests certain middle ground between the requirement of "*machine-readability*" enabling automated processes while at the same time granting the rightsholders the freedom to choose means available to them.[47]

and Data Mining: A Copyright Carol IN Rosati, E. The Routledge Handbook of EU Copyright Law. Ed. Eleonora Rosati. Abingdon. 2021. ISBN: 9780367436964. P. 30. Hugenholtz, B. The New Copyright Directive: Text and Data Mining (Articles 3 and 4) [online]. Kluwer Copyright Blog. 2019. Available at: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/ [Accessed on 31.12.2024].

[43] Recital 8 and 18 CDSM Directive.

[44] Explanatory memorandum (Gesetzesbegründung) of the German Government (Bundesregierung) to its legislative proposal implementing the CDSM Directive: Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesbegründung: Besonderer Teil. No. 19/27426. Page 95. Available at https://dip.bundestag.de/vorgang/.../273942 [Accessed on 31.12.2024].

[45] Ibid.

[46] Explanatory memorandum (Důvodová zpráva) of the Czech Government to the Act. No. 429/2022 Coll. (amending the Czech Copyright Act implementing the CDSM Directive). Section § 39c.

[47] However, it is very important to highlight that – as already mentioned above – the question of "machine-readability" was only tackled by the court in obiter dictum of the judgement whereas although the court shared its legal opinion on the question at hand, it also explicitly noted that whether the defendant can rely on the TDM exception under Art. 4 CDSM Directive "*does not need to be conclusively determined*" which slightly undermines the precedential weight of the argumentation.

14 The court applied a rather *pro-rightsholder* interpretation as it set the benchmark of "*machine-readability*" relatively low which however imposes very high demands on the users relying on the TDM Exception when decoding such reservations. The court has however not tackled the issue of potential unreliability of Gen AI which may prevent such users from consistently and reliably identifying reservations in all cases.[48] As a result, while such reservations may *in most cases* be indeed decoded by generative AI capable of understanding natural language, the accuracy of decoding is unlikely to be flawless (for example reliability will likely vary depending on the specific generative AI model[49] or language of the reservation[50]). This uncertainty exposes generative AI developers to legal risks of potential copyright infringements despite applying their best efforts and state-of-the-art technologies. On the other hand, the failure to adequately present a reservation in a machine-readable form with sufficient reliability should not disadvantage users relying on the TDM exceptions who might not be able to reliably decode such reservation despite applying state-of-the-art technologies but should rather go to the detriment of the rightsholders who have the power and control as to how they implement and express their reservations.

15 Although he rightsholders to set the tone of the "*appropriate means*" as they decide how to implement their reservations, the recently adopted AI Act[51] obliges the providers of so-called general-purpose AI models[52] to put in place a policy to comply with Union copyright law, and in particular to "*identify*" ... "*through state of the art technologies*". reservations of rights.[53] In *Robert Kneschke v. LAION,* German court used a reference to the AI Act while assessing whether publicly available declarations in human language may constitute a machine-readable exception.[54] In the author's view, the interplay with Art. 53 of the AI Act could offer a valuable solution for addressing challenges under Art. 4 of the CDSM Directive. While Art. 53 of the AI Act applies specifically to providers placing general-purpose AI models on the EU market and may not cover all providers of generative AI

[48] Not to mention that this interpretation creates a „chicken-and-egg" dilemma, as generative AI capable of understanding natural language cannot be developed without access to sufficiently broad high-quality datasets.

[49] Iorliam, Aamo & Ingio, Joseph. (2024). A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing. Journal of Computing Theories and Applications. Volume 2. 10.62411/jcta.9447.

[50] Reliability of Gen AI decoding the reservation may for example largely depend on language of the reservation as some Gen AI models have higher reliability in English language but lower reliability in other languages.

[51] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act or also AI Act).

[52] Defined in Art. 3 AI Act as "*an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market*".

[53] Providers of general-purpose AI models shall inter alia (i) draw up technical documentation (including also information on the data used for training, testing and validation and how the data was obtained and selected); (ii) put in place a policy to comply with Union copyright law, and in particular to identify and comply with, including through state of the art technologies, a reservation of rights expressed pursuant to Art. 4(3) CDSM Directive; and (iii) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model as follows from Art. 53 AI Act. These requirements shall apply within 12 Months after the AI Act comes into force. Finally, respecting opt-outs from the TDM exception is an explicit part of the GPAI model providers' obligation to comply with EU copyright law as follow from Art. 53(1)(c) of the AI Act. As a result, GPAI models trained with material in violation of valid opt-outs are not compliant with the AI Act and may not be put into service or placed on the market in the EU. Recital 106 of the AI Act further justifies the requirement by competition grounds while explaining the necessity to ensure a level playing field among providers of general-purpose AI models where no provider should be able to gain a competitive advantage by applying lower copyright standards. Therefore, we can expect that within the upcoming 12 months, remaining developers of generative artificial intelligence shall follow the trend set by OpenAI and shall introduce their recommendations on implementation of the reservation from the TDM exception which shall make it easier for rightsholders to effectively implement their reservations. Concurrently, new obligations of publishing a sufficiently detailed summary about the content used for training shall make it easier for rightsholders to establish unlawful use of their content in case the reservation has not been duly complied with.

[54] Specifically, the court assessed the question of whether and under what specific conditions a reservation of use expressed in "natural language" can also be considered "machine-understandable" and noted it must always be answered based on the technical developments prevailing at the relevant time of use of the work. Subsequently the court referred to "state-of-the-art technologies" under the AI Act and concluded that these "state-of-the-art technologies" undoubtedly include, in particular, AI applications capable of comprehending text written in natural language".

models utilizing copyright-protected content within the EU, the state-of-the-art technologies employed by these providers could set a precedent eventually influencing how courts interpret and apply Art. 4 CDSM Directive.

16 As explained above, the CDSM Directive aims to strike a balance between the interests of users of text and data mining (to be able to conduct automated analysis of data) and the interests of rights holders (to protect their rights). As a result, while users of text and data mining should indeed be expected to employ state-of-the-art technologies to decode reservations, rightsholders' "*express*" reservations in "*machine-readable*" formats should, in the author's view, achieve a *reliable level* of machine interpretability. This might require the reservation to be presented in a sufficiently binary form that enables such advanced technologies to *reliably* decode its content leaving no room for doubt. This may be reflected for example by a standardized formulas (despite being written in a natural human language) which could be for example similar to standardized open-source license terms. On the contrary, the author believes that vague terms and conditions generally prohibiting scraping or bot access without expressly invoking reservation of rights from the TDM Exception (mainly those applied prior to TDM Exceptions coming into effect) should in most cases in fact not be able to achieve the level of "*express*" reservation in "*machine-readable*" form fulfilling the required level of reliability of its decoding. For instance, in the case assessed by the German court, the plaintiff's reservation used a rather generic wording prohibiting *"use automated programs .. for purposes of ... scraping"* but did not expressly refer to text and data mining.[55] Moreover, the court noted that these terms were published on the websites as early as 13 January 2021, before the CDSM Directive was implemented in Germany on 20 May 2021. This timing suggests that the reservation may not have been intended to address the TDM exception. Such an interpretation, however, might conflict with the requirement for "*expressly*" reserving rights in "*machine-readable*" means, which in the author's view implies that a reservation should unequivocally be understood as the rightsholder's intention to prevent text and data mining, leaving no room for doubt.[56]

17 However, as of today, no such sufficiently unified language of such reservation exists despite some attempts to introduce unified formulas.[57] Such unification could be for example established by independent bodies having sufficient authority to influence the global market.[58]

III. Existing State-of-the-Art Technologies Enabling to Express Rightsholder's Opt-Out

18 Open Future research aptly differentiates between "*unit-based*" and "*location-based*" identifiers enabling to place the express rights reservations either at a high level, affecting all applicable content available for example under a given website, or reservations affecting each content item individually.[59] Among

[55] Specifically, the court referred to the following wording on the defendant's website: *"RESTRICTIONS: YOU MAY NOT: (...) 18. Use automated programs, applets, bots or the like to access the XXX.com website or any content thereon for any purpose, including, by way of example only, downloading content, indexing, scraping, or caching any content on the website."*

[56] For example, when requesting ChatGPT (version 4o) using various prompts to provide an answer whether the wording applied in the case at hand presents a valid reservation within the meaning of Art. 4 CDSM Directive, ChatGPT was not able to provide clear answer – out of three prompts, in one case ChatGPT responded positively, in one case provided vague answer and in one case responded negatively. Although this itself does not exclude the machine-readability, it somewhat underlines the possibility of Gen AI providing different conclusions. For example: Prompt: "*Website published terms and conditions containing wording below. Has the owner of the website expressly reserved its rights by machine-readable means under article 4 para 3 of the EU CDSM Directive? "RESTRICTIONS: YOU MAY NOT: (...) 18. Use automated programs, applets, bots or the like to access the XXX.com website or any content thereon for any purpose, including, by way of example only, downloading content, indexing, scraping, or caching any content on the website."* Answer: "*The wording you provided restricts the use of automated tools to access the website but does not seem to expressly reserve rights through machine-readable means, as required under Article 4(3) of the EU Directive 2019/790 on Copyright in the Digital Single Market (CDSM Directive). ... Based solely on the provided text, the website owner has not expressly reserved their rights under Article 4(3) by machine-readable means. To comply with the Directive, the owner would need to implement additional technical measures beyond this contractual language.*"

[57] Keller/Warso, 'Defning Best Practices for Opting Out of ML Training' (29 Sep 2023), OpenFuture Policy Brief #5; Available online at: www.openfuture.eu/wp-content/uploads/2023/09/Best-_practices_for_optout_ML_training.pdf [Accessed on 31.12.2024]. Keller/Warso, 'Considerations for Opt-out Compliance Policies' (16 May 2024), Open Future Policy Brief #6 (2024), available at https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024].

[58] For example, German Explanatory memorandum proposed to incorporate such wording to Impressum. Czech SPIR recommended standardized wording for website header.

[59] Keller/Warso, 'Considerations for Opt-out Compliance Policies' (16 May 2024), Open Future Policy Brief #6

those location-based identifiers is the mostly cited method of implementing the reservation from TDM exception is Robots.txt.[60] Alternatively, TDM fields in the HyperText Transfer Protocol (HTTP) Response header, TDM Metadata in HTML Content,[61] or various forms of access restrictions denying access to automated bots also come into consideration or expressions via terms and conditions of the website.[62] In addition, there can be numerous types of "*unit-based*" identifiers depending on type of content – for example TDM Metadata in EPUB files or metadata or watermarking of various types of media files.[63] Location-based identifiers are suitable mainly for those rightsholders who manage their own domains or sites, while those unit-based may be suitable for independent files especially when expecting subsequent spreading the respective files on the internet.[64]

19 Technical measures continuously evolve and will continue to evolve in the future. For example, Goole announced its plan to explore additional machine-readable means for web publishers[65] and Spawning AI created a Do Not Train registry and recently published the new option of ai.txt[66] which

(2024), available at https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024].

[60] Robots.txt is based on principles of good faith not technically preventing a robot from accessing the site, but merely expressing the intention not to allow automated robots access (primarily the case of Robots.txt or information embedded in the website header). Since the CDSM Directive solely requires that such reservation must (i) be machine-readable and (ii) express the rightsholder's will not to allow text and data mining, even voluntary expression should be sufficient. Nevertheless, the question whether voluntary measures can be considered as effectively expressing such reservation is controversial. Hugenholz names Robots.txt as a typical example of technical restrictions expressing reservation within the meaning of Art. 4 of the CDSM Directive, while Ducato and Strowel express arguments based on the InfoSoc Directive against such interpretation. Hugenholtz, B. The New Copyright Directive: Text and Data Mining (Articles 3 and 4) [online]. Kluwer Copyright Blog. 2019. Available at: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/ [Accessed on 31.12.2024]. R. Ducato and A. Strowel, 'Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to "Machine Legibility"', CRIDES Working Paper Series (2018) 10.13140/RG.2.2.15392.84482. Available at SSRN: https://ssrn.com/abstract=3278901 or http://dx.doi.org/10.2139/ssrn.3278901 [Accessed on 31.12.2024].

[61] See for example W3C TDMRep Final Community Group Report of 2 Feb 2024. Available at: https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/ [Accessed on 31.12.2024]. Keller/Warso, 'Considerations for Opt-out Compliance Policies' (16 May 2024), Open Future Policy Brief #6 (2024), available at https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024]. Hanjo Hamann, Artificial Intelligence and the Law of Machine-Readability: A Review of Human-to-Machine Communication Protocols and their (In)Compatibility with Art. 4(3) of the Copyright DSM Directive, 15 (2024) JIPITEC 102 para 1.

[62] Other measures, on the contrary, may directly block access to the given website when identifying automated crawlers through various bot-detection measures (namely CAPTCHA, browser challenges, browser fingerprinting, etc.) or enable access solely to verified human users accessing the content (namely password protections or similar access restrictions). Explicit denial of access to the given website e.g. by displaying error window (either after previous recognition of automated user based on bot-detection measures or after failure to pass log-in or registration path) could possibly also serve as a means of expressing such reservation within the meaning of Art. 4 of the CDSM Directive. However, implementation of these measures is not always user-friendly and desirable for the rightsholders. On the other hand, the sole implementation of bot-detection measures (for example CAPTCHA or browser challenges) without subsequently disabling access or expressing the intention not to grant such access in any way, could hardly have such legal relevance due to the absence of expression of rightsholder's will. There are further technical restrictions used to recognize bots and tactics aimed to make bot access more complicated, such as for example rate-limiting or crawl delay. However, since such measures solely to indirectly complicate bot access but do not clearly express the website holder's intention not to allow access via automated means, such could accordingly hardly have such legal relevance.

[63] See for example W3C TDMRep Final Community Group Report of 2 Feb 2024. Available at: https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/ [Accessed on 31.12.2024]. Open Future, Open Future policy brief #6: Considerations for opt-out compliance policies by AI model developers. Available at: https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024].

[64] Open Future, Open Future policy brief #6: Considerations for opt-out compliance policies by AI model developers. Available at: https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024].

[65] In June 2023, Google suggested an option to explore additional machine-readable means for web publishers and to attempt finding new alternatives to robots.txt in connection with artificial intelligence and other emerging technologies. A principled approach to evolving choice and control for web content. Google Blog. Available at: https://blog.google/technology/ai/ai-web-publisher-controls-sign-up/ [Accessed on 31.12.2024].

[66] Spawning is an independent third party that created a Do

was already cited by the French Data Protection Authority in terms of scraping publicly available personal data.[67] Other examples of such new means could be the TDM Reservation protocol (TDMRep)[68] or DeviantArt's noai meta-tags.

20 In addition, there may be other means specific for various member states within the EU. For example, German explanatory memorandum suggests that the reservation can be included in the imprint of a given website (*Impressum*) - which is a section typical for German websites – or terms and conditions, as long as such reservation is machine-readable.[69] As explained therein, the purpose and intention of the regulation is to give the rightsholders the opportunity to prohibit such use while at the same time ensuring that automated processes, which are a typical for text and data mining, can truly be carried out automatically for content that is accessible online.[70] Czech Association for Internet Development[71] recommends – besides Robots.txt – to place opt-out related wording to website footer which has been followed by some rightsholders in the Czech Republic.[72] French collective management society SACEM announced in its statement dated 12 October 2023 that it is opting out of machine learning training for the works in its repertoire.[73] Interestingly, Spanish Ministry of Culture and Sport recently published for public consultation a draft Royal Decree (*Proyecto de Real Decreto*) on Extended Collective Licensing introducing the idea of collective management of copyright-protected works in the development of AI models.

IV. Robots.txt and its Technical Limitations

21 Robots.txt is often cited as a typical example of technical restrictions expressing reservation within the meaning of Art. 4 of the CDSM Directive.[74] However, there are numerous practical constrains associated with using Robots.txt to express reservation from the TDM exception (especially for purposes of preventing use of data for generative AI training).

22 Robots.txt (or also called the *Robots Exclusion Protocol*) is a simple text file containing rules on which crawlers may access which parts of a site.[75] Robots.txt is based on voluntary basis meaning it does not technically block the automated access, but merely expresses the rules for access introduced by the given website. Robots.txt consists of set of rules stipulating the following information: (i) to whom the rule applies (the "user agent"); (ii) which directories or files that agent can access; and (iii) which directories or files that agent cannot access.[76] Interestingly, Robots.txt has been published in 1994[77] and *defacto* become

Not Train registry intended to provide machine readable opt-outs to AI model trainers.

[67] Commission nationale de l'informatique et des libertés (CNIL) ; La base légale de l'intérêt légitime: fiche focus sur les mesures à prendre en cas de collecte des données par moissonnage (web scraping); Guidance issued on 10 July 2024, Available at: https://www.cnil.fr/fr/focus-interet-legitime-collecte-par-moissonnage [Accessed on 31.12.2024].

[68] TDM Reservation Protocol (TDMRep); Available online at: https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/ [Accessed on 31.12.2024].

