Searching for Harmonised Rules: Understanding the Paradigms, Provisions, and Pressing Issues in the Final EU AI Act

Hannah Ruschemeier & Jascha Bareis

Abstract

This analysis provides an overview of the enactment of the final European regulation about harmonised rules on artificial intelligence (AI Act). The AI Act establishes the first legally binding horizontal regulation on AI. The paper follows an interdisciplinary approach in combining legal scrutiny with political analysis in order to clearly define and explain the rationale, overall structure, and the shortcomings of the provisions. We understand the crafting of the AI Act as a reaction to the growing centralisation and power of non-European platforms in developing and providing AI systems, and the EU's geopolitical and normative aspirations to shape the adoption of this technology. Overall, this analysis seeks to familiarise researchers from other disciplines (from tech to policy) with the complex regulatory structure and logic of the AI Act. The analysis is structured into three major parts: first, analysing the regulatory necessity in introducing a coercive regulatory framework; second, presenting the Act's regulatory concept with its fundamental decisions, core provisions, and risk typology; and, lastly, critically analysing the shortcomings, tensions, and watered-down assessments of the Act.

1. Regulating AI: an introduction[1]

The enactment of the Regulation (2024/1689) about harmonised rules on artificial intelligence (hereafter, the AI Act), adopted on May 21, 2024 by the Council of the 27 EU member states, establishes the first legally binding horizontal act on artificial intelligence (AI). The adoption of enforceable and binding legal requirements relating to the regulatory subject matter of AI marks a milestone in the diverse development of normative require-

1 Many of the legal aspects were first developed in by Ruschemeier (2023).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ments for this nascent technology. Different institutions, and regulatory levels and subjects are involved in the discussion on normative requirements for AI. To date, no country has enacted a comprehensive legal framework for AI following a horizontal approach, and no international treaty providing uniform international guidelines is currently in force.[2] The international regulation of AI is more of a patchwork than a jigsaw puzzle, due to the different approaches of different states, associations of states, non-governmental organisations (NGOs), and other institutions, if only due to the variety of different competences (Ruschemeier, 2023b).

At first glance, AI is not an unusual subject for regulation: legal regulation in particular has always dealt with new technological developments, uncertainties, or global impacts, as exemplified by environmental and technology law. However, there is a growing international consensus that existing rules at different levels are insufficient for the effective regulation of AI. The reasons for this are manifold and lie in the socio-technical implications of AI, the wide individual, systemic, and residual risks that AI systems can embody, the power centralisation around a few developers and providers, and its ever-evolving technical specificities. Current legal and policy initiatives are faced with the difficulties of keeping pace with these challenges. From a regulatory and societal perspective, the dangers of AI systems grow in line with their use, as greater adoption can impact such protected interests as fundamental rights, democratic processes, inclusion, or public safety. These affected legal interests are not new and are not only threatened by AI applications. However, due to AI's growing pervasiveness in everyday spheres of life in the social, legislative, military, health, and intimate domains, regulators must be able to carefully weigh the risks and potentials.

Given this continuity of technological development, AI is not the *new* disruptive force befalling society suggested by certain private and public narratives (Bareis and Katzenbach, 2022). Rather, its uptake depicts a growing societal leaning on algorithmic automation, continuously reshaping human relationships, with new forms of intimacies (e.g., recommender systems in dating apps), social orders (e.g., the power of Big Tech in providing and controlling digital infrastructure), and knowledge authorities (e.g.,

2 Other countries, such as China or the US, have also forwarded AI regulatory initiatives, including the 2023 Chinese "Interim measures of the management of generative artificial intelligence services" or the 2023 US executive order on "Safe, secure and trustworthy AI". However, these interventions address only selective areas, and thus do not have the scope and depth of the horizontal and comprehensive AI Act.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

societal trust in large language model (LLM) chatbots, such as ChatGPT, to provide *knowledge*). Unsurprisingly, such a cross-cutting technology as AI impact various areas of law, including product safety law, consumer protection law, copyright law, data protection law, protection of fundamental rights, private liability law, criminal attribution issues, and labour law. Thus, AI is by no means being used in a legal vacuum that now urgently requires new, detailed regulation in every area. For example, the Digital Services Act (DSA) (Regulation 2022/2065) does not explicitly mention AI, but aims to create a "safe and trustworthy" online environment, which is threatened by the way digital platforms operate (Art. 1 DSA). This includes the use of AI to display and moderate content (see, for example, the requirements for recommender system transparency in Art. 27 DSA).