[69] Explanatory memorandum (Gesetzesbegründung) of the German Government (Bundesregierung) to its legislative proposal implementing the CDSM Directive: Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesbegründung: Besonderer Teil. No. 19/27426. Page 88. Available at https://dip.bundestag.de/vorgang/.../273942 [Accessed on 31.12.2024].

[70] Ibid.

[71] Czech Association for Internet Development – in *Czech* as Sdružení pro internetový rozvoj (abbreviated as "SPIR").

[72] SPIR press release: Online vydavatelé se vymezují proti vytěžování dat umělou inteligencí. [online]. Spir.cz. Available at: https://www.spir.cz/online-vydavatele-se-vymezuji-proti-vytezovani-dat-umelou-inteligenci/ [Accessed on 31.12.2024].

[73] Although in the author's view such CMO's declaration placed on its own website can hardly fulfill the requirements of a valid express reservation in machine-readable means (without appropriate legal basis in the law). SACEM press release: Pour une intelligence artificielle vertueuse, transparente et équitable, la Sacem exerce son droit d'opt-out. [online]. societe.sacem.fr. Available at: https://societe.sacem.fr/actualites/notre-societe/pour-une-intelligence-artificielle-vertueuse-transparente-et-equitable-la-sacem-exerce-son-droit [Accessed on 31.12.2024].

[74] Hugenholtz, B. The New Copyright Directive: Text and Data Mining (Articles 3 and 4) [online]. Kluwer Copyright Blog. 2019. Available at: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/ [Accessed on 31.12.2024]. As will be further outlined below, Robots.txt is currently recommended by key market players as a means to avoid being scraped in connection with AI training.

[75] As follows from the Google guidelines for developers accessible online at https://developers.google.com/search/docs/crawling-indexing/robots/robots_txt or also at http://www.robotstxt.org/robotstxt.html [Accessed on 31.12.2024]

[76] Google Developers: Introduction to Robots.txt. Available at: https://developers.google.com/search/docs/crawling-indexing/robots/intro [Accessed on 31.12.2024].

[77] The standard, initially RobotsNotWanted.txt, allowed web developers to specify which bots should not access their website or which pages bots should not access. The internet was small enough in 1994 to maintain a complete list of all

a standard shortly after. There are the following historical descriptions of Robots.txt.: (i) the original 1994 A Standard for Robot Exclusion document[78]; and (ii) a 1997 Internet Draft specification A Method for Web Robots Control[79], further expanded by standard RFC 9309 Robots Exclusion Protocol.[80] As David Pierce said for the Verge, "*GPTBot has become the main villain of robots.txt because OpenAI allowed it to happen*" whereas "*it did all of this after training the underlying models that have made it so powerful*".[81] However, Robots.txt is not further actively developed.[82] In terms of potential future development, Google, while officially supporting Robots.txt as the means of expressing bot access rules, last year noted via its VP of trust Danielle Romain that "*We recognize that existing web publisher controls were developed before new AI and research use cases .... We believe it's time for the web and AI communities to explore additional machine-readable means for web publisher choice and control for emerging AI and research use cases.*"[83]

1. Generally Prohibiting all Text and Data Mining via User Agent Line Blocking all Bot Access?

23 Robots.txt differentiates specific terms for selected users (in the "*User-agent*" line of the Robots.txt) and URLS which may or may not be accessed (in the "*Disallow/Allow*" line of the Robots.txt).[84] The default rule usually is that a user agent can crawl any page or directory not blocked by a disallow rule. However, by generally blocking all automated access via robots.txt, such website could prevent Google[85] and other search engines from accessing and indexing the given website or could negatively impact how such website appears in search results in search engines, which considering the functioning of the internet might an undesirable scenario. As a result, rightsholders only rarely choose to disallow all bot access via Robots.txt.

24 "*User-agent*" line of Robots.txt allows to apply different reservation on various users (e. g. by allowing Google to crawl and index a website and prohibiting specific crawlers to scrape the website). The rightsholder may choose a "*whitelist*" of crawlers who may access the site[86] or *vice versa* a "*blacklist*" of crawlers who may not access the site. Such approach could be a reasonable solution for rights holders. However, such approach requires knowing the list of whitelisted or blacklisted users and knowing how to specifically identify such users in the "*User-agent*" line (to establish the machine-readability of the information for potential bots accessing such Robots.txt).

25 Robots.txt however does not enable prohibiting a specific purpose or means of use, i.e. prohibit any kind of text and data mining by any crawlers. Theoretically the user agent line could also identify group of crawlers, nevertheless such approach makes it even more difficult to decode the Robots.txt and could thus prevent the machine-readability of the "*User-agent*" line. An example may be the recommendation of the Czech Association for Internet Development recommending adding "*Machine Learning*" to "*User-agent*" line to prohibit large language models from accessing the site for AI training.[87] This approach has been subsequently implemented by some rightsholders in the Czech Republic[88], however, there are no available data as to whether this approach has been followed by AI companies.

bots; server overload was a primary concern.

[78] A Standard for Robot Exclusion, document dated 30 June 1994 published at: http://www.robotstxt.org/orig.html [Accessed on 31.12.2024].

[79] A Method for Web Robots Control; document dated 4 December 1994; published at: http://www.robotstxt.org/norobots-rfc.txt [Accessed on 31.12.2024].

[80] Koster/Illyes/Zeller/Sassman, 'Standard RFC 9309: Robots Exclusion Protocol', as of Sep 2022, Available at: Rfc-editor.org/rfc/rfc9309.html [Accessed on 31.12.2024].

[81] Pierce, D. The text file that runs the internet. The Verge (2024) [online]. Available at: https://www.theverge.com/24067997/robots-txt-ai-text-file-web-crawlers-spiders [Accessed on 31.12.2024].

[82] What about further development of /robots.txt? Robots.org. [online]. Available at: http://www.robotstxt.org/faq/future.html [Accessed on 31.12.2024].

[83] Romain, D. A principled approach to evolving choice and control for web content. Google Blog. [online]. Available at: https://blog.google/technology/ai/ai-web-publisher-controls-sign-up/ [Accessed on 31.12.2024].

[84] Koster/Illyes/Zeller/Sassman, 'Standard RFC 9309: Robots Exclusion Protocol', as of Sep 2022, Available at: Rfc-editor.org/rfc/rfc9309.html [Accessed on 31.12.2024].

[85] As follows from Google guidelines for developers accessible online at https://developers.google.com/search/docs/crawling-indexing/robots/intro [Accessed on 31.12.2024].

[86] Nevertheless, rightsholders should, where sought, allow automated crawling through a website containing terms and conditions (especially where websites are protected by such technical restrictions) in order to enable a search through a website containing terms of use via automated means if the rightsholder wishes to apply these.

[87] SPIR press release: Online vydavatelé se vymezují proti vytěžování dat umělou inteligencí. [online]. Spir.cz. Available at: https://www.spir.cz/online-vydavatele-se-vymezuji-proti-vytezovani-dat-umelou-inteligenci/ [Accessed on 31.12.2024].

[88] See for example official Czech Press Agency under ctk.cz/robots.txt or also some Czech media platforms including idnes.cz/robots.txt. [Accessed on 31.12.2024].

26 In addition, there are numerous other practical constrains associated with proper implementation and proper decoding of rightsholder's opt out. For example, it is market-standard that crawlers search for Robots.txt solely on the top-level directory of a site.[89] However, the CDSM Directive does not introduce any such requirement and thus even files and information hidden in lower levels can be legally effective.

2. Identifying Scrapers in User-Agent Line?

27 Another issue associated with proper decoding of Robots.txt is the standardisation of its content as Robots.txt requires identification of the scraper in the User-agent line to be effectively implemented. However, it is the scrapers themselves who set their own name.[90] After strike of lawsuits in the USA, top AI market players have set the trend of publishing the recommended way to opt-out from their AI training.[91] This approach however requires rightsholders to monitor instructions published by all viable scrapers and currently also significantly disadvantages those AI developers, who take this step of proactively publishing their recommendations on their websites against those who do not do so (since as follows from the Originality.AI analysis explained below, websites tend to follow such recommendations and restrict use of their data to such user agents).

28 For example, on 7 August 2023 OpenAI published on its website a recommendation on how to disallow their GPTbot from accessing a website as follows:

> "*To disallow GPTBot to access your site you can add the GPTBot to your site's robots.txt:*
> *User-agent: GPTBot*
> *Disallow: /*"[92]

29 On 28 September 2023, Google announced a Google-Extended, a new control for web publishers[93] which enables to place "*Google-Extended*" to user-agent line of Robots.txt of rightsholder's websites to prevent its content to be used to train Bard (later re-named to Gemini) and Vertex AI generative APIs and future generations of models that power those products.

30 Common Crawl, non-profit foundation producing and maintaining an open repository of web crawl data,[94] published its recommended structure of Robots.txt to prevent Common Crawl from crawling a website and recommended implementing "*CCBot*" to the user-agent line.[95] According to a study published in 2020, OpenAI's GPT-3 was trained using data mostly collected from Common Crawl.[96] On the other hand, Common Crawl is used for a variety of other purposes unrelated to generative artificial intelligence.[97]

31 In June 2024, another key AI market player Anthropic AI[98], developer of large language model called Claude, published its recommendation for placing "*ClaudeBot*" to the user-agent line of Robots.txt.[99]

32 Shortly prior to the above, on 7 July 2023 Czech Association for Internet Development[100] issued a recommendation to rightsholders on how to implement the reservation from the TDM exception within Robots.txt as follows:

> "*User-agent: MachineLearning*

[89] See, for example, a recommendation in the Google guidelines for developers accessible online at https://developers.google.com/search/docs/crawling-indexing/robots/robots_txt [Accessed on 31.12.2024].

[90] Koster/Illyes/Zeller/Sassman, 'Standard RFC 9309: Robots Exclusion Protocol', as of Sep 2022, Available at: Rfc-editor.org/rfc/rfc9309.html [Accessed on 31.12.2024].

[91] No Robots(.txt): How to Ask ChatGPT and Google Bard to Not Use Your Website for Training. Electronic Frontier Foundation. [online]. Available at https://www.eff.org/deeplinks/2023/12/no-robotstxt-how-ask-chatgpt-and-google-bard-not-use-your-website-training [Accessed on 31.12.2024]. How to block AI crawlers with robots.txt. Netfuture. [online]. Available at https://netfuture.ch/2023/07/blocking-ai-crawlers-robots-txt-chatgpt/ [Accessed on 31.12.2024].

[92] GPTBot. Available at https://platform.openai.com/docs/gptbot [Accessed on 31.12.2024].

[93] An update on web publisher controls. Google Blog. [online]. Available at: https://blog.google/technology/ai/an-update-on-web-publisher-controls/ [Accessed on 31.12.2024].

[94] Common Crawl is a non-profit foundation founded with the goal of democratizing access to web information by producing and maintaining an open repository of web crawl data that is universally accessible and analyzable by anyone. Common Crawl, CCBot. [online]. Available at: https://commoncrawl.org/ccbot [Accessed on 31.12.2024].

[95] Common Crawl, CCBot. [online]. Available at: https://commoncrawl.org/ccbot [Accessed on 31.12.2024].

[96] Brown, T.T., et al., Language Models are Few-Shot Learners. Available at: https://arxiv.org/pdf/2005.14165 [Accessed on 31.12.2024].

[97] Common Crawl, Use cases. [online]. Available at: https://commoncrawl.org/use-cases [Accessed on 31.12.2024].

[98] Anthropic has developed a family of large language models (LLMs) named Claude as a competitor to OpenAI's ChatGPT and Google's Gemini.

[99] Does Anthropic crawl data from the web, and how can site owners block the crawler? [online]. Available at: https://support.anthropic.com/en/articles/8896518-does-anthropic-crawl-data-from-the-web-and-how-can-site-owners-block-the-crawler [Accessed on 31.12.2024].

[100] Czech Association for Internet Development – in *Czech* as Sdružení pro internetový rozvoj (SPIR) -

Disallow: /"[101]

33 As can be seen from Robots.txt implemented by some media companies[102], many have implemented these solutions recommended by these AI companies. In 2023, Originality.AI analysed the top 1000 websites in the world to identify which sites are already blocking GPTBot[103] and later added also the CCBot, Google-Extended bot and anthropic-ai. As of June 2024, OriginalityAI found that 350 out of the 1000 websites, i.e. 35 %, block GPTBot, 216 out of the 1000 websites, i.e. 21,60% block CCBot, 126 out of the 1000 websites, i.e. 12.60 % block Google-Extended bot and 84 websites out of 1000 websites, i.e. 8.40% block anthropic.ai. As Originality.AI originally noted, "*it is not clear if "anthropic-ai" and "claude-web" would be effective as there has been no documentation from Anthropic*." (although in the meantime Anthropic published its recommendation). [104]

34 As a result, technical limitations of Robots.txt solution inevitably lead to the consequence that those companies which take this step of proactively publishing the identification of their scrapers are more likely to be excluded by rightsholders from use of their data. On the contrary, those scrapers who are not known to the rightsholders are less likely to be covered in rightsholders reservations. This result however does not seem to be fair as it is disadvantageous for those companies who publish their User agent instructions and motivates the other not to voluntarily publish this information.

35 Potential solution to the above technical limitations could be either a completely new solution designed to implement TDM exception and express rightsholder's rules for use of content for AI training. For example, the European Commission recently announced its plan to conduct a feasibility study on the creation of a central registry where rights holders could opt out from TDM. The purpose of the study is to assess both the opportunity and feasibility of developing a work-based registry of content identifiers and associated metadata that would support – whether centrally or within a federated network– the effective expression of TDM opt-outs and facilitate their identification by AI developers. This could be a possible solution which might however require robust technical solution (which is to be explored by the aforementioned feasibility study).[105]

36 Alternatively, if Robots.txt is to be used, generic wording of User Agent line enabling to express reservation from TDM exception without applying differing rules for various scrapers could appear to be fair and workable solution. For example, as Open Future Policy Brief suggests, these could take the form of wildcard user-agent names such as **-genai, *-tdm, *-aiuser*[106] or the form of *MachineLearning* as suggested by Czech SPIR.[107] Alternatively - instead of such binary opt-out/non-opt-out approach allo - such unified vocabulary could introduce even more granular taxonomy of use cases for rightsholders to opt out from.[108] This solution could for example enable rightsholders to prohibit TDM for generative AI training but allow use for other forms of AI.[109] However, such solution could be even more complicated to unify which is the main issue in the existing technological landscape.

[101] Recommendation of Czech Association for Internet Development. Online vydavatelé se vymezují proti vytěžování dat umělou inteligencí. [online]. Available at https://www.spir.cz/online-vydavatele-se-vymezuji-proti-vytezovani-dat-umelou-inteligenci/ [Accessed on 31.12.2024].

[102] Media companies' websites are typically those publicly available websites who can be expected to publish copyright-protected content.

[103] AI Bot Blocking. OriginalityAI. [online]. Available at https://originality.ai/ai-bot-blocking [Accessed on 31.12.2024].

[104] Note: In the meantime, Anthropic AI published its recommendation on implementing „ClaudeBot" within Robots.txt.

[105] Study to assess the feasibility of a central registry of Text and Data Mining opt-out expressed by rightsholders, Accessible under File No. EC-CNECT/2025/OP/0002 in the EU Funding & Tenders Portal. Available online at: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/tender-details/8726813a-bd9b-4f58-8679-01c80f7a1abf-CN?isExactMatch=true&order=DESC&pageNumber=1&pageSize=50&sortBy=startDate [Accessed on 20.03.2025].

[106] Keller/Warso, 'Considerations for Opt-out Compliance Policies' (16 May 2024), Open Future Policy Brief #6 (2024), available at https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024].

[107] Recommendation of Czech Association for Internet Development. Online vydavatelé se vymezují proti vytěžování dat umělou inteligencí. [online]. Available at https://www.spir.cz/online-vydavatele-se-vymezuji-proti-vytezovani-dat-umelou-inteligenci/ [Accessed on 31.12.2024].

[108] Ibid.

[109] For example, C2PA approach distinguishes between data_mining, ai_training, ai_generative_training, and ai_inference. See standards introduced by the Coalition for Content Provenance and Authenticity (C2PA). Available at: https://c2pa.org/. [Accessed on 31.12.2024]. Other approaches (such as Spawning's products and the DeviantArt no-ai meta tag) are specifically targeted at (generative) AI training, while others (such as TDMRep) are explicitly aimed at the full spectrum of text and data mining. See Keller/Warso, 'Considerations for Opt-out Compliance Policies' (16 May 2024), Open Future Policy Brief #6 (2024), available at https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024].

37 Once the AI Act comes into effect, all providers placing general-purpose AI models on the market in the EU will be obliged to publish their policies on how they comply and identify with rightsholder's opt-out. It is not yet clear whether these policies will follow the same path taken by OpenAI, Common Crawl, Google and others and will contain instructions for User Agent line to prevent their scrapers accessing the respective content. Recently the AI Office published the first draft Code of Practice for general-purpose AI models for public consultation which however solely suggests that "*Signatories will only employ crawlers that read and follow instructions expressed in accordance with the Robot Exclusion Protocol (robots.txt)*". The Code of Practice undergoes multiple rounds of consultations and is expected to be finalized before May 2025.

V. Burden of Proof & Logging Evidence of Valid Opt-Out

38 The reservation from the TDM exception should in the author's view be effective after being placed at the respective website. Prior to that moment the TDM exception applies without such condition that rightsholders expressly reserved their rights. Although as for example Peter Mézei aptly points out "*the directive neither prompts nor excludes that such reservations should be carried out ex ante (preceding the mining) or ex post (following the mining).*" while noting that "*TDM might happen quicker than an ex-ante reservation could have been expressed. Consequently, ex post reservations shall not be automatically excluded from the scope of Art. 4(3).*".[110] On the contrary, for example Czech Explanatory Memorandum explicitly highlights that reservation applies solely for future use and cannot apply retrospectively.[111] Such conclusion may follow also from past tense forms used in some member state laws implementing the CDSM Directive - for example in the German[112], Czech[113], Austrian[114] implementation. In addition, requiring the developer to do so does not seem to be proportionate in case the developer has lawfully relied on an exception from copyright protection allowing to *retain reproductions for as long as is necessary for the purposes of text and data mining.*[115] *Ex-ante* reservations also correspond to technological reality as once an AI model is trained, the copyright protected content can hardly be retrospectively removed from the original training data. As Open Future Policy Brief notes, for each version of AI model, there could be some sort of *opt-out cut-off date, after which new opt-outs will no longer affect the model's training* whereas such cut-off date could be transparently communicated once AI model is released.[116]

39 However, the existence of a reservation as of

[110] Mezei, Péter, A saviour or a dead end? Reservation of rights in the age of generative AI (January 15, 2024). European Intellectual Property Review, 2024, 46(7), p. 461-469. Available at SSRN: https://ssrn.com/abstract=4695119 or http://dx.doi.org/10.2139/ssrn.469511. [Accessed on 31.12.2024]. Page 8.

[111] Explanatory memorandum (Důvodová zpráva) of the Czech Government to the Act. No. 429/2022 Coll. (amending the Czech Copyright Act implementing the CDSM Directive). Section § 39c.

[112] § 44b (3) of the German Urheberrechtsgesetz stating that *„(3) Nutzungen nach Absatz 2 Satz 1 sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat. Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in maschinenlesbarer Form erfolgt.*" or as translated to English: *„(3) Uses pursuant to paragraph 2 sentence 1 shall only be permitted if the rightholder has not reserved the right of use. A reservation of use in the case of works accessible online shall only be effective if it is made in machine-readable form.*"

[113] § 39 c (2) of the Czech Copyright Act stating that "*(2) Ustanovení odstavce 1 se nepoužije pro rozmnoženiny díla, jehož autor si užití podle odstavce 1 výslovně vyhradil vhodným způsobem; v případě díla zpřístupněného podle § 18 odst. 2 strojově čitelnými prostředky.*" or as translated to English: „*2) The provision of paragraph 1 does not apply to reproductions of the work, the author of which has expressly reserved the use according to paragraph 1 in an appropriate manner; in the case of a work made available in accordance with § 18 paragraph 2 by machine-readable means*"

[114] § 42 h of the Austrian Urheberrechtsgesetz stating that "*(6)Jedermann darf für den eigenen Gebrauch ein Werk vervielfältigen, um damit Texte und Daten in digitaler Form automatisiert auszuwerten und Informationen unter anderem über Muster, Trends und Korrelationen zu gewinnen, wenn er zu dem Werk rechtmäßig Zugang hat. Dies gilt jedoch nicht, wenn die Vervielfältigung ausdrücklich verboten und dieses Verbot in angemessener Weise durch einen Nutzungsvorbehalt, und zwar etwa bei über das Internet öffentlich zugänglich gemachten Werken mit maschinenlesbaren Mitteln, kenntlich gemacht wird. Eine Vervielfältigung nach diesem Absatz darf aufbewahrt werden, solange dies für die Zwecke der Datenauswertung und Informationsgewinnung notwendig ist.*" or as translated to English "*(6) Anyone may reproduce a work for their own use in order to automatically evaluate texts and data in digital form and to obtain information on patterns, trends and correlations, among other things, if they have lawful access to the work. However, this shall not apply if reproduction is expressly prohibited and this prohibition is appropriately indicated by a reservation of use, for example in the case of works made publicly accessible via the Internet by machine-readable means. Reproduction in accordance with this paragraph may be retained as long as this is necessary for the purposes of data analysis and information retrieval.*"

[115] Art. 4 (2) CDSM Directive.

[116] Keller/Warso, 'Considerations for Opt-out Compliance Policies' (16 May 2024), Open Future Policy Brief #6 (2024), available at https://openfuture.eu/wp-content/uploads/2024/05/240516considerations_of_opt-out_compliance_policies.pdf [Accessed on 31.12.2024].

certain moment in time may be practically difficult to prove in potential dispute without for example time-stamped evidence proving the existence of reservation from the TDM exception as of certain specific moment in time. In the event of a potential dispute, rightsholders as potential plaintiffs might be claiming copyright infringement whereas scrapers as potential defendants might be claiming that TDM exception applies. Therefore, the rightsholders will likely bear the burden of proof that copyright infringement occurred whereas scrapers will likely bear the burden of proof of lawful use of content and thus proving that TDM exception applies. German explanatory memorandum suggests that the burden of proof for the absence of a reservation shall be born the user who is relying on such exception.[117] Löbling, Handschigl, Hofmann and Schwedhelm are of the view that "*TDM user bears the onus of proof, mandated by the phrasing of paragraph 3 ("are permitted only if they have not been reserved*"), although acknowledge that "*copyright holder is accountable for properly expressing their opt-out decision*".[118] This question will remain to be addressed by civil procedural rules which differ in EU member states.

D. Remarks on the Interplay between TDM Exceptions under Art. 3 and 4 CDSM Directive Considering Practicality of Gen AI Development

40 Datasets are not always created by the same legal entities which are developing artificial intelligence. On the contrary, datasets are often populated by various third parties or non-profit organisations and only subsequently cleansed, adjusted and used by AI companies to train Gen AI.[119] This follows for example from limited publicly available information suggesting that some large language models might have been trained on datasets such as Common Crawl, LAION, BookCorpus, Wikipedia, WebText.[120] Each of these examples implements different purpose and *modus operandi* – for example Common Crawl is a non-profit organisation publishing a dataset consisting of raw web page data, metadata extracts, and text extracts collected from publicly available websites since 2008[121], LAION on the other hand provides publicly and free of charge a dataset for image-text pairs consisting of hyperlinks to images or image files publicly accessible on the Internet as well as other information related to the respective images, including an image description.[122] However, while these repositories – due to their non-profit nature - might themselves rely on TDM Exception for research purposes – those AI companies using their data might not. Although the Common Crawl Foundation proclaims to comply with Robots.txt and *no follow policies* of the scraped websites (for these purposes the Common Crawl Foundation even issued its own Robots.txt guidance recommending implementing "*CCBot*" to the user-agent line[123]), at the same time Common Crawl's publicly available Terms of use explicitly limit Common Crawl's liability for third party IP infringements and explicitly state that Crawled Content may be subject to separate terms of use or terms of service from the owners of such Crawled Content.[124] These aspects add additional layer of complexity in potential disputes over lawfulness of text and data mining. For example, in *Robert Kneschke v. LAION* the court tackled solely the use of protected content by LAION (as the defendant) but subsequent use of LAION datasets by AI developers was not part of the case.[125]

41 Both TDM exceptions are by virtue of the definition of text and data mining limited to actions aiming to *generate information.* Such requirement is stemming from the legal definition of text and data mining as a legal term defined in the CDSM Directive as "*any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations*".[126] As indicated in the preamble of the

[117] Explanatory memorandum (Gesetzesbegründung) of the German Government (Bundesregierung) to its legislative proposal implementing the CDSM Directive: Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesbegründung: Besonderer Teil. No. 19/27426. Page 88. Available at https://dip.bundestag.de/vorgang/.../273942 [Accessed on 31.12.2024].

[118] Löbling, L., Handschigl, Ch. Hofman, K., Schwedhelm, J. Navigating the Legal Landscape: Technical Implementation of Copyright Reservations for Text and Data Mining in the Era of AI Language Models. 14 (2023) JIPITEC 499 para 12.

[119] Generally, preparation of dataset for AI training involves very thorough process involving data cleansing, de-duplication and other measures aiming to enhance dataset quality.

[120] Brown, T.T., et al., Language Models are Few-Shot Learners. Available at: https://arxiv.org/pdf/2005.14165 [Accessed on 31.12.2024]. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I., Language Models are Unsupervised Multitask Learners, OpenAI, 2019.

[121] Common Crawl, *Common Crawl Overview*, available at: https://commoncrawl.org/overview [Accessed on 31.12.2024].

[122] LG Hamburg, Urteil vom 27. September 2024 – 310 O 227/23 (Robert Kneschke v. LAION).

[123] Common Crawl, CCBot. [online]. Available at: https://commoncrawl.org/ccbot [Accessed on 31.12.2024].

[124] Terms of Use of Common Crawl, available online at: https://commoncrawl.org/terms-of-use [Accessed on 31.12.2024].

[125] Although partially mentioned in the *obiter dictum.*

[126] Article 2 (2) CDSM Directive.

CDSM Directive, the text and data mining exceptions aim to encourage innovation in both the public and private sectors as legislators acknowledge its benefits in enabling the processing of large amounts of information with a view to "*gaining new knowledge and discovering new trends*".[127] Narrowing the definition of text and data mining solely to the purpose of *generating information* reflects the overarching goal of the CDSM Directive. For example, as noted in the German explanatory memorandum, the purpose of the text and data mining covered by the exception does not cover actions aimed at collecting and storing content to create parallel digital archives.[128] German court in *Robert Kneschke v. LAION* offered interesting perspective and interpreted the requirement of generating new information very broadly. The court applied TDM Exception with an explanation that the defendant undertook the reproduction action for the purpose of extracting information about "*correlations*" to compare the image content with the image description already stored in the text using an available software application. The court noted that although the creation of the dataset itself may not yet be associated with a knowledge gain, it is a fundamental step aimed at using the dataset for the purpose of later knowledge acquisition. The court held as sufficient that the dataset was undisputedly published for free and thus made available, particularly to researchers working in the field of artificial neural networks. However, the court considered as irrelevant whether such other researchers are commercial enterprises or non-profit undertakings.

42 However, although such interpretation has positive impact on innovation allowing such organisations to create and publish open-source datasets, such interpretation might not hold up. As explained above, some organisations might be merely populating publicly available data and publishing the respective datasets for non-profit research purposes, however, not train AI or generate new information themselves. On the contrary, such dataset created for non-profit purposes may be subsequently used by companies developing Gen AI on a for-profit basis. That could however mean that strictly speaking, companies creating datasets are *stricto sensu* not generating new information and on the contrary, reproductions made by companies developing Gen AI on a for-profit basis cannot be covered by the research exception. In addition, such argumentation had justification with respect to LAION as it does not publish the original works but solely hyperlinks and concurrently indeed provides analysis of the correlations. The same modus operandi however might not apply to other dataset publishers.

43 In instances where an AI model is initially developed under a non-profit framework, adheres to removing original datasets post-training, and later transitions into commercial use, the initial reproduction or extraction activities could technically still fall within the TDM exception under the CDSM Directive for non-commercial research purposes. However, this exception would strictly apply only to those preliminary reproduction and extraction actions within the non-profit stage. Any subsequent activities, including storage of original raw data or dissemination of copyrighted material within AI outputs that might arise due to data memorization, fall outside this exception as further described below. Lastly, if companies that create datasets are found to infringe on copyright, such infringement could potentially compromise the legality of AI companies' subsequent use of the datasets. Even if these AI companies duly rely on the TDM exception under Art. 4 of the CDSM Directive, initial copyright infringement might lead to unlawful access, conflicting with the lawful access requirement outlined in Articles 3 and 4 of the Directive.

E. Does the TDM Exception Really Provide an Answer? Is it Technically Possible to Train Gen AI but Prevent Verbatim Extracts of Training Data in Gen AI Outputs?

44 Due to the limited scope of 3 and 4 CDSM Directive, both TDM Exceptions cover solely the acts of reproduction but not subsequent modifications or communication to the public / reutilization of the original data.[129] Specifically, TDM exception covers

The following key elements of text and data mining can be derived from this legal definition: (i) automated analytical techniques; (ii) analysis of text and data in digital form; (iii) aim intended to generate information (including patterns, trends and correlations).

[127] Recital 8 and 18 CDSM Directive.

[128] Explanatory memorandum (Gesetzesbegründung) of the German Government (Bundesregierung) for a legislative proposal implementing the CDSM Directive: Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesbegründung: Besonderer Teil. Page 88. Available at https://dip.bundestag.de/vorgang/.../273942 [Accessed on 31.12.2024]. Page 88.

[129] E Rosati, The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects, available at http://www.europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf. [Accessed on 31.12.2024]. Novelli, Claudio and Casolari, Federico and Hacker, Philipp and Spedicato, Giorgio and Floridi, Luciano, Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity (January 14, 2024). Available at SSRN: https://ssrn.com/abstract=4694565 or http://dx.doi.org/10.2139/ssrn.4694565 [Accessed on 31.12.2024].

(i) the *right of reproduction* of copyrighted works[130], databases[131], and on-demand press publications[132]; (ii) *the right of extraction* of a whole or a substantial part of databases covered by the sui generis database rights[133]; and (iii) the right to reproduction and the right to adaptation of computer programs[134].[135]

45 As follows from claims filed in the US and UK[136], plaintiffs often claim not only use of their works in connection with AI training but also further dissemination of their works within AI outputs which in terms of EU law would exceed the scope of right of reproduction and may constitute a communication to the public (as for example follows from the complaint filed by The New York Times Company against Microsoft and OpenAI or class action complaint filed by the US Authors Guild against Microsoft and OpenAI).

46 The act of text and data mining occurs at the early stage of model development. During this phase, the model is trained on such datasets. Although large language models might not be technically storing the original datasets and raw data used for training; such models may sometimes retain and produce verbatim snippets or other identifiable data elements due to a phenomenon known as data memorization. Data memorization occurs for example when specific data points, such as text or images, are repeatedly encountered during training, leading the model to "*memorize*" these elements, sometimes resulting in output that closely resembles or directly mirrors segments of the original data.[137] As Carlini concluded "*Memorization significantly grows as we increase (1) the capacity of a model, (2) the number of times an example has been duplicated, and (3) the number of tokens of context used to prompt the model*".[138]

47 Although TDM exceptions may serve as a legal basis authorizing use of protected content for purposes of AI training, they might not justify subsequent reuse the respective content in case generative AI models produce verbatim snippets of original works.[139] Practical solution may be implementation of additional measures. For example, deduplication[140] of training data which is considered to be one of available countermeasures against data memorization[141] whereas "*the core idea is to remove any duplicated content—e.g., repeated documents—because duplicated content is much more likely to be memorized. However, deduplication does not guarantee that a model will not still memorize individual (deduplicated) examples*. In addition, applying various types of output filters may prevent further dissemination of the protected content within AI outputs such as retroactive censoring or memfree decoding which explicitly "*prohibit the model from emitting a sequence if it is contained (entirely or partially) in the training dataset*".[142] For example, GitHub's Copilot, a language

[130] Article 2 InfoSoc Directive.

[131] Article 5(a) Database Directive.

[132] Article 15(1) CDSM Directive.

[133] Article 7(1) Database Directive.

[134] Articles 4(1)(a) and (b) InfoSoc Directive.

[135] The scope of exception under Article 4, CDSM Directive is broader than the exception under Article 3 of the CDSM Directive (i.e. TDM for scientific purposes), which unlike Article 4 of the CDSM Directive does not cover the right to reproduction and the right to adaptation of computer programs.

[136] See footnote 5.

[137] Biderman, S., Prashanth, U. S. S., Sutawika, L., Schoelkopf, H., Anthony, Q., Purohit, S., & Raff, E., 2023. *Emergent and Predictable Memorization in Large Language Models.* arXiv preprint arXiv:2304.11158v2 [cs.CL]. Available at: https://doi.org/10.48550/arXiv.2304.11158 [Accessed on 31.12.2024]. Huang, J., Yang, D., & Potts, C., 2023. Demystifying Verbatim Memorization in Large Language Models. Stanford University. Available at: https://arxiv.org/ [Accessed on 31.12.2024]. Carlini, N., Ippolito, D., Jagielski, M., Lee, K., Tramèr, F., & Zhang, C., 2023. Quantifying Memorization Across Neural Language Models. Google Research and Cornell University. Available at: https://arxiv.org/ [Accessed on 31.12.2024]. Albert Ziegler, "GitHub Copilot research recitation" Github blog, 30 June 2021; Available at: https://github.blog/2021-06-30-github-copilot-research-recitation [Accessed on 31.12.2024]. Daphne Ippolito, Florian Tramèr, Milad Nasr, Chiyuan Zhang, Matthew Jagielski, Katherine Lee, Christopher A. Choquette-Choo, and Nicholas Carlini, 'Preventing Verbatim Memorization in Language Models Gives a False Sense of Privacy', arXiv (2023), arXiv:2210.17546v3 [cs.LG], pp. 1–26. Gowthami Somepalli, Vasu Singla, Micah Goldblum, Joans Geiping & Tom Goldstein, "Diffusion Art or Digital Forgery? Investigating Data Replication in Diffusion Models" (2023) https://arxiv.org/abs/2212.03860 [Accessed on 31.12.2024]. Nicholas Carlini. Jamie Hayes, Milad Nasr, Matthew Jagielski, Vikash Sehwag, Florian Tramèr, Borja Balle, Daphne Ippolito & Eric Wallace, "Extracting Training Data from Diffusion Models" (2023).