The regulation of AI takes different forms: traditional legal regulation can define preventive prohibitions, repressive sanctions, or requirements to act. It can apply existing regulations or create new ones; early ethical proposals can relate to moral requirements, which can, however, become the basis for legal regulation; technical requirements, such as standardisation norms, often create de facto obligations (Veale, Matus and Gorwa, 2023). Consequently, the need for new legislation must be carefully assessed and, if laid open, regulatory gaps should be filled to meet regulatory objectives. For example, the GDPR (Regulation 2016/679)[3] is reaching its limits in terms of the regulation of data-driven technologies, such as predictive analytics or generative AI, since the regulatory object is the single data processing of data belonging to an identifiable data subject.

Through this chapter, we seek to familiarise researchers from other disciplines with the regulatory structure and key requirements of the AI Act, and to critically reflect and analyse the Regulation's fundamental decisions through our interdisciplinary approach. Our conceptual take to the AI Act combines sociological and political analysis with the legal scrutiny of the provisions, thus making the analysis fruitful to legal, policy, social, and technology scholars. To meet these aims, the analysis is structured into three parts:

- First, *Regulatory necessity* introduces the Act's inception. We open our analysis of the AI Act in recognition of the rise and perpetuation of larger power structures, attested to by the pervasive roll-out of AI through ecosystems of platforms and clouds controlled by a few international

3 For more information about the GDPR, see Chapter 14 'EU data protection law in action: introducing the GDPR' by Julia Krämer.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Big Tech companies (van der Vlist, Helmond and Ferrari, 2024). Given the global influence of US and Chinese tech companies in global AI development, we underpin our legal analysis with a short depiction of the EU's geopolitical and normative aspirations, which influenced the overall crafting of the AI Act.

- This larger political embedding of the AI Act leads us to the second part, *Regulatory concept of the AI Act.* This part presents the Regulation's core provisions in addressing the different scopes of application. Here, we also dive into the various risk-categorisations and their subsequent regulatory prescriptions, reaching from no restrictions to forbidden practices for market deployment.
- Finally, *Critical analysis* reflects upon the shortcomings, tensions, and watered-down assessments of the AI Act. We argue that these largely stem from the Act's overall conflictual aspiration to combine fundamental rights protection with a risk-regulatory assessment of harms for products, while simultaneously aiming towards a harmonised and internationally competitive and resilient common AI market.

2. Part I: Regulatory necessity

2.1 Regulating AI is regulating power

Common regulatory objectives for AI are often described as "fairness", "transparency", "explainability", "trustworthiness", "safety", "protection of fundamental rights", "sustainability" and "fostering innovation" (Hacker, 2018; Malgieri and Pasquale, 2024; Goh and Vinuesa, 2021; Stahl et al, 2022). However, further to these desirable and laudable goals, there is a further rationale to create new regulatory requirements for AI. The regulation of AI is the regulation of societal power, and thus a truly constitutional and public interest issue, because the rule of law serves to simultaneously legitimise and limit power (for a general overview, see Summers, 1998). Power dimensions in AI applications are manifold: the centralisation of infrastructure, AI models, and data appropriation in the hands of a few Big Tech players; the "black boxing" of AI systems, where people cannot understand, explain, or comprehend the path to a system's output which decides upon them; or the individual and highly systemic dangers that AI systems can cause without providers taking accountability (see also Guijarro Santos, 2023).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Firstly, the key players in AI technologies, who have urged state to take action and provoked the legal policy debate on AI regulation in the first place, are large global technology companies. The development and application of AI is not limited to the private sector: open source initiatives, NGOs, government institutions, and scientific research also play key roles in the development and dissemination of AI applications. However, the technologies dominating the market and discussed in public discourse are primarily those developed and deployed by private sector actors and are embedded in their platforms. Therefore, it would be vital to make transparent the purpose of the economic profit of these actors, who all too often foster a deregulatory agenda.

Despite the privatisation of AI, it is by no means impossible that many people (can) benefit from it, or that the technology could be used for the greater common good. However, pervasive power structures are created when states and users are forced to rely on private companies for the use of AI. The current structural dependency on Big Tech players for infrastructure provision, model development, maintenance, and auditing is creating lock-in effects. Indeed, as stated by Whittaker (2021, p. 35): "These companies control the tooling, development environments, languages, and software that define the AI research process – they make the water in which AI research swims". With the recent development towards foundation models – i.e., very large pre-trained models on which such popular applications as ChatGPT or Midjourney run – the centralisation of AI is increasing further (Burkhardt and Rieder, 2024; van der Vlist, Helmond and Ferrari, 2024). Big Tech use their platforms as bottlenecks in AI development and provision, assuming a gatekeeper position to certain apps. For example, the COVID-19 tracing apps could only be successfully launched through the Google and Apple app stores (Bock et al, 2020), or ChatGPT can only be used in the Open AI or Microsoft Azure ecosystems, following Microsoft's investments into Open AI. While the functioning of the tracking apps does not directly fall under the definition of AI under the AI Act (Art. 3 (1)), the risks to digital sovereignty through a heavy reliance on private digital infrastructures is transferable and growing with AI ecosystems. Currently, there are already discussions about the use and implementation of LLMs in the public sector. For instance, Microsoft announced its intension to implement generative AI in many Office 365 applications, a software which is heavily used by public authorities despite its non-compliance with the General Data Protection Regulation (GDPR) – indeed, it is generally perceived as too big to not use (Ruschemeier, no date; EDPS, 2024).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Secondly, the power dimension is present with these Big Tech companies executing data appropriation of users, essentially an *assetisation* of citizens with the lure of free-to-use services, a business model also called *service-for-profile* (Elmer, 2003; Mager, Norocel, and Rogers, 2023). Alphabet (Google's parent company) collects data on the behaviour of users of its various services, allowing it to build detailed profiles and predictions of consumer preferences. These sensitive data can then be sold to third parties and advertisers (Ridgway, 2023). Meta (formerly Facebook, Inc.) personalises its algorithm to display content and collects data to an extent to which users are generally unaware (Arias-Cabarcos, Khalili and Strufe, 2023). Hence, these private players exploit extremely large user bases to fuel and train their AI models to offer service for *free*. This endows them with considerable predictive power, having insights in the most intimate, sensitive social and political spheres – which is historically unprecedented for the private sector – ranging from highly sensitive information, such as creditworthiness, to sexual orientation or health status (Mühlhoff, 2023; Ruschemeier, 2024a; Mühlhoff and Ruschemeier, 2024a; Ruschemeier, no date). Often, consent is not even requested: Open AI's ChatGPT only works as well as it does because it was developed by trawling almost the entire internet for publicly available information on which to train its model (Ruschemeier, 2023c). The European Court of Justice (ECJ) recently ruled that Facebook's business model – namely, financing through individualised advertising – does not in itself constitute a legitimate interest in the mass processing of personal data (Meta v Bundeskartellamt, 2023).

Thirdly, the ubiquity of these digital processes and the proliferation of AI also carry epistemological implications: how are decisions that govern over people procedurally made? How can they be contested? How is knowledge generated and given authority (Ruschemeier, no date; Hong, 2020)? Such production of the perception of knowledge is a pervasive exercise of epistemic power, with users granting excessive trust in machine-based decision suggestions (Ruschemeier, 2023d; Hondrich and Ruschemeier, 2023). Empirical studies have shown that users do so even if they know nothing about the underlying training data or, perhaps more gravely, if they are aware that they are confronted with a biased AI (Krügel, Ostermaier and Uhl, 2022). LLMs provide eloquent sounding answers, and have been pervasively hyped as knowledge models (indeed, ChatGPT's slogan reads: "Ask me anything!"), intentionally leaving the functionality of the probabilistic models working with tokens, and not hermeneutically with meaning, in the dark. (Bareis, 2024). Probabilistic models process data based

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

on statistical likelihood. These models have no *understanding* of neither the prompts nor outputs they generate, and can thus generate nonsensical content (termed "hallucinations") (Metz, 2023). Moreover, this publicly produced misconception leads to a crisis of knowledge, as synthetically generated content is currently flooding the internet and is being indexed as "knowledge" by search engines. This provokes an epistemological crisis. As argued elsewhere, this could lead to our inability to identify trustworthy information even when we find it (Bareis, 2023c).