[138] Carlini, N., Ippolito, D., Jagielski, M., Lee, K., Tramèr, F., & Zhang, C., 2023. Quantifying Memorization Across Neural Language Models. Google Research and Cornell University. Available at: https://arxiv.org/ [Accessed on 31.12.2024].

[139] Rosati, Eleonora, Infringing AI: Liability for AI-generated outputs under international, EU, and UK copyright law (August 31, 2024). European Journal of Risk Regulation, Available at SSRN: https://ssrn.com/abstract=4946312 or http://dx.doi.org/10.2139/ssrn.4946312 [Accessed on 31.12.2024].

[140] Data deduplication has arisen as a pragmatic countermeasure against data memorization (Lee et al., 2021; Kandpal et al., 2022; Carlini et al., 2022). The core idea is to remove any duplicated content—e.g., repeated documents—because duplicated content is much more likely to be memorized.

[141] Lee et al., 2021; Kandpal et al., 2022; Carlini et al., 2022.

[142] Daphne Ippolito, Florian Tramèr, Milad Nasr, Chiyuan Zhang, Matthew Jagielski, Katherine Lee, Christopher A. Choquette-Choo, and Nicholas Carlini, 'Preventing

model-based code assistant, adopts similar measures and offers users to "*block suggestions matching public code*".[143] However, previous research indicates that even when a model is restricted from emitting any output with snippets of verbatim memorization, the model might still leak some parts of training data.[144] For example, research testing GitHub Copilot which implemented retroactive censoring shows that "*Copilot's filter can easily be bypassed by prompts that apply various forms of "style-transfer" to model outputs, thereby causing the model to produce memorized (but not verbatim) outputs*".[145] On the other hand, such "style-transfer" outputs may significantly less likely constitute copyright infringement than verbatim snippets depending on the level of autonomy of the creation and dependency on the pre-existing content.[146] Such assessment however depends on case-by-case basis taking into account also involvement of the user prompting the LLM.[147] in such case the burden of proof of the respective copyright infringement lies with the rightsholders potentially claiming such infringement.

48 As a result, even when duly and lawfully applying TDM exception for purposes of text and data mining to facilitate generative AI training, AI models may still face significant challenges and difficulties to rely on text and data mining within the legal borderlines of copyright laws.

F. Concluding Remarks

49 This paper highlighted practical challenges tied to TDM exceptions, which may inevitably come up in disputes over AI-related copyright infringements. For example:

- Machine-readable reservation allowing rightsholders to opt-out from for-profit TDM exception may hit the barrier of lacking standardisation.

- The CDSM Directive does not define the required level of "*machine-readability*" for rightsholders' reservations. German court noted that "*machine-understandability*" may be sufficient depending on technical developments at the relevant time of use. With such justification the court considered even terms and conditions in human language as machine-readable since such terms may be decoded by generative AI. German court in Robert Kneschke v. LAION noted that "these "state-of-the-art technologies" undoubtedly include, in particular, AI applications capable of comprehending text written in natural language" which might however not achieve sufficient level of reliability and thus applying these conclusions would pose significant risks for AI companies relying on such technologies to decode rightsholders' opt out.

- However, in order to strike a balance between the interests of users of text and data mining (to be able to conduct automated analysis of data) and the interests of rights holders (to protect their rights), this rightsholders' "*express*" reservations in "*machine-readable*" formats should, in the author's view, achieve *sufficiently reliable level* of machine interpretability which might not be achieved when relying on Gen AI decoding terms and conditions written in natural language. This might require the reservation to be presented in a sufficiently *standardized form* that enables such advanced technologies to *reliably* decode its content leaving no room for doubt. This may be reflected for example by standardized formulas (despite being written in a natural human language) – for example similarly as open-source licensing terms.

- Robots.txt is a key tool for expressing reservations but its simplicity can lead to technical limitations and unintended side effects. Prohibiting all bot access via Robots.txt affects website indexing by search engines, making it largely impractical.

- Currently, Robots.txt cannot block specific uses like text and data mining; it only allows naming specific scrapers in the user-agent line. Some AI market players set the trend of publishing instructions for the user-agent line to block their scrapers and opt-out from their AI training. However, this requires rightsholders to monitor all viable scrapers and disadvantages those AI companies who publish these instructions (since websites typically follow these recommendations if published and restrict data use for specified user agents) while practically favouriting those who do not (as the rightsholders do not know how to identify them in the User agent line).

Verbatim Memorization in Language Models Gives a False Sense of Privacy', arXiv (2023), arXiv:2210.17546v3, pp. 1–26. [Accessed on 31.12.2024].

[143] Ibid.

[144] Ibid.

[145] Ibid.

[146] Novelli, Claudio and Casolari, Federico and Hacker, Philipp and Spedicato, Giorgio and Floridi, Luciano, Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity (January 14, 2024). Available at SSRN: https://ssrn.com/abstract=4694565 or http://dx.doi.org/10.2139/ssrn.4694565 [Accessed on 31.12.2024].

[147] Ibid.

- AI Act implementations might bring clarity by once providers of general-purpose AI models publish TDM compliance policies following state-of-the-art technologies, though this applies only to companies marketing such models in the EU.

- Potential disputes will also inevitably involve practical and procedural challenges, such as determining the extent of each party's burden of proof and how to demonstrate that a reservation was made at a specific point in time.

50 Consequently, given the practical and technical limitations discussed in this paper, developing a clear market standard solution that both AI developers and rightsholders can adhere to would be highly beneficial. Standardized TDM identifiers will enable to streamline opt-out processes and will reduce costs and increase legal certainty for both rightsholders and AI companies.

51 Nevertheless, since TDM exceptions allow solely the acts of reproduction / extraction but not subsequent modification and use – even if TDM part of AI training is resolved, AI companies will still have to carefully tackle the risks of any data memorization which may lead to producing verbatim snippets of training data which would not be legitimized by the TDM exceptions. As a result, even when duly and lawfully applying TDM exceptions to legitimize use of data for generative AI training, AI models may still face significant challenges and difficulties when scraping copyright protected content without a license from the rightsholders.

52 To the very end, machine-readable reservations allowing rightsholders to opt out of for-profit TDM exceptions could grant the rightsholders significant power, potentially leading to widespread withdrawal from AI training. This might deprive the EU public of future AI innovations using high quality datasets while at the same time not enabling the authors from benefitting therefrom (for example by offering their content in exchange for remuneration). Solutions such as machine-readable licensing models or collective management, could offer a balanced compromise between protecting rightsholders' rights and fostering AI development. Such solutions would however either require significant legislative changes or robust licensing frameworks and data spaces enabling to acquire license via automated means.

Push Notifications under E-Privacy Law: A Review and Outlook on the Interplay between Data Protection Law, E-Privacy Law and other Legal Acts

by **Tristan Radtke ***

Abstract: Push notifications are widely used to inform users directly about messages, news and offers. Although the opt-in mechanisms implemented by all providers of push notifications might suggest straightforward compliance with e-privacy law, this popular phenomenon is a good example to discuss the current and future challenges under European e-privacy and data protection law. The use of push notifications raises intriguing legal questions under the e-privacy directive, the General Data Protection Regulation (EU) 2016/679 (GDPR) and the law of unfair commercial practices. The focus here is on questions related to the interaction of these different legal acts, the requirements for legal bases as well as the relationship between a consent requirement and the push notification permissions granted through system permission prompts on the devices. A closer look is necessary for the requirements of the e-privacy Directive with regard to the storage of information on the device, unsolicited communication and the question of whether push notifications constitute electronic mail or other forms of communication. Against this background, this article explores the complex legal landscape surrounding push notifications, addresses these legal challenges, and provides standards for push notifications using different scenarios. Finally, the article concludes with a discussion on how the current legal framework handles such an important phenomenon and considers what to expect from a potential e-privacy Regulation in this regard.

Keywords: Push Notifications; E-Privacy Directive; Consent Requirement, Data Protection Law

Recommended citation: Tristan Radtke, Push Notifications under E-Privacy Law: A Review and Outlook on the Interplay between Data Protection Law, E-Privacy Law and other Legal Acts, 16 (2025) JIPITEC 107 para 1

A. Introduction

1 Push notifications are brief, alert-style messages sent by app providers, including websites, to user devices such as smartphones or personal computers. These notifications are designed to inform users about updates and reminders, provide promotional content, or prompt users to an action, even when the app is not actively in use. Push notifications address individual users directly by delivering content to their devices. The combination of these phenomena raises challenges under data protection law and the law of unfair commercial practices.

I. Data flows

2 The specifics of the data flow vary according to the device and the operating system in use. However, the data flow involved in the delivery of a push notification can be summarized as follows:[1] The app of

* Dr. Tristan Radtke, LL.M. (NYU) is a research assistant (Akademischer Rat a.Z.) at the Chair for Law and Regulation of the Digital Transformation (Prof. Dr. Boris P. Paal, M.Jur. (Oxford)), TU Munich – School of Social Sciences and Technology, Department of Governance. The author would like to thank Prof. Dr. Boris P. Paal, M.Jur. (Oxford) for his valuable comments on a previous version of this article and fruitful discussions on this topic.

[1] See 'Setting up a remote notification server' (*Apple*

the app provider is installed and launched on the device by the users. Once the app is launched, the app provider can ask users to allow push notifications through an operating system permission prompt. Users can change the format of the push notification in the operating system settings, but not the content or frequency of the notifications.

3 The launch of the application initiates the registration process with a push notification provider, which generates a unique token for the specific application on the particular device. The push notification provider depends on the device and its operating system. For iOS and other Apple devices, it is the Apple Push Notification (APN)[2] service, for Android devices, it is often Firebase Cloud Messaging,[3] for web push notifications in Firefox, it is the Mozilla Web Push[4] service. However, there might be additional service providers in the middle between the app provider and the push notification provider in order to facilitate the process and provide a framework for sending push notifications on different platforms.

4 Once the device token has been generated, it is the responsibility of the app provider to transmit this token to its own servers and link it with other identifiers (e.g., with the user's account information).

5 When the app provider wishes to issue a push notification via their servers, the server of the app provider requests the push notification provider to initiate the process by submitting the message, the modalities and the app's device token. Such modalities may include information about the expiration of the notification after a certain period of time during which the device was offline and the notification could not be delivered (e.g., 30 days).[5] Furthermore, the app provider could specify the priority of a notification.[6]

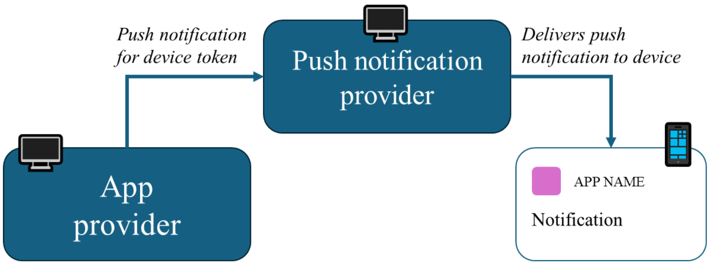

Push notification for device token
Push notification provider
Delivers push notification to device
App provider
APP NAME
Notification

Figure 1: Process of initiating a push notification from the server of the app provider to the user's device (simplified)

6 This analysis focuses on the initiation of the push notification process from the perspective of the app provider. As the push notification provider is only responsible for delivering the content *provided by the app provider*, the role and obligations of the latter under applicable law will not be considered in this article.

II. Analysis based on three Scenarios

7 To illustrate the lawfulness of push notifications, this article focuses on three specific scenarios in the context of connected cars:

(1) An app is connected to a vehicle and can be used to view certain metrics about the vehicle's condition. The push notification informs the user that a new software update for the software of the vehicle is available.

(2) As in Scenario 1, but the push notification informs the user of an available update for the mobile app.

(3) As in Scenario 1, but the push notification contains promotional advertising on a discount available for a vehicle software upgrade.

8 These Scenarios will be used to analyze the legal challenges posed by device access under the e-privacy Directive (see below B. I.), communication with the user under the e-privacy Directive (see below B. II.) and the processing of personal data in general under the GDPR (C.). Further requirements arising from the UCP and ecommerce Directive will then be addressed (see below D.).

Developer) <https://developer.apple.com/documentation/usernotifications/setting-up-a-remote-notification-server> accessed 15 November 2024; 'Registering your app with APNs' (*Apple Developer*) <https://developer.apple.com/documentation/usernotifications/registering-your-app-with-apns> accessed 15 November 2024.

[2] 'Registering your app with APNs' (*Apple Developer*) <https://developer.apple.com/documentation/usernotifications/registering-your-app-with-apns> accessed 15 November 2024.

[3] 'Firebase Cloud Messaging' (*Google Firebase*), <https://firebase.google.com/docs/cloud-messaging> accessed 15 November 2024.

[4] 'Web push notifications in Firefox' (*Firefox Support*, 9 February 2023) <https://support.mozilla.org/en-US/kb/push-notifications-firefox> accessed 15 November 2024.

[5] 'Sending notification requests to APNs' (*Apple Developer*) <https://developer.apple.com/documentation/usernotifications/sending-notification-requests-to-apns> accessed 15 November 2024.

[6] *Apple Developer* (n 5).

B. E-Privacy Directive

9 The e-privacy Directive 2002/58/EC, as amended by Directive 2009/136/EC, addresses primarily privacy concerns with respect to electronic communication services and "particularise[s] and complement[s]" the GDPR insofar (cf. art. 1(2) e-privacy Directive, art. 94(2) GDPR). In light of the stipulations set forth in art. 95 of the GDPR, which establishes that the provisions of the e-privacy Directive prevail over the general GDPR,[7] this analysis will initially focus on the e-privacy Directive and subsequently address the GDPR requirements.

10 However, despite the e-privacy Directive being *lex specialis* to the GDPR, the e-privacy Directive takes a slightly different approach. As the name of the Directive suggests, the e-privacy Directive is primarily concerned with the protection of privacy with regard to devices and the confidentiality of communications (arts. 7, 8 EU Charter of Fundamental Rights, hereinafter Charter), rather than merely data protection (art. 8 Charter).[8] The here relevant arts. 5(3) and 13 e-privacy Directive are primarily concerned with the protection of the private sphere, including users' devices in the context of electronic communication.[9] The national provisions implementing the e-privacy Directive have to be interpreted in accordance with the Directive.

11 In the near future, the e-privacy Regulation, which has not yet to be agreed upon,[10] could replace the e-privacy Directive. Although the precise details of the successor provision to art. 5(3) remain uncertain, there are indications that the e-privacy Regulation will adopt a provision similar to art. 13, potentially with only a few modifications.[11]

I. Access to the User's Device under Art. 5(3) E-Privacy Directive

1. Scope of Art. 5(3) E-Privacy Directive

12 According to art. 5(3) e-privacy Directive, the storage of information and the access to information on the terminal equipment of the user (e.g., a smartphone) is subject to limited specific legal bases: (1) the consent of the user, (2) the necessity for transmissions or (3) the necessity for the provision of a requested service. This applies to information on any type of device medium including the RAM for temporary storage.[12] The ECJ places emphasis on the language "information" and interprets art. 5(3) e-privacy Directive broadly to cover both personal and non-personal data.[13]

13 Delivering a push notification involves temporarily storing its content on the device, which constitutes storing information on the user's terminal equipment under art. 5(3) of the e-privacy Directive. As with

[7] Recital 173 GDPR; Christoph Werkmeister in Frenz Jürgen Säcker and Torsten Körber (eds), *TK – TTDSG* (4th edn, dfv 2023), s 25 TTDSG para 40; Daniel A Pauly in Boris P Paal and Daniel A Pauly (eds), *DS-GVO BDSG* (3rd edn, CH Beck 2021), art. 95 DS-GVO para 2.

[8] cf. art. 1(1), recital 12 e-privacy Directive; Achim Klabunde and Martin Selmayr in Eugen Ehmann and Martin Selmayr, *DS-GVO* (3rd edn, CH Beck 2024), art. 95 DS-GVO para 10; Alexander Golland in Jürgen Taeger and Detlev Gabel (eds), *DSGVO – BDSG – TTDSG* (4th edn, dfv 2022), art. 95 DSGVO para 9; Vagelis Papakonstantinou and Paul De Hert in Indra Spiecker gen. Döhmann and others (eds), *General Data Protection Regulation* (CH Beck and Nomos 2023), art. 95 para 2. On the terms privacy and data protection Lee A Bygrave, 'Privacy and Data Protection in an International Perspective' (2010) 56 Scandinavian Stud L 165.