The business models, structural dependencies, socio-technical interactions, and, not least, the pervasiveness of scale described above mean that previous regulatory approaches are no longer effective in all cases. Where power is involved, the potential for social improvement is as obvious as the risk of abuse. According to the precautionary principle, certain particularly risky products and processes may be preventatively subject to legal regulation if they threaten important legal and public interests (Sandin, 1999). As with any transformative technology, it has often been argued that the challenge with AI is that some impacts are difficult, impossible, or even unknowable to foresee. However, with these pervasive societal effects of AI already present (and, indeed, known) this argument should not exempt politics from accountability. The law-lagging moment with AI is politically produced and a well-studied case (Doezema and Frahm, 2023). The precautionary principle gives politics the mandate to intervene in the name of public interest. Law must not socially be lagging, but leading.

2.2 EU taking a stance in the geopolitical AI arena

In recent years, a number of initiatives have emerged globally to define values and principles for the ethical development and use of AI. A multitude of international and supranational bodies, such as the Organisation for Economic Co-operation and Development (OECD, 2019), have proposed principles for standards of "trustworthy" AI. Likewise, the United Nations (UN, 2023) published the "Governing AI for humanity" report in late 2023. These reports are mostly based on abstract ethical principles useful for providing orientation on the safeguards, rights, and principles deemed to be protected in the international realm. Still, non-binding recommendations, policy papers, soft law, or ethical principles are often criticised for being ineffective because they are non-binding and therefore unenforceable (Mittelstadt, 2019). So far, the private sector, dominated by US Big

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Tech companies, has largely ignored all proposals and lobbied aggressively against regulation, which also became very visible in the final phase of the European AI Act legislative process (Bareis, 2023a; Ruschemeier and Mühlhoff, 2023). Hence, ethical principles give normative orientation, but can quickly be watered down and often lack teeth.

The strivings of the European AI Act are embedded in a global AI race, with nations and their companies identifying AI as a core present and future enabler technology. Moreover, the EU envisions that AI shall transform the common internal market into an international competitive player, competing over global market shares and innovation (Krarup and Horst, 2023; Paul, 2023; Smuha, 2021). States approach AI not as a mere technology, but also as a strategic asset in the geopolitical positioning against rivalling economic (and military) actors, such as China, or the US and their Big Tech companies (Bächle and Bareis, 2022; Bächle and Bareis, 2025; Kello, 2017). When discussing the formation of the AI Act, it should be kept in mind that its formation falls into a global paradigm where tech policy has been highlighted by states as a pivotal realm to advance and harness sovereignty and a claim to first mover clout (Broeders, Cristiano and Kaminska, 2023).

The "European way" of tech-policy is subsumed by the European Commission (EC) as a necessity for achieving its own tech sovereignty. The Council of the European Union defines this strategic autonomy as the "ability to act autonomously when and where necessary and with partners whenever possible" (Mogherini, Timmermans, and Domecq, 2016, p. 4). EC president Ursula von der Leyen referred to this paradigm of strategic autonomy through stressing that: "Tech sovereignty describes the capability that Europe must have to make its own choices, based on its own values, respecting its own rules" (European Commission, 2020a). These statements echo endeavours of a de-risking strategy, essentially acknowledging the fragile balancing act of protecting Europe's AI market without retreating into a paradigm of protectionism in questions of economic trade, sensitive technology exchange, and military development (Rodríguez Codesal, 2024). In this context, the "European Chips Act" (European Commission, 2023) is situated with the proclaimed aim to support Europe's AI infrastructure, subsidising the European semiconductor industry and encouraging companies to invest so as to decrease dependencies on Taiwan, the US, Japan, or China.

For the EU especially, which is a supranational entity unifying 27 sovereign member states under the principle of subsidiarity (Art. 5 (3)

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

TEU), the harmonisation of standards and policy is a complex and lengthy process. The significant efforts and prioritisation of the EC, which hails itself as the first "geopolitical Commission" (von der Leyen, 2019) to tackle the AI Act, can also be understood as a reaction to the tedious EU constitutional integration process that was substantially gridlocked. The then-curtailed treaty of Lisbon was marked by a multitude of obstacles in the ratification process in the early 2000s, complicating further constitutional integration from an inward union perspective. Hence, on constitutional, military, and geopolitical stances, the EU's power is limited in finding joint positions and reacting quickly and effectively. It is rather by the power of "commanding the weight of the internal market" that the EU can execute "regulatory power in the international domain" (Broeders, Cristiano and Kaminska, 2023, p. 1265). In market policy questions, European integration is, as historically grown from its foundation of a coal and steel community (ECSC), the deepest, with clear delegated roles and coercive power for EU institutions. It is this context where the DSA, DMA, and AI Act are embedded, attempting to strengthen the unity of the European member states with a common AI rule book in order to meet a geopolitical competitive environment. Whether the so-called "Brussels effect" – that is, the hope that EU's AI regulation will have the desired impact on the global diffusion and standard-setting beyond its own borders (Siegmann and Anderljung, 2022) – remains to be seen.

2.3 Coming into being: from ethical guidelines to legal regulation

Next to these imperatives of an *outward* international competitive situation for AI market shares and the political aim for *inwards* legal harmonisation against fragmented national policy, the EU sees itself as a proponent of safeguarding consumer protection within the single market and the fundamental rights of individuals.

This very normative pillar of the EU's self-identity is legally enshrined with the EU charter of Fundamental Rights. Additionally, the ethical alignment is evidenced by the AI's framework of "human-centric ethics", "fundamental rights impact assessment" (see Section 3, below) and the fashion-

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ing of *trustworthiness* throughout the European AI documents.[4] Although the ethical considerations are non-binding and not passed via a democratic process, they have influenced the roadmap of AI legislation. Here, the role of high-level ethics groups in sketching the path for AI legislation is particularly noteworthy. The principle setting by expert groups is an important trajectory in understanding how the coercive AI Act came into being. The European Group on Ethics in Science and New Technologies (EGE) published a report (EGE, 2018) on "Artificial intelligence, robotics and 'Autonomous Systems'", calling "for the launch of a process that would pave the way towards a common, internationally recognised ethical and legal framework for the design, production, use and governance of artificial intelligence (...)". In a clearly prescriptive call, the EGE "urges the European Union to place itself at the vanguard of such a process and calls upon the European Commission to launch and support its implementation" (2018). Frahm and Schiølin (2023) understood these early AI ethics reports by convening expert committees as instruments of socio-technical sense-making and ordering of the EU's position on AI, as well as the rise of the principle of "European technological sovereignty", which the EUC henceforth embraced. The subsequent adherence to AI ethical principles and values subsumed under the notion of a "trustworthy AI ecosystem" were adopted by the High-Level Expert Group (HLEG) on AI in 2019 (AI HLEG, 2019) and normatively underpinned the formation of the AI Act.

It is not only in the field of AI that legally binding requirements and ethical proposals influence each other as different dimensions of normativity: ethical standards are based on the legal system, while the law translates ethics into enforceable requirements (Ruschemeier and Mühlhoff, 2023). For example, the HLEG's "Ethical guidelines for trustworthy AI" advance three central criteria that all AI systems should fulfil: legality, ethical compliance, and robustness.[5] At the national level, the German Data Ethics Commission proposed a risk-adaptive regulatory approach in its report (Datenethikkommission, 2019) on algorithmic systems, which is now being implemented in a similar form at the European level.

4 For example, the 2020 Assessment List for Trustworthy AI (ALTAI) (European Commission, 2020b) or the 2020 white paper issued by the EUC (European Commission, 2020c).

5 However, trust is not actually defined in any of the EU documents, which neither reflect whether "trust" is actually the correct term or a conceptual misfit in this context (Bareis, 2024).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

In the AI Act, the focus now lies on the protection of health, safety, and fundamental rights, while there are almost no references to ethical guidelines left in the binding part of the Act. Indeed, only Art. 60(3) requires that the testing of high-risk systems in real world conditions should be made without prejudice to any ethical review required by Union or national law, which is a special provision for supporting innovation via regulatory sandboxes. The second mention of ethical considerations can be found in Art. 95, which outlines codes of conducts with specific, but voluntary, requirements. These voluntary guidelines can include applicable elements provided for in Union ethical pillars in order to establish "trustworthy AI" (Art. 95(2) AI Act). Beyond the explicit mentioning of ethical guidelines, the AI Act no longer includes specific ethical considerations, instead remaining silent on value aspects. There remain many open normative questions that wait for instantiation and concretisation. For example, when are biases in AI systems problematic (following which understanding of anti-discrimination?), or what makes an AI system really "fair" (given the myriad contradictory fairness principles) or "trustworthy" (can technology be trustworthy at all, or just reliable?) (see discussions in Bareis, 2024; Laux, Wachter and Mittelstadt, 2023; Orwat et al, 2024; Wong, 2020)?