[9] cf. Case C-673/17 *Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband e.V. v Planet49 GmbH*, Opinion of AG Szpunar, ECLI:EU:C:2019:246, para 107; recital 24 e-privacy Directive; EDPB, 'Opinion 5/2019 on the interplay between the e-privacy Directive and the GDPR, in particular regarding the competence, tasks and powers of data protection authorities' (12 March 2019), paras 25-28; Werkmeister (n 7) 1; Carlo Piltz, 'Das neue TTDSG aus Sicht der Telemedien' [2021] CR 555, 560; Hanloser, 'Telekommunikation-Telemedien-Datenschutz-Gesetz' [2021] ZD 121, 121.

[10] The e-privacy Regulation was originally intended to come into force at the same time as the GDPR in 2018. Due to different views on the proposal within the EU institutions, probably also and especially with regard to tracking, no agreement has been reached to date and the proposal has just been withdrawn. For an overview, see e.g. Martin Selmayr and Eugen Ehmann in Ehmann and Selmayr (n 8) Introduction 130.

[11] See the draft of the e-privacy Regulation, Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)' 6087/21 (10 February 2021); providing an overview Christina Etteldorf, 'A New Wind in the Sails of the EU e-privacy-Regulation or Hot Air Only? On an Updated Input from the Council of the EU under German Presidency' (2020) 6 Eur Data Prot L Rev 567; Louisa Specht in Louisa Specht and Reto Mantz (eds), *Handbuch Europäisches und deutsches Datenschutzrecht* (CH Beck 2019), s 9 para 13.

[12] EDPB, 'Guidelines 2/2023 on Technical Scope of Art. 5(3) of e-privacy Directive' (14 November 2023), para 37.

[13] Case C-673/17 *Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband e.V. v Planet49 GmbH* (ECJ, 1 October 2019), ECLI:EU:C:2019:801, paras 69-70.

cookies, the fact that they are automatically deleted after a certain period of time does not preclude the presumption that information is being "stored".[14] It may be argued against an interpretation of art. 5(3) e-privacy Directive, which covers temporary storage of a displayed information, that such a broad interpretation would cover any website that is stored on a user's device in order to display its content. This was probably not the intention of the legislator. However, storing such information initiated directly by the user does not constitute provider-initiated storage as required under art. 5(3) of the e-privacy Directive.

14 Furthermore, information originating from the device such as the device token or the version of the installed app, could be accessed in order to deliver a push notification, which would also be considered access to information on the device.[15]

2. Exceptions from Consent Requirement

15 Art. 5(3) e-privacy Directive provides two exceptions to the principle that access to or the storage of information on the user's terminal equipment is prohibited. Where these exceptions apply, providers are not required to obtain the user's consent.

a.) Necessity for Transmission of a Communication

16 The first exception permits storage or access if it is necessary "for the sole purpose of carrying out the transmission of a communication over an electronic communications network" (art. 5(3)(2)(alt. 1) e-privacy Directive).

17 This is the case for device identifiers, without which communication could not be delivered.[16] If the service provider were to access the device token on the device before delivering a particular push notification in order to enable the delivery of the push notification at hand, that access would be covered by the exception.

18 However, it is not considered necessary to access the device to *store* the content of the push notification on the device in order to carry out the *transmission*. Art. 5(3)(2)(alt. 1) e-privacy Directive must be read restrictively in order to leave some room for the exception of the service explicitly requested by the user (alt. 2, as discussed under b.).[17] If the storage of any information were covered by the exception for transmission, there would be no need for the exception for the service requested by the user or for a consent requirement.

19 In all three Scenarios, the access to the identifiers stored on the device is covered by the exception laid down in art. 5(3)(2)(alt. 1) e-privacy Directive. For the storage of the content of the push notification, the provider has to rely on another exception.

b.) Information Society Service Explicitly Requested by the User

20 Second, any storage or access "strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service" is permitted (art. 5(3)(2)(alt. 2) e-privacy Directive). Information society services are defined in art. 1(1)(b) Directive (EU) 2015/1535 as "any service normally provided for remuneration, at a distance, by electronic means and at the individual request of a recipient of services". This covers any app-related services and push notifications, regardless of whether they are actually provided for remuneration.[18] Since an *explicit* request for each communication is not required for a service to be considered an information society service, the installation of an app may be generally sufficient for the purposes of the definition with respect to future push notifications.[19]

21 Any interpretation of art. 5(3)(2)(alt. 2) e-privacy Directive has to give sufficient consideration to the elements "strictly necessary", "explicitly requested" and the determination of the respect service and its scope.[20] The test for the "strictly necessary" prong is whether the specific service could not be provided at all without the storage of or access to the information.[21] The element of an explicit request of the service is met if the user has the reasonable expectation that information will be stored or accessed on his device, if this part of the service is used and thus "requested".[22] In order not to undermine the general consent requirement under art. 5(3)(1) e-privacy Directive, the part of the

[14] cf. *Planet49 GmbH* (n 13) 75.

[15] cf. EDPB (n 12) 55.

[16] WP29, 'Opinion 04/2012 on Cookie Consent Exemption' (7 June 2012), 3.

[17] cf. WP29 (n 16) 2-3.

[18] Case C-484/14 *Tobias Mc Fadden v Sony Music Entertainment Germany GmbH* (ECJ, 15 September 2016), ECLI:EU:C:2016:689, paras 41-42; Stefan Ernst in Paal and Pauly (n 7), art. 4 DS-GVO para 143.

[19] cf. Ernst (n 18) 147.

[20] cf. WP29 (n 16) 5.

[21] Recital 66 Directive 2009/136/EC.

[22] Adrian Schneider in Simon Assion (ed), *TTDSG* (1st edn, Nomos 2022), s 25 para 40; cf. WP29 (n 16) 8; recital 47 GDPR.

service must be considered granularly in terms of its function.

22 In the first and second Scenario, there may be some doubt as to whether the user explicitly requested the specific update information service. While it can be assumed that the user has requested the vehicle connection service, the user has not explicitly requested to be informed via push notifications on available updates. However, there is a closer link to the provision of the vehicle connection service in the case of essential updates, where the use of the service would be disrupted if not installed on time. In such cases, the sending of a push notification is covered by art. 5(3)(2)(alt. 2) e-privacy Directive.

23 In addition, the permission given through the system prompt can be considered as a request for the respective service. By authorizing push notifications for the app, the user expects to receive such push notifications. The question of whether the respective notifications can still be considered "explicitly requested" in accordance with the user's legitimate expectation depends on the scope of the notification's purposes pursued with the app and the frequency with which notifications are sent. If, as in Scenarios 1 and 2, a vehicle connectivity app only sends relevant connectivity notifications, these are still covered by the explicit request. However, supplementary advertising messages as in Scenario 3 may be assessed differently.

24 In Scenario 3, the small-scale analysis requires that the information on discounts for additional vehicle features be considered as a separate service or as a separate part of the same service. The discount notification promotes a service that is subject to a separate contract. The information on the option to conclude another contract is not expected by the user when the app is installed and the vehicle connection features are activated or when the user gives permission to receive push notifications in general.

3. Consent Given by the User

25 In the absence of any applicable exceptions, the service provider may rely on the user's freely given and informed consent (art. 5(3)(2), recital 17 e-privacy Directive).[23] The consent required under the e-privacy Directive generally adheres to the same principles as those set out in the GDPR.[24] The operating system's permission prompt for allowing push notifications could potentially function as a consent prompt, which will be assessed below.

a.) Standards in Comparison to the GDPR

26 Art. 5(3)(2) e-privacy Directive requires that the user "has given his or her consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing". This could be interpreted as adopting the consent requirements as provided for under the Data Protection Directive 95/46/EC and, nowadays, the GDPR (art. 94(2)(1) GDPR) without any modifications. However, an alternative interpretation of the language in art. 5(3)(2) e-privacy Directive could be that it refers only to the general *information* requirements of processing (e.g., arts. 4(11), 13-14 GDPR).

27 The different interpretations are relevant with regard to the information about the right to withdraw the consent (art. 7(3)(3) GDPR). If the reference is limited to specific information and does not encompass the information on the right to withdraw the consent under art. 7(3)(3) GDPR, the requirements for the consent could be met more easily by the system permission prompts (see below b.). Nevertheless, several aspects indicate that the reference includes the information on the right to withdrawal under the GDPR: the language employed in art. 7(3)(3) GDPR ("inter alia") as well as the interest of the user in withdrawing the consent and being informed about it and the need for a unified standard under GDPR and e-privacy Directive, which can seamlessly interlock in their application.[25]

28 Accordingly, this consent is subject to the same standards set out in the GDPR, including information on the right of withdrawal.

b.) Operating System's Permission Prompt

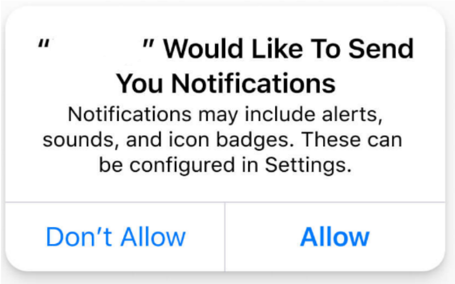
29 The push notification permission prompt triggered by the app provider as in Figure 2 and Figure 3 constitutes a valid consent if the GDPR requirements are met. According to art. 6(1)(a) GDPR, the controller is not required to obtain the consent directly; rather, any party, including push notification or app store

[23] *Planet49 GmbH* (n 13) 50-65.

[24] *Planet49 GmbH* (n 13) 60 et seqq.

[25] Stefan Hanloser in Sibylle Gierschmann and Ulrich Baumgartner (eds), *TTDSG* (1st edn, CH Beck 2023), s 25 TTDSG para 79; Peter Schmitz in Martin Geppert and Raimund Schütz (eds), *Beck'scher Kommentar zum TTDSG* (5st edn, CH Beck 2023), s 25 TTDSG para 46; LfD Niedersachsen, 'Handreichung: Datenschutzkonforme Einwilligung auf Webseiten' (November 2020), p 3 <https://lfd.niedersachsen.de/startseite/themen/internet/datenschutzkonforme-einwilligungen-auf-webseiten-anforderungen-an-consent-layer-194906.html> accessed 15 November 2024; Diana Ettig in Taeger and Gabel (n 8) s 25 TTDSG para 34; cf. *Planet49 GmbH* (n 13) 60-64; Schneider (n 22) 32.

service providers such as Apple and Google, can obtain the consent for the specific purpose on the controller’s behalf.[26]

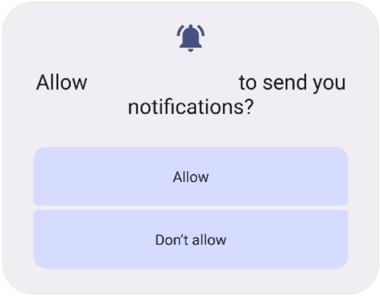


"   " Would Like To Send You Notifications

Notifications may include alerts, sounds, and icon badges. These can be configured in Settings.

Figure 2: Example of a previous iOS push notification permission prompt.

30 It is possible that details of the design of the permission prompt window may have an impact on the validity of the consent. For example, the display of the prompt in a previous iOS version (Figure 2), highlights the “Allow” option in bold. This highlighting of the option to give consent could arguably be considered stirring[27] as part of a dark pattern, potentially[28] impairing the voluntariness of the consent (art. 4(11) GDPR) and violating the principle of fairness (art. 5(1)(a) GDPR).[29]



Allow   to send you notifications?

Figure 3: Example of Android push notification permission prompt.

31 The prompt in the most recent iOS versions and the Android prompt as in Figure 3 does not highlight one of the options. Aside from this issue, there are two additional challenges that need to be considered in order for the permission granted via the system prompt to be considered a valid consent under the e-privacy Directive and the GDPR.

32 Firstly, the system permission prompt does not distinguish between different categories of push notifications such as information on available updates (see Scenarios 1 and 2), reminders, and advertising (see Scenario 3). The access to device data or the storage of data for the purpose of sending push notifications with such entirely different content is subject to different purposes within the meaning of art. 6(1)(a) GDPR. A general consent to all such notifications is incompatible with the “specific” prong in art. 6(1)(a) GDPR and the “freely given” prong in art. 4(11) GDPR.[30]

33 Secondly, it is evident from Figure 2 and Figure 3 that the system permission prompts often fail to sufficiently inform users on the right of withdrawal and the implications for the processing. While users are able to change the settings for push notifications through the operating system’s permission settings, they must be informed of this option prior to providing their consent (art. 7(3)(3) GDPR).

34 It could be argued that a general reference to the settings, as illustrated in Figure 2, suffices (“These can be configured in Settings”). However, the language in art. 7(3) GDPR clearly requires (explicit) information on the right to withdrawal and that it does not affect the lawfulness of the processing prior to the withdrawal. Such information is typically not provided in the system permission prompts.

35 Furthermore, one could argue that the majority of smartphone users are aware of the option to change their push notification settings. However, other than

[26] Tristan Radtke, *Gemeinsame Verantwortlichkeit unter der DSGVO* (Nomos 2021) 395-397; Richard Jansen and Fabian Kreis, ‘Herausforderungen bei der Datenverarbeitung im Rahmen der NEVADA Share & Secure Strategie der Automobilindustrie‘ [2020] RAW 19, 24; cf. Case C-40/17 *Fashion ID GmbH & Co. KG v Verbraucherzentrale NRW e.V.* (ECJ, 29 July 2019), ECLI:EU:C:2019:629, paras 99-102.

[27] EDPB, ‘Guidelines 03/2022 on Deceptive design patterns in social media platform interfaces: how to recognise and avoid them – Version 2.0’ (14 February 2023), paras 50-53.

[28] Taking the view that the consent would still be valid in such cases Schneider (n 22) 30. On the issue in general, Ettig in Taeger and Gabel (n 8) s 25 TTDSG para 30.

[29] EDPB (n 27). See also for online platforms art. 25, recital 67 of the Digital Services Act; Pascal Schumacher, Lennart Sydow and Max von Schönfeld, ‘Cookie Compliance, quo vadis? Datenschutzrechtliche Perspektiven für den Einsatz von Cookies und Webtracking nach TTDSG und e-privacy-VO’ [2021] MMR 603, 608 with further references.

[30] EDPB, ‘Guidelines 05/2020 on consent under Regulation 2016/679 – Version 1.1’ (4 May 2020), paras 42-45, 55-61.

art. 13(4) GDPR, art. 7(3) GDPR does not provide an exception in cases where the data subject has already been provided with the necessary information. Instead, it requires that the information be provided in any case.

36 It can be reasonably concluded that the system permission prompt does not satisfy the conditions for valid consent under the GDPR. In such cases, the responsibility for obtaining the user's consent, with the exception of the transmission process and the explicitly requested service, lies with the respective app provider.

c.) Separate Consent by the App Provider

37 If the app provider wants to rely on the user's consent as the legal basis, the app provider could implement a separate process initiated subsequent to the permission prompt, which complies with the information requirements and allows users to choose between the push notifications for different purposes.

4. Summary

38 The exceptions laid down in art. 5(3)(2) of the e-privacy Directive cover identifiers for the transmission process as well as notifications on some essential updates for the vehicle software or the mobile app. The permission obtained through the system permission prompt does not constitute a valid consent given that those prompts do not address the specific purposes and often lack sufficient information on the right of withdrawal. However, depending on the design of the app and the services offered, the general permission given by the user through the system permission prompt, could be considered an explicit request of such service and would thus be the basis for push notifications. In other cases, the app provider is required to obtain the consent of the user separately, in accordance with the consent standards set forth in the GDPR.

II. Unsolicited Communications under Art. 13 E-Privacy Directive

39 Art. 13 e-privacy Directive does not focus on the user's device per se, rather, it focuses on messages as unsolicited communications reaching the user's sphere.

1. Scope of Art. 13 E-Privacy Directive

40 Art. 13 e-privacy Directive establishes a consent requirement for communication via means such as electronic mail for the purposes of direct marketing, with the exception of the promotion of similar products or services following a sale (art. 13(1),(2) e-privacy Directive). With regard to other forms of unsolicited communication and means other than electronic mail, the Directive leaves the concrete approach to the Member States (art. 13(3) e-privacy Directive). Member States may elect to implement either an opt-in or a mechanism for excluding users who do not wish to receive the communications. However, as apparent from the draft of the e-privacy Regulation from 2021, under a future e-privacy Regulation, the distinction between electronic mail and other forms of electronic communication may become almost obsolete.[31]

41 With regard to push notifications, it has to be determined whether they, firstly, constitute a form of direct marketing within the meaning of art. 13 e-privacy Directive. If this were not the case, art. 13 e-privacy Directive would not apply to push notifications. Secondly, it has to be assessed whether push notifications are considered either as electronic mail (art. 13(1),(2) e-privacy Directive) or as other forms of communication (art. 13(3) e-privacy Directive).

a.) Direct Marketing

42 The e-privacy Directive does not provide a definition of the term direct marketing. The definition of the similar term of advertising under art. 2(a) Directive 2006/114/EC addresses traders only.[32] However, the definition in art. 2(d) UCP Directive considering direct marketing as form of commercial

[31] See art. 16 of the draft of the e-privacy Regulation, Council of the European Union (n 11).

[32] Helmut Köhler in Helmut Köhler, Joachim Bornkamm and Jörn Feddersen (eds), *Gesetz gegen den unlauteren Wettbewerb* (42nd edn, CH Beck 2024), s 7 UWG para 149.

communication, an ECJ judgement[33] and guidelines by the authorities[34] suggest that direct marketing means the communication addressed directly and individually to a person in connection with the promotion, sale or supply of a product or service.

43 The direct marketing prong hinges on the content of the push notification and whether this direct communication promotes products or services. While in Scenario 3, the content itself constitutes direct marketing, for push notifications in other Scenarios the classification depends on the link to the supply of services. However, mere information without any connection to the promotion of services (as it is the case with regard to editorial information)[35] or the fact that the push notification reminds the user of the app – which is already installed – is not sufficient to constitute direct marketing.[36] Furthermore, information that the provider is legally obliged to provide may lack the promotional intent required for the marketing requirement (e.g., on necessary updates).[37]

44 However, strong indicators of a link to the supply of an *additional* service include: advertising for third parties in the app, if the app allows users to subscribe to additional services (e.g., subscription to over-the-air-updates, in-car internet access, battery capacity upgrade) or if the app is used to gain commercial advantage by analyzing user behavior – regardless of whether the user has given his or her consent under data protection law. If one of these non-exhaustive factors is present and there is a strong link to the content of the push notification, the push notification could be interpreted as relating to the promotion of (additional) services. This is because the direct marketing requirement must be interpreted broadly in line with the above definition.[38]

b.) Electronic Mail and other Forms of Communication

45 Push notifications could be considered either as electronic mail with a strict consent requirement or as other forms of communication with opt-in or opt-out requirement depending on the legislation of the respective Member State.

46 The term “electronic mail” is defined as “any text, voice, sound or image message sent over a public communications network which can be stored in the network or in the recipient’s terminal equipment until it is collected by the recipient” (art. 2(2)(h) e-privacy Directive). In recital 67 of the amending Directive 2009/136/EC, it is stated that the term “electronic mail” should be interpreted in a broad sense and that it should also apply to SMS, MMS, and similar means of electronic communication.[39] From these sources and the comparison with SMS and MMS, it can be concluded that mail requires an inbox as a local or online collection of received messages. This inbox typically prompts the user to go through the messages as a list, which makes it more likely for advertising to be noticed by the user than it is the case for other means of communication.[40]

47 Push notifications do not utilize an inbox in the same way as email or SMS; both of which may be the subject of a push notification. The respective operating system does indeed collect the push notifications and provides the user with an overview of the notifications received. However, this categorized overview does not adhere to the conventional rules of an inbox and is instead selective and temporal in nature.

48 For example, limits might apply to an app or website sending multiple push notifications to a user,[41] the app provider could suppress the notification from being displayed beforehand[42] or the push notification could be discarded before delivery if the user’s device is offline for a long time.[43] Unlike emails, a notification is often deleted as soon as the corresponding app is opened. The notification might not even be stored until retrieval by the user’s device. In addition, push notification providers have

[33] Case C-102/20 *StWL Städtische Werke Lauf a.d. Pegnitz GmbH v eprimo GmbH* (ECJ, 25 November 2021), ECLI:EU:C:2021:954, para 47.

[34] DSK, ‘Orientierungshilfe der Aufsichtsbehörden zur Verarbeitung von personenbezogenen Daten für Zwecke der Direktwerbung unter Geltung der Datenschutz-Grundverordnung (DS-GVO)‘ (February 2022), 3.

[35] cf. Köhler (n 32) s 2 UWG para 2.70.

[36] cf. Köhler (n 32) s 2 UWG para 2.36.

[37] Hans-W. Micklitz and Martin Schirmbacher in Gerald Spindler and Fabian Schuster (eds), *Recht der elektronischen Medien* (4th edn, CH Beck 2019), s 7 UWG para 173; but see Köhler (n 32) s 2 UWG para 2.52, in particular for misleading information.

[38] E.g., Christian Alexander in Peter W Heermann and Jochen Schlingloff (eds), *Münchener Kommentar zum Lauterkeitsrecht* (3rd edn, CH Beck 2020), s 5a UWG para 106; Köhler (n 32) s 2 UWG para 2.42; Case I ZR 57/05 (BGH, 19 April 2007), para 27; cf. *StWL Städtische Werke Lauf a.d. Pegnitz GmbH v eprimo GmbH* (n 33) 47-48.

[39] *StWL Städtische Werke Lauf a.d. Pegnitz GmbH v eprimo GmbH* (n 33) 38-39.

[40] cf. *StWL Städtische Werke Lauf a.d. Pegnitz GmbH v eprimo GmbH* (n 33) 41.

[41] E.g., *Firefox Support* (n 4).

[42] E.g., ‘com.apple.developer.usernotifications.filtering’ (*Apple Developer*) <https://developer.apple.com/documentation/bundleresources/entitlements/com_apple_developer_usernotifications_filtering> accessed 15 November 2024.

[43] E.g., *Apple Developer* (n 5); ‘About FCM messages’ (*Google Firebase*) <https://firebase.google.com/docs/cloud-messaging/concept-options> accessed 15 November 2024.

announced summaries of push notifications, and it remains to be seen whether the system may further filter push notifications in this context.[44]

49 Overall, push notifications do not follow the typical inbox and list procedure but are rather selectively and temporarily stored for the user. Instead of being relevant until the user reacts by replying, forwarding or deleting the message, the push notification is usually only relevant for a short time frame and serves as a reminder in connection with the respective app.

50 At first sight, the ECJ's finding that the display of randomly generated and only temporarily stored advertising within an inbox suffices[45] could support the view that any temporarily stored message nevertheless falls within the scope of art. 13(1) e-privacy Directive. However, the ECJ did primarily contest the fact that the temporarily stored advertising was displayed as part of the inbox for electronic mail.[46] As push notifications are displayed separately and not as part of an inbox, this argument cannot be applied to push notifications.

51 Thus, push notifications do not constitute electronic mail.[47]

2. Requirements

52 As push notifications fall within the scope of art. 13(3) e-privacy Directive,[48] providers are obliged to comply with the applicable implementation at the level of the Member State. This entails either obtaining consent or refraining from sending push notifications to users, who do not wish to receive the notifications. The latter may be indicated, e.g., by the app's settings, which allow users to specify the types of notifications they wish to receive, or to indicate whether they wish to be informed about previous notifications. This is subject to the condition that the tracking of such reactions is permissible under data protection law.

53 Insofar as the push notification provider allows users to choose an interruption level (e.g., "passive", "active", "time sensitive", and "critical" for iOS users),[49] the selection of an inappropriate interruption level has to be considered when assessing the compliance with the user's wish. This is because the language in art. 13(3) e-privacy Directive considers the specific unsolicited communication ("these communications"), which allows for the consideration of the specific circumstances of such a message. The general classification of different means of communication (e.g., art. 13(1) e-privacy Directive) and the consideration of a certain *circumstance* in art. 13(2) e-privacy Directive supports this finding.

C. GDPR

54 The GDPR lays down requirements for the processing of personal data and sets requirements for push notifications to the extent personal data is processed.

55 In light of the e-privacy Directive's status as *lex specialis* and the GDPR's as *lex generalis* (art. 95 GDPR),[50] the GDPR does not impose additional requirements pertaining to the legal basis for accessing a user's device and storing information including potential unsolicited communication on the user's device.[51] However, the sending of a push notification entails the processing of personal data prior to and subsequent to the access to the user's device. In such instances, the GDPR, and in particular art. 6 GDPR, applies.[52]

[44] 'Introducing Apple Intelligence, the personal intelligence system that puts powerful generative models at the core of iPhone, iPad, and Mac' (Apple, 10 June 2024) <https://www.apple.com/newsroom/2024/06/introducing-apple-intelligence-for-iphone-ipad-and-mac/> accessed 15 November 2024.

[45] *StWL Städtische Werke Lauf a.d. Pegnitz GmbH v eprimo GmbH* (n 33) 63.

[46] *StWL Städtische Werke Lauf a.d. Pegnitz GmbH v eprimo GmbH* (n 33) 46.

[47] Taking a different view Julia Höltge, 'Werbung über mobile Push-Dienste' [2015] ITRB 223, 223.

[48] For art. 13 in general, EDPB, 'Guidelines 8/2020 on the targeting of social media users – Version 1.0' (2 September 2020), 17.

[49] 'Managing notifications' (*Apple Developer*), <https://developer.apple.com/design/human-interface-guidelines/managing-notifications> accessed 15 November 2024.

[50] Papakonstantinou and De Hert (n 8) 1.

[51] EDPB (n 9) 40; Tilman Herbrich and Elisabeth Niekrenz, 'Privacy Litigation Against Real-Time Bidding Data-driven online marketing: Enforcing the GDPR by protecting the rights of individuals under civil law' [2021] CRi 129, para 50; Carlo Piltz in Peter Gola and Dirk Heckmann (eds), *Datenschutz-Grundverordnung* (3rd edn, CH Beck 2022), art. 95 DS-GVO para 23; Golland (n 8) 23. Taking another view Maximilian Becker, 'Consent Management Platforms und Targeted Advertising zwischen DSGVO und e-privacy-Gesetzgebung' [2021] CR 87, para 55.

[52] EDPB (n 9) 40-41; Herbrich and Niekrenz (n 51) 65; Wolf-Tassilo Böhm and Valentino Halim, 'Cookies zwischen e-privacy und DS-GVO – was gilt? – Anforderungen an die Verwendung von Cookies nach der aktuellen Rechtsprechung' [2020] MMR 651, 653; cf. Ettig in Taeger and Gabel (n 8) s 25 TTDSG para 12.

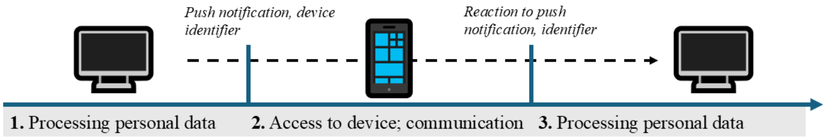

Push notification, device identifier
Reaction to push notification, identifier
1. Processing personal data
2. Access to device; communication
3. Processing personal data

Figure 4: Underlaying technical steps for the interplay between the GDPR (1. and 3.) and the e-privacy Directive (2.).

I. Scope of the GDPR

56 Pursuant to art. 2(1) GDPR, the GDPR applies to any processing of personal data, which is broadly defined as any operation which is performed on "any information relating to an identified or identifiable natural person ('data subject')" (art. 4(1), (2) GDPR). According to the case law of the ECJ, the information constitutes personal data from the perspective of the controller as defined in art. 4(7) GDPR, if the controller has available "means likely reasonably to be used either by the controller, [...] or by any other person, to identify that person, without, however, requiring that all the information enabling that person to be identified should be in the hands of a single entity".[53]

57 In all Scenarios, the push notification is sent to a specific device of a data subject. The processing of the push notification in order to transmit and deliver it to the device entails the processing of personal data, including the device token and the content of the notification. In general, the app provider links the device token to account data or other user data on its servers. Even if the respective app provider does not have access to the device tokens but is nevertheless able to trigger a general push notification to all registered devices, the natural person in question can be identified by the app provider through the use of reasonable means. In fact, the app provider can request further information about the specific device and potentially the user at any time.

58 Thus, the GDPR applies to the process of sending and receiving a push notification except for the final step of the storage on the user's device. In light of the fact that the app provider determines the purpose and means of the processing by initiating the processing,[54] including the decision on the user as the data subject, the content, and other modalities, the app provider is to be regarded as controller under art. 4(7) GDPR. However, if another person or entity exerts influence in its own interest over the push notification (e.g., extensive filtering by the push notification provider or a third party pays for advertising), such person or entity might be considered the controller or joint controller under art. 26 GDPR.[55] This applies even in cases where such person or entity lacks access to personal data.[56]

II. Legal Basis

59 Pursuant to art. 6 GDPR, any processing activities must be supported on a legal basis. In the context of push notifications, the following legal bases are particularly relevant: the data subject's consent (art. 6(1)(a) GDPR), the necessity for the performance of a contract (art. 6(1)(b) GDPR) and the balancing of interests (art. 6(1)(f) GDPR).

1. Consent

60 In order to obtain consent in accordance with arts. 4(11), 6(1)(a) GDPR,[57] the standards and requirements set out above apply (see B. I. 3.). Consent under the e-privacy Directive and for the upstream and downstream processing operations under the GDPR can be jointly[58] obtained if all operations serve similar, specific purposes[59] within the meaning of arts. 4(11), 5(1)(b), art. 6(1)(a) GDPR.

2. Performance of a Contract

61 Nevertheless, the processing is also lawful if it is "necessary for the performance of a contract to which

[53] Case C-319/22 *Gesamtverband Autoteile-Handel eV v Scania CV AB* (ECJ, 9 November 2023), ECLI:EU:C:2023:837, para 45; Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* (ECJ, 19 October 2016), paras 42-43.

[54] cf. *Fashion ID* (n 26) 75, 78.

[55] cf. Case C-25/17 *Tietosuojavaltuutettu, Jehovan todistajat – uskonnollinen yhdyskunta* (ECJ, 10 July 2018), ECLI:EU:C:2018:551, para 68.

[56] Case C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* (ECJ, 5 June 2018), ECLI:EU:C:2018:388, para 38.

[57] See *Planet49 GmbH* (n 13).

[58] cf. Björn Steinrötter, 'Anforderungen an die Einwilligung des Internetnutzers beim Setzen und Auslesen von Cookies' [2020] GPR 106, 109.

[59] cf. Marion Albers and Raoul-Darius in Heinrich Amadeus Wolff, Stefan Brink and Antje v. Ungern-Sternberg (eds), *BeckOK Datenschutzrecht* (48th edn, CH Beck, 1 May 2024) art. 6 DS-GVO para 32; Giovanni Sartor in Spiecker gen. Döhmann and others (n 8), art. 6 para 19.

the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract" (art. 6(1)(b) GDPR). Necessity must be determined from an objective perspective and has to take into account the main obligations under the contract as mutually agreed by the parties.[60] Features that are merely useful for the performance of the contract, such as personalized advertising in a social network rather than non-personalized advertising, are thus not covered by art. 6(1)(b) GDPR.[61]

62 In the first Scenario, there are likely several contractual relationships depending on the individual circumstances. For example, the user may have bought the vehicle and thereby concluded a consumer contract for the sale of a good (cf. Directive (EU) 2019/771). The user may have entered into a separate contract for the use of connected vehicle services and a contract for the downloading and utilization of the smartphone app (cf. Directive (EU) 2019/770). In order to determine the contractual relationships in question, it is crucial to ascertain whether such vehicle network and smartphone app services form an integral part of the contract for the purchase of the vehicle (cf. art. 3(4) Directive (EU) 2019/770).[62]

63 In the case of security updates, the vendor of the vehicle or the app provider may be under a contractual obligation to provide such updates (e.g., art. 8(2) Directive (EU) 2019/770). Providing information on these updates via push notification is a secure and admissible way of notifying the user of the available update. In this case, it could be considered that the processing is necessary for compliance with the legal obligation deriving from art. 8(2) Directive (EU) 2019/770 and within the meaning of art. 6(1)(c) GDPR. However, in accordance with art. 6(3)(1) GDPR, the specific purposes of the processing must be clearly outlined in the legal basis.[63] The national implementation of art. 8(2) Directive (EU) 2019/770 does not provide for a *notification* obligation and also presupposes a contractual relationship. Therefore, recital 38(1),(2) Directive (EU) 2019/770 refers particularly to the legal basis of the necessity for the performance of the contract as laid down in art. 6(1)(b) GDPR.

64 However, the necessity of processing for the performance of the contract depends on whether the data subject is not sufficiently informed by other means. For example, the notification may not be considered necessary if the vehicle or, in the case of Scenario 2, the app store automatically provides and installs the app updates in due time and the connection to the vehicle in the meantime is maintained (i.e., the update is not urgent).

65 Assuming that this requirement is met, the installation of updates is often linked to the continuous provision of the vehicle connection services. As a result, the processing is necessary for the performance of the respective contracts.

66 With regard to Scenario 2, similar considerations apply. Updates that are essential for maintaining the connection to the vehicle, which form the main purpose of the app and the respective contract, and the information conveyed via a push notification may be covered by art. 6(1)(b) GDPR, provided that there are no more effective means of installing the update.

67 In the third Scenario, the information on the availability of the discount could be considered useful for the user. However, such an upgrade is not part of the same contract (see above B. I. 2. b., and the vehicle connection services can be provided without the information on the upgrade. Therefore, processing for a push notification in Scenario 3 would not be considered necessary.

3. Balancing Interests

68 In particular for the third Scenario, art. 6(1)(f) GDPR could be considered the applicable legal basis. Art. 6(1)(f) GDPR permits processing which "is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child". Accordingly, the ECJ requires (1) a legitimate interest, (2) for which the processing is necessary, and (3) the interests and rights of the data subjects must not override the legitimate interest.[64]

[60] Case C-252/21 *Meta Platforms Inc. and others v Bundeskartellamt* (ECJ, 4 July 2023), ECLI:EU:C:2023:537, para 98; EDPB, 'Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects – Version 2.0' (8 October 2019), paras 30-32.

[61] *Meta Platforms* (n 60) 102.

[62] See Tristan Radtke, 'Das Recht des Streamings im Vergleich mit dem herkömmlichen Kaufrecht' in Gregor Albers and Hanjo Hamann (eds), *Vertrieb und Vertrag auf der Schwelle zur Dienstleistungswirtschaft* (Mohr Siebeck, forthcoming).

[63] Marion Albers and Raoul-Darius in Wolff, Brink and v. Ungern-Sternberg (n 59) art. 6 DS-GVO para 48; for examples for legal obligations see EDPB, 'Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects' (8 October 2019), paras 44 (example 4) and 47 (example 6).

[64] Case C-597/19 *Mircom International Content Management*

69 Any economic interest fulfils the requirement of the legitimate interest (cf. recitals 47, 48 GDPR). This is particularly the case where there is a contractual relationship between the controller and the data subject (recital 47(2) GDPR). The reasonable expectation of the data subject (cf. recital 47(3),(4) GDPR) is an important factor in determining the weight of the interests of the data subject.

70 In Scenarios 1 and 2, the legitimate interest of the app provider in informing users of software updates to maintain the services is particularly strong if the information on the update is necessary to maintain the security of the vehicle and the app (cf. recital 49 GDPR). From the user's perspective, there is an interest in the protection of their personal data (art. 8 Charter) and the right to be protected against unsolicited communications on their devices, including the processing prior to delivering such communications (cf. art. 7 Charter). In light of the aforementioned considerations, it is reasonable to conclude that the rights and interests of data subjects do not prevail in Scenarios 1 and 2, contingent on the design of the app and the user expectations shaped by it (see above under B. I. 2. b.), as well as the frequency of processing (i.e., the frequency of notifications).

71 The processing of personal data for marketing purposes, including direct marketing, can also serve a legitimate interest (recital 47(7) GDPR).[65] Nevertheless, data subjects have the right to object to the processing at any time, without giving reasons (art. 21(3) GDPR). In the light of the aforementioned, even the occasional dissemination of information via push notifications regarding discount offers within the app, as in Scenario 3, may be justified as form of direct marketing on the basis of art. 6(1)(f) GDPR.

72 However, the interests of data subjects may prevail if the content and timing of the advertising is personalized in such a way that it is based on excessive behavioral targeting in the form of profiling[66] (arg. art. 35(3)(a), recital 60(3),(4) GDPR) or if third parties process personal data in connection with push notification advertising for third party services.

4. Summary

73 Consent appears to be a practical way forward under the GDPR, as it can be combined with the consent required under the e-privacy Directive. However, the processing of personal data for necessary and urgent updates, as potentially in Scenario 1 and 2, may be necessary for the performance of the contract. Notifications in those Scenarios regarding less urgent updates might be based on the balancing of interests. In Scenario 3 and other scenarios, under certain conditions, the balancing of interests or, in any case, the consent obtained by the app provider allows for the processing to prepare the delivery of a push notification and for post-delivery processing.

III. Further Requirements

74 In addition to its existing obligations, the controller is subject to further requirements under the GDPR. It is important to note that such requirements pertaining to push notifications are interlinked with those discussed above. For instance, controllers must comply with the data processing principles under art. 5 GDPR, including the lawfulness under art. 5(1)(a) GDPR, and must inform data subjects pursuant to arts. 13, 14 GDPR. Default settings for push notifications within an app (e.g., for fine-tuning the content and frequency of notifications) must be designed in compliance with data processing principles such as data minimization (art. 25(1),(2), art. 5(1)(c) GDPR).

D. Further Regulation

75 In regard to push notifications and their content, the relevant legislation, such as the UCP Directive and the ecommerce Directive (both as amended), provides less specific provisions.

I. UCP Directive

76 The amended UCP Directive 2005/29/EC subjects commercial practices to additional requirements. The term "commercial practice" refers to "any act, omission, course of conduct or representation, commercial communication including advertising and marketing, by a trader, directly connected with the promotion, sale or supply of a product to consumers" (art. 2(d) UCP Directive). Consequently, the requirements set forth in the UCP Directive apply to push notifications that are directly connected to the promotion of the app and the provided services

& Consulting (M.I.C.M.) Limited v Telenet BVBA (ECJ, 17 June 2021), ECLI:EU:C:2021:492, para 106; case C-13/16 *Valsts policijas Rīgas reģiona pārvaldes Kārtības policijas pārvalde v Rīgas pašvaldības SIA „Rīgas satiksme“* (ECJ, 4 May 2017), ECLI:EU:C:2017:336, para 28.

[65] See also Becker (n 51) 66.

[66] Article 29 Data Protection Working Party, 'WP251rev.01' (6 February 2018), 14-15.

(cf. above under B. II. 1. a.).[67]

77 For example, art. 7(2) UCP Directive considers the absence of an indication of commercial intent within a commercial practice to be a misleading omission. This is the case where the commercial intent is neither apparent from the context nor identified. The commercial link between the push notification and the promotion of the app is typically apparent from the name and icon of the app as included in the push notifications. However, in instances where an additional commercial intent is not apparent from the context, as may be the case in Scenario 3, the app provider is required to identify this commercial intent, e.g., by declaring the push notification as "advertising".[68]

78 The misleading use of app name and icon to mislead users into believing that another provider is responsible for the notification and the app may be prohibited under art. 5(1),(5), Annex I(13) UCP Directive without prejudice to claims under intellectual property law. This is particularly relevant for apps under less strict scrutiny of app store providers, e.g., third party apps provided on alternative app distribution platforms.[69]

79 The sending of persistent and unwanted solicitations is prohibited under art. 5(1),(5), Annex I(26) UCP Directive. This practice is of less relevance for push notifications, as users are able to indicate their wish regarding push notifications and, furthermore, to prevent an app and its provider from sending the user push notifications by means of the operating system settings. However, if the app provides for finer adjustments to notifications and does not respect the indicated settings, this prohibition could apply.

II. Ecommerce Directive

80 The amended ecommerce Directive 2000/31/EC applies to information society service providers within the meaning of art. 1(1)(b) Directive (EU) 2015/1535. Such services include apps and sent push notifications (see above B. I. 2. b.).

81 In accordance with art. 5 ecommerce Directive, app providers are obliged to make information such as the name of the provider easily, directly and permanently accessible. In the case of push notifications, this requirement is satisfied if the relevant app interface allows for the information to be accessed with ease.[70]

82 However, in addition to the attribution of the push notification to a particular application and its associated interface, the identification requirement set forth in art. 6(b) of the ecommerce Directive also necessitates the assignment of a unique and distinctive combination of an app name and app icon.

83 Similar to art. 7(2) UCP Directive, commercial communication has to be clearly identifiable as such under art. 6(a) ecommerce Directive.

E. Conclusion

84 Push notifications have become an important means to inform users directly. Although sending push notifications might appear straightforward given the permissions obtained by each app, these notifications raise complex legal issues, particularly under the e-privacy Directive and potentially under a future e-privacy Regulation.

85 Despite the impression the system permission prompts for push notifications might give, such permissions do not constitute a valid consent under art. 5(3) e-privacy Directive for the temporary storage of the notification on the user's device. In this respect, push notifications clearly demonstrate the requirements for consent in the interaction between the e-privacy Directive and the GDPR.

86 Nevertheless, contingent on the configuration of the app and the scope of services it offers, as well as the frequency of notifications, such permission may be construed as an explicit request for the notification service, in accordance with art. 5(3)(2) of the e-privacy Directive. Accordingly, the interpretation of the concept of a service, whether narrow or broad, is crucial for the application of art. 5(3) of the e-privacy Directive. Push notifications with marketing and advertising content, by contrast, regularly require consent under the e-privacy Directive.

[67] Micklitz and Schirmbacher (n 37); cf. Boris Paal and Dominik Nikol, 'Spendenwerbung durch E-Mail-Direktmarketing zwischen UWG und DSGVO' [2023] GRUR 781, 784 for the relationship between commercial practice and direct marketing.

[68] In detail Tristan Radtke, 'Disclosure Requirements for Influencer Marketing in the U.S. and Germany' (2022) 12 JIPEL 141, 147-154. See also art. 5(5), Annex I(11) UCP Directive for advertorials, which is of less relevance for push notifications.

[69] See recently for Apple devices 'About alternative app distribution in the European Union' (*Apple*) <https://support.apple.com/en-us/118110> accessed 15 November 2024.

[70] cf. Case I ZR 228/03 (BGH, 20 July 2006).

87 In light of the case law of the ECJ and the meaning and purpose of the characteristic of electronic mail, push notifications are not to be considered as electronic mail within the meaning of art. 13(2) e-privacy Directive. Thus, app providers must comply with the requirements of art. 13(3) e-privacy Directive as implemented by the Member States when sending push notifications. In instances where Member States have elected to implement an alternative approach that excludes users who do not wish to receive communications, as opposed to opt-in, a relative approach is applied, allowing for consideration of the circumstances of the individual communication. The specific purpose of the communication, the frequency and the application settings have been identified as such relevant circumstances. As things stand at present, an e-privacy Regulation may abandon special provisions for electronic mail and establish uniform standards for electronic communication.

88 The GDPR applies to the processing of personal data both before and after the delivery of the push notification. Consequently, the e-privacy Directive and the GDPR are complementary and require a clear distinction between the individual storage of information and processing activities. For such processing activities, the legal basis under the GDPR is often the necessity for the performance of a contract under art. 6(1)(b) GDPR or the balancing of interests under art. 6(1)(f) GDPR. With regard to contracts for connected products, the different contracts need to be assessed carefully taking into account regulations such as the Directive (EU) 2019/770. The balancing of interests requires consideration of factors similar to those under the e-privacy Directive. Forms of direct marketing might fall within art. 6(1)(f) GDPR. For other cases, the processing can be based on the consent under art. 6(1)(a) GDPR, which can often fulfil the consent requirements under the e-privacy Directive at the same time.

89 This complex interplay of e-privacy and data protection is further compounded by other legal acts, such as the UCP Directive. The resulting transparency requirements assume particular significance with regard to apps that are downloaded via unofficial app stores, a phenomenon that has only recently become possible on Apple devices.

Copyright and Generative AI: Opinion

by **Séverine Dusollier, Martin Kretschmer, Thomas Margoni, Peter Mezei, João Pedro Quintais, Ole-Andreas Rognstad ***

Executive Summary: The ECS considers that the current development of generative artificial intelligence (AI), under the regulatory framework set up by the Directive on Copyright in the Digital Single Market (CDSM) of 2019 and the AI Act of 2024 (Regulation (EU) 2024/1689), leaves legal uncertainties and several open questions. The following issues require, in the view of the ECS, urgent consideration by the European Union:

1. The determination of the scope of the text and data mining (TDM) exception: the exception enacted in Arts. 3 and 4 of the CDSM Directive at a time when the Generative AI development could not have been fully anticipated, can be interpreted as covering some operations of training of a Generative AI model, but certainly not all aspects or stages of the life cycle of AI models and systems, from curating a dataset for training to the generation of an image, text or other media, by users. The exact scope of the TDM exception, and hence the copyright status of acts carried out at each stage of development and operation of Generative AI models and systems, should be further studied and analysed. That would require a decision as to whether acts of reproduction or public communication occur and which actors are liable for such acts. Under such an assessment, the possibility of commercial use of models trained for scientific research and the effect of the exercise of the opt-out provided by Art. 4 CDSM Directive, on the availability of lawfully accessible sources for the research exception provided by Art. 3 CDSM Directive, merit particular attention.

2. The content of the obligation under Art. 53(1)(c) of the AI Act related to the reservations of rights: in particular, the technologies that can be used to express the opt-out should be identified and regularly reviewed; the rightholders entitled to opt-out and the opt-out modalities, including the timing and the location, should be clarified.

3. The scope and modalities of the transparency obligation laid down by Art. 53(1)(d) of the AI Act: in particular, the relevant information to be included in the summary and the impact of the transparency obligation on the assessment of the lawful access criterion contained in Arts. 3 and 4 CDSM Directive should be clarified.

4. The privileges for research and for open source models: the importance of research and the key role of open source data and software in the field of AI should guide the interpretation of the CDSM Directive and the AI Act. This would lead to needed clarification of some of their provisions, with the objective of preserving the fundamental rights of research, academic freedom and education. The uncertainties raised by the Hamburg court decision in the LAION case, as to the interface between Art. 3 and Art. 4 of the CDSM Directive, should particularly be addressed in order to avoid general purpose AI (GPAI) model providers relying on training for the purposes of research, hereby escaping the more restrictive frame of the exception of Art. 4.

5. The articulation between the CDSM Directive and the AI Act: the CDSM directive is a private law instrument organizing a protection of private rights on a territorial basis, whereas the AI Act is a public law that regulates the safety of AI products, as a condition for importation and use in the EU. That raises several issues in the articulation of both legislative texts, notably the territorial scope of the obligations imposed, the entities covered by the different obligations, the effect of the AI Office's voluntary Code of Practice, the distinct modes of enforcement of the obligation laid down by the CDSM Directive and by the AI Act. These points should be clarified.

6. The fair remuneration of authors and performers for all acts of exploitation of their works and performances occurring in the life cycle of Generative AI models and systems (including when an opt-out from the application of Art. 4 CDSM Directive has been exercised and when their works or performances are included in a dataset that has been licensed to an AI provider) needs to be reaffirmed as a fundamental principle of the EU acquis. The Commission should look at the best ways to ensure such a remuneration, including remuneration rights or other compensation mechanisms, in concert with Member States.

Recommended citation: Séverine Dusollier, Martin Kretschmer, Thomas Margoni, Peter Mezei, João Pedro Quintais, Ole-Andreas Rognstad, Copyright and Generative AI: Opinion of the European Copyright Society, 16 (2025) JIPITEC 121 para 1

* The European Copyright Society (ECS) was founded in 2012 with the aim of creating a platform for critical and independent scholarly thinking on European Copyright Law and policy. Its members are scholars and academics from various countries of Europe, seeking to articulate and promote their views of the overall public interest on all topics in the field of authors rights, neighbouring rights and related matters. The ECS is neither funded nor instructed by any particular stakeholders. Its Opinions represent the independent views of a majority of ECS members.

Background

Before the advent and public availability of generative AI tools such as ChatGPT, Stable Diffusion, MidJourney Dall-E, GitHub or Udio, the Directive (EU) 2019/790 of 17 April 2019 on copyright and related rights in the Directive on Copyright in the Digital Single Market (CDSM Directive) enacted two exceptions to copyright and related rights to allow for text and data mining (TDM) of protected subject matter: one for purposes of scientific research, the other for any other purpose. In that latter case, the rightholders are entitled to reserve the right to authorise such TDM by opting out of the application of the exception. In 2024, in the context of a growing concern that generative AI tools could produce texts, images, music or films and impact copyright protection and remuneration of creators and artists, the Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act), in its final stage of negotiation, included obligations to providers of general-purpose AI (hereafter GPAI) models to provide transparency as to the datasets used for training their models and to put in place a standard policy for the exercise of opt-out by copyright and related rights owners.

In parallel with the growth of litigation in the US and Europe, controllers of aggregate copyright works (such as news publishers and outlets, stock images companies, or other types of content) are striking deals with technology firms about (often exclusive) access for AI model training. At the same time, authors, artists, performers are receiving new contract types from publishers, producers and collective management organisations (CMOs). These compete for assignments or clarifications about rights to train, which the respective intermediaries aim to license on to technology companies.

Due to the discussions surrounding the adequate manifestation of the CDSM Directive opt-out provision and other challenges, including in other jurisdictions, in effect the AI training space is already moving to licensing as a default.

Without contesting what has been achieved by the CDSM Directive and by the AI Act, the ECS considers that the rapid development of generative AI technology associated with the emergence of a licensing market for specific datasets, highlight some remaining uncertainties and bring new challenges that require EU intervention.

The ECS is also following with attention the drafting of the General-Purpose AI Code of Practice initiated by the Working Groups set up by the EU AI Office, particularly in relation to copyright. However, we would like to point out several pending questions, issues and uncertainties related to the combined application of both legislative texts to generative AI models and systems and copyright protection. Those remaining issues and uncertainties are of great policy relevance and are critical to innovation and to the sustainability of a distinct European creative sphere. The EU copyright *acquis* is founded in the fundamental rights framework established by the Treaties and the Charter of Fundamental Rights of the EU (Charter). The AI Act, which necessarily operates within the same framework, is perhaps even more explicit in its fundamental rights enabling objectives. Art. 1 of the AI Act puts this clearly when it states that its purpose is to "… promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection …".

As is well established by the case law of the CJEU, neither of the fundamental rights established in the Charter is absolute or prevails over the others, but they are in a constant need of balancing and dialogic conversation. The ECS believes that in the balancing exercise needed to address the many possible tensions across the multifaceted actors in the AI life cycle the following elements should operate as guiding principles:

- The interests of human authors and performers;
- The interests of users and of the wider public, anchored in the fundamental rights framework established by the Treaties and the Charter, as reminded by Art. 1 of the AI Act;
- The enhancement of research and innovation.

1. The application of the text-and-data-mining exception to generative AI operations

Although the issue has been disputed, the ECS holds the view that the TDM exceptions in Arts. 3 and 4 CDSM Directive are applicable to the development of generative AI models (a type of GPAI model as per the AI Act), albeit not necessarily covering all aspects of it. While it is not *prima facie* obvious that the reproduction right in Art. 2 of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society (Infosoc Directive) or the extraction right of Art. 7 of Directive 96/9/EC (Database Directive) apply to any case of training of generative AI models, Recital 105 of the AI Act nevertheless presupposes that "any use of copyright protected content requires the authorisation of the rightholder concerned unless relevant copyright exceptions and limitations apply". At the same time, this Recital as well as substantive provisions of the AI Act (e.g. Art. 53(1)c)) are based on the premise that the TDM provisions of the CDSM Directive are applicable to the development of generative AI systems or GPAI models. The broad phrasing of the definition of TDM in Art. 2(2) of the Directive ("any automated analytical technique aimed at *analysing text and data* in digital form in order to ***generate*** information which includes but is *not limited to* patterns, trends and correlations"; emphasis added here) supports this position.

However, although generally relevant, the TDM provisions of the CDSM Directive should not be considered as resolving all the problems concerning the use of works protected by copyright and other subject matter. The provisions of Arts. 3 and 4 of the CDSM Directive do not necessarily apply to all aspects or stages of the life cycle of a GPAI model or a generative AI system. Rather, these TDM exceptions cover different (but not necessarily all) aspects of the training stage of GPAI models, to the extent that those activities qualify as: (i) acts of TDM under the broad definition in Art. 2 of the CDSM Directive, and (ii) reproductions of protected subject matter of the type mentioned in Art. 3 and 4 of the CDSM Directive.

This raises two important questions. First, does the concept of TDM and its exceptions cover all activities taking place leading up to and including the training stage of a GPAI model? Second, is the TDM regime relevant for acts taking place once a model is trained and when outputs are generated?

Regarding the first question, the AI Act takes a clear position on the copyright-relevant nature of TDM, as already stated above. Recital 105 also mentions the rights reservation mechanism in Art. 4(3), noting that, where applicable, GPAI model providers must abide by this mechanism if they want to conduct TDM on those materials, namely by implementing the principles agreed upon by the Code of Practice to be established by the EU AI Office.

From the perspective of EU law, therefore, carrying out TDM on copyright-protected content appears in most cases to amount to reproducing a work.[1] As such, TDM requires authorization from the rightholder, or it must benefit from a copyright exception, such as those in Art. 3 and 4 of the CDSM Directive. The question that arises is whether all copyright-relevant reproductions and extractions involved in the training and development of an AI model qualify as TDM. What appears clear is that the TDM exception does not cover subsequent acts of communication to the public or the making available of TDM results. Indeed, the scope of the TDM exceptions covers only acts of reproduction and extraction. Furthermore, Art. 3(2) and 4(2) of the Directive put clear boundaries on the subsequent uses of copies of works or other subject matter made pursuant to the TDM exceptions.

Regarding the post-training operations, it should be noted that the TDM definition and the TDM exceptions do not apply to any acts taking place at a stage following to training the model. This means that they do not cover the integration of a trained GPAI model into an "AI system," its "placing on the market," "making available on the market," or its "putting into service" in the EU. They also do not cover the generation of outputs by an AI model or system.

All these activities may be relevant for copyright purposes, as they may involve restricted acts and subsequent copyright infringement. There is significant legal uncertainty about the copyright status of acts that have been labelled as memorization at the model level, as well as regurgitation, extraction, and reconstruction at the output generation stage. The integration of a dataset constituted or of a model trained under the research-related TDM exception (art. 3 of the CDSM Directive) in a GPAI model or system made available for commercial purposes is another issue that has been recently dealt with by the Hamburg District Court (the LAION case)[2] in a manner that raises many questions. These issues deserve further research and clarification.

As a result, the TDM exception and its assessment should be considered separately from the commercial exploitation, effects, or harms to creators stemming from generative AI outputs. In other words, such commercial exploitation and competition with or substitution for human (non-AI-assisted or generated) creations are not relevant – as a matter of law – to the assessment of the exception in Art. 4 of the CDSM Directive. This also has consequences for the assessment of the exception under the three-step test, as the qualification of a conflict with the normal exploitation and the assessment of unreasonable prejudice to rightholders must be considered in the context of TDM related to the training of an AI model, rather than in relation to the exploitation that takes place once the model is

[1] This policy choice of including any technical, even if fugitive, fixation of a work within the scope of reproduction right, made by the EU lawmaker as early as the 1991 directive on computer programs, could have been different and remains challenged by several copyright scholars, including some signatories of the present opinion.

[2] Landgericht Hamburg, 27 September 2024, AS. 310 O 227/23.

trained and placed on the market. By contrast, the commercial exploitation that takes place during the training stage – e.g., licensing of datasets by rightholders for third parties to carry out TDM – might be relevant to the assessment of Art. 4 of the CDSM Directive. This is a point of legal interpretation of EU law, rather than a normative pronouncement on its desirability.

2. The content of the obligation under Art. 53(1)(c) of the AI Act

According to this rule, GPAI model providers must *put in place a policy to comply with EU copyright law* in particular to identify and comply with, including through state of the art technologies, the reservations of rights (i.e. "opt-out") expressed pursuant to Art. 4(3) of the CDSM Directive.

This provision includes two main prongs: on the one hand, the requirement to 'put in place' a policy document; and, on the other hand, to identify and comply with opt-out mechanisms, that is, in essence, to guarantee the compliance with the CDSM Directive.

As regard the first prong, 'putting in place' a policy shall not only mean *drawing up* such a document, but GPAI model providers shall also *keep such policy up-to-date*, they shall also *implement* their commitments per the policy document, and, finally, to *publish* the policy document. The latter shall be understood in a broader sense: GPAI model providers shall provide access to the policy document to the general public, rather than solely to the AI Office. This is evident from the language of the AI Act itself. Art. 53(1)(a) of the AI Act introduces a limited publication obligation ('upon request, to the AI Office and the national competent authorities'); whereas Art. 53(1)(c) does not include any such limitation.

As regards the second prong, GPAI model providers' policy, in line with *effet utile*, only if capable to guarantee that rightholders can effectively opt-out their contents from the training of GPAI models.

Based on that, the Commission and the AI Office, and particularly the Working Group on transparency and copyright-related rules, have already started to work on a Code of Practice to provide guidance on (a) the *scope and modalities of the said policy requirement*; (b) the modalities and methods of the opt-out mechanism that will be considered *compliant with Art. 53(4) AI Act*.

In that process, the *effect of the compliance of GPAI model providers with the obligation under Art. 53(1)(c) AI Act* on the consideration of whether they are compliant with the rights reservation rule under Art. 4(3) CDSM Directive, should be ascertained and a special clarification is needed regarding various sub-topics. First, as provided for by Art. 56(8) of the AI Act, technologies to be used for the expression of rights reservation need to be regularly reviewed, in order to avoid the danger that a specific technological solution becomes mandatory and to ensure instead that all state-of-the-art solutions might be deployed in practice. Second, rightholders entitled to opt-out and the opt-out modalities should be expressly determined; an issue that has special importance in light of the numerous alternatives for opt-outs (developed by GPAI model providers and/or independent third parties) and the growing number of "press-release-like" reservation of rights by CMOs or licensees, e.g. publishing houses. Third, the timing of the reservation of rights should be discussed; that is, whether opt-outs preceding or following the mining of text or data are compliant with the *acquis*. Finally, the location of the expression of the reservation should be clarified; that is, whether opt-out at the source-level where the protected subject matter is stored or from where it has been made lawfully accessible and/or at the work-level, that is, via the developers' website/reservation mechanism, are covered by the *acquis*.

3. The scope and modalities of the transparency obligation laid down by the Art. 53(1)(d) of the AI Act

These rules require GPAI model providers to draw up and make publicly available a sufficiently detailed summary about the content used for training of the GPAI model (including of the generative type), according to a template provided by the AI Office.

First, our arguments expressed in the previous point on 'drawing up', 'publish' and making the relevant document 'meaningful' apply mutatis mutandis under Art. 53(1)(d) of the AI Act. Similarly, under *effet utile*, the summary shall include relevant information about how and when the providers respected opt-outs required by Art. 4(3) of the CDSM Directive.

From a copyright perspective, it is also crucial that the Commission and the AI Office clarify *how this requirement of the AI Act influences the assessment of the lawful access criterion* (or even criteria) underpinning Arts. 3 and 4 CDSM Directive and what exact information GPAI model developers shall disclose as regards such access to training data.

4. The privileges for research and for open source models

The AI Act acknowledges the importance of research in the field of AI as well as the use of AI in research activities. It therefore establishes that it does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development (Article 2(6)). The AI Act also does not apply to any research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service (Article 2(8)). These are important provisions. Yet, given the very strict definition of research, their practical effect, particularly in the context of public-private partnerships in research, remains to be ascertained.

In a similar vein, the AI Act recognizes that software and data, including models, released under a free and open-source license can contribute to research and innovation in the market and can provide significant growth opportunities for the Union's economy. Accordingly, the AI Act does not apply to AI systems released under such licenses, unless they are placed on the market or put into service as prohibited AI or as high-risk AI systems or as certain AI systems subject to specific transparency obligation (Art. 50). GPAI models under free and open-source licenses are excluded only from the provisions of Art. 53(1)(a) and (b), but must comply with those under letter c (the Policy) and letter d (the Summary). Essentially, the documentary obligations of Art. 53(1)(a) and (b), together with the exclusion for certain third parties in relation to the high-risk AI value chain (Art. 25(4)) are the only actual exemptions favoring free and open-source AI. Considering the restrictive definition adopted in the AI Act that excludes any form of monetization – a considerable deviation from the generally accepted definitions of free and open source software – the real effect of the provision, similarly to the case of research, remains unclear.

A very specific issue was highlighted by the LAION case, brought before the Hamburg district court: the potential use by commercial players of datasets mined on the basis of Art. 3 CDSM Directive. Whereas this aspect would deserve a dedicated treatment, the analysis needs to take into account, as argued above, that both the copyright *acquis* and the AI Act are grounded in the fundamental rights framework established by the Treaties and the Charter. Scientific research, academic freedom and the right to education are central in this framework and their preservation must be ensured. In the specific case of LAION, the dataset prepared did not contain the actual works needed for the successive phase of model training, but only information about their location. A GPAI model provider interested in exploiting this "preselection" would need a proper legal basis to access those sources. This legal basis would likely be Art. 4 of the CDSM Directive or a contractual agreement, in case the opt-out provided for in Art. 4(3) has been exercised. In the opinion of the ECS this approach, already logically following from the regulatory framework put in place by the interface between the Arts. 3 and 4 of the CDSM Directive and Art. 53 of the AI Act, represents a proportionate balance in the protection of the different fundamental rights at stake.

5. The relationship between the AI Act (a Regulation) and the CDSM Directive with respect to the enforcement of the copyright-related provisions

Recital 108 clarifies that the AI Act does not affect the enforcement of copyright rules as provided for under Union law; several recitals and provisions mention that the AI Act is both without prejudice to Union copyright law or meant to assist in compliance with EU copyright law.

Copyright law is an area of private law where civil enforcement is left to the owners of copyright and related rights. In part, the relationship between the AI Act and copyright law is just a clarification and assertion of such existing private interests of legal subjects, leaving enforcement within the national regimes of Member States, harmonised by the TDM provisions of the CDSM Directive.

However, an important (and entirely new) set of obligations in the AI Act need to be understood as meta-laws at the EU level. As Peukert suggests, they resemble "horizontal meta-obligations of hosting service and search engine providers under the Digital Services Act (DSA) who also have to put in place various mechanisms to act on or prevent the presence or findability of illegal content".[3]

The obligations of the AI Act about transparency and compliance with opt-out provisions of Art. 4(3) CDSM Directive imposed on the GPAI models providers are presented as if they have an *extraterritorial* effect, and could apply to the training of models outside of the EU. Breaches of obligations lead potentially to administrative fines (up to 3% of the annual total worldwide turnover or EUR 15 000 000, whichever is higher), i.e. a public law remedy rather than private enforcement.

[3] A. Peukert (2024) Copyright in the Artificial Intelligence Act – A Primer, GRUR International, 73(6), 2024, 497–509, at p. 502.

The core of the AI Act, before the late introduction of rules covering foundation models and generative AI as GPAI in Chapter 5, introduced extraterritorial implications via the concept of the *AI value chain*. Under Art. 25 of the AI Act (Chapter 3, High-Risk AI Systems, Responsibilities along the AI value chain), the prohibitions and obligations for high-risk AI systems apply to any "distributor, importer, deployer or other third-party". However, these do not apply to development activity that takes place before the release and they do not include copyright obligations.

With respect to the copyright-related meta-obligations under Art. 53 of the AI Act, extraterritorial application relies on a supporting Recital 106 that demands compliance with EU law on copyright and related rights "regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI models take place." This recital arguably goes beyond the legal provision it supports, potentially dislodging the territoriality principle of copyright law,[4] under which the provisions of EU copyright law do not apply outside its jurisdiction.

Furthermore, the entities carrying out copyright-relevant acts such as TDM-type reproductions, may not be model providers. That means their activities, such as those by Common Crawl (for web scraping) or LAION (for dataset preparation), will not fall under the GPAI chapter of the AI Act at all.

The European Commission is side-stepping the issue with the AI Office's voluntary Code of Practice under the instruction of Art. 56(1) of the AI Act. The extraterritorial effect of the (draft) Code's provision is indirectly obtained by wording that it applies "to all phases of the development of a general-purpose model, including data collection, training, testing and placing on the market" (Measure 2.1: Draw up and implement an internal copyright policy, second draft published 19 December 2024, Rules related to Copyright, AI Act Art. 53(1)(c)). Consequently, the life cycle approach of the Code of Practice will enable providers to demonstrate compliance with the AI Act, suggesting a complex form of voluntary extraterritoriality.

The introduction of value chain and life cycle concepts, combined with a mix of private and public law enforcement is new to the copyright sphere and needs to be thought through carefully. While the AI Act currently does not envisage private enforcement (e.g. a claim for damages from copyright and related rights owners), it may be fruitful to explore analogies with competition law where findings of anti-competitive behaviour may lead to private action for damages.

6. The fair remuneration of authors and performers

Finally, a market is already developing for licensing of copyrighted works and other protected subject-matter, particularly to provide high-quality datasets for training generative AI models and systems (as demonstrated by recent examples of licensing partnerships between AI operators and press publishers, news outlets or images databases producers) and will continue to develop. Therefore, the question of a fair remuneration of authors and performers in compliance with the fundamental principle laid down by Art. 18 of the CDSM Directive needs to be addressed. The following principles should in our view apply to ensure that authors and performers are associated with any exploitation of their works and performances in generative AI operation:

- Art. 18 of the CDSM Directive mandates, as a general principle, that authors and performers receive an appropriate and proportionate remuneration for acts of exploitation of their works and performances in all relevant stages of operation of generative AI models and systems (from training to post-training commercial exploitation of generative AI models, as well as exploitation of generated content similar to their works or performances).

- When their works or performances are part of a collection of works that is specifically licensed to a generative AI model provider as a training dataset, the producer of such a collection, database or news publications needs to ensure an appropriate and proportionate remuneration to authors and performers of content included in the licensed dataset.

- When, after having opted out from the application of the TDM exception, under the conditions laid down by Art. 4 CDSM Directive and Art. 53(1)(d) of the AI Act, rightholders enter into licensing agreements to authorise TDM of works and other protected subject matter by generative AI model providers, some appropriate and proportionate remuneration should be provided to authors and performers when they have transferred or licensed their rights to such rightholder. Since remuneration in such a case of training on massive numbers of works and performances might be rather minimal for authors and performers or difficult

[4] João Pedro Quintais, The AI Act, Copyright and extraterritoriality, Kluwer Copyright Blog, 28 November 2024.

to determine, it would be useful to investigate and identify the legal options left to Member States or adopted at the EU level to organise some other forms of appropriate compensation (such as a residual remuneration right or collective remuneration models existing in several Member States, or, beyond the copyright regime, other compensation mechanisms such as a financial contribution to cultural funds/activities or to the impacted creative sectors).

Disclaimer. *ECS member Prof. Alexander Peukert is currently chairing the sub-working group on the copyright-related provisions of the EU General Purpose AI Code of Practice under the AI Act. He did not participate in the drafting of this Opinion and takes no position on its contents.*

jipitec

Journal of
Intellectual Property,
Information Technology,
and Electronic Commerce
Law

www.jipitec.eu