Despite the provisions, however, the recitals explicitly point out the objective of promoting the European human-centric approach to AI and stress the Union's goal to be a global leader in the development of "secure, trustworthy and ethical AI", as stated by the European Council. It ensures the protection of ethical principles, as specifically requested by the European Parliament (Recital 8). Recital 27 refers and explains the ethical guidelines for trustworthy AI developed by the HLEG (human agency and oversight; technical robustness and safety; privacy; data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; and accountability). The recital states that: "Without prejudice to the legally binding requirements of this Regulation and any other applicable Union law, those guidelines contribute to the design of coherent, trustworthy and human-centric AI, in line with the Charter and with the values on which the Union is founded". However, it should be noted that these recitals do not form part of the Regulation's bind text, but are rather used for interpretation and guidance. Some obligations for high-risk systems *can* be linked to the ethical considerations, such as the provisions on human oversight or data governance. However, these will ultimately be specified by the private standardisation organisations (for a more detailed discussion, see Sections part III, 3, 3.1). Recital 28 also

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

refers primarily to codes of conduct, although, again, these can be used on a voluntary basis. Despite being explicitly mentioned, the impact of the ethical guidelines as an interpretative guide is rather limited. It is striking how little of the ethical pillars, initially greatly stressed by the HLEG, is left in the final AI Act and incorporated into binding law.

3. Part II: Regulatory concept of the AI Act

The following section introduces the regulatory concept of the AI Act by explaining its regulatory structure (3.1), the scope of application (3.2), the important categories of forbidden and high-risk systems (3.3), and the oversight and governance structure (3.4).

The AI Act constitutes a legislative act of the EU in the form of the Regulation (Art. 288(2) TFEU). From this, it follows that the normative provisions are entirely binding and directly applicable in all Member States. EU regulations take precedence over national laws in case of conflict. Most aspects of the AI Act are fully harmonised, but there are opening clauses for the Member States, such as on the prohibition of certain systems under national law.

3.1 Regulatory structure

The general goal of the AI Act is to set harmonised rules for the development, use, and marketisation of AI in the European single market. Its regulatory aim is described as:

> ... to promote the uptake of human centric and trustworthy artificial intelligence (AI) while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (the "Charter"), including democracy, the rule of law and environmental protection, to protect against the harmful effects of AI systems in the Union, and to support innovation. This Regulation ensures the free movement, cross-border, of AI-based goods and services, thus preventing Member States from imposing restrictions on the development, marketing and use of AI systems, unless explicitly authorised by this Regulation. (Recital 1, AI Act)

The explicit reference to health and safety shows how the Act is mostly a product safety regime with additional references to fundamental rights due to its heavy references to the harmonised framework of product safety law

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

in the EU, especially the New Legislative Framework[6] (NLF) (European Commission, 2008). Consequently, the AI Act is part of a larger package to further regulate product safety for AI and other products, such as the new Machine Regulation (Council of the EU, 2023) or the Toys Directive (Directive 2009/48/EC)).

Furthermore, the AI Act is part of the Commission's digital strategy (European Commission, 2024), which includes other important legislative acts, such as the DSA and the Digital Markets Act (DMA). While its legislation was mostly parallel to the discussion and enactment of the DSA and DMA, the latter two regulations are fundamentally different. The DSA and DMA aim to regulate such intermediaries as social media platforms and search engines in the digital sphere, and create special obligations for very large online platforms and search engines, such as Meta, Instagram, TikTok, Bing, and Google (see Art. 33 et seq. and Art. 3 DMA addressing "gatekeepers"). The AI Act, on the other hand, does not primarily address Big Tech players, but rather focuses on public sector applications, (cf.[7] Annex III). This raises the question of whether the Regulation sufficiently addresses the power aspects of private actors. Additionally, the AI Act does not specifically consider the position of the actors, unlike the regulatory categories of "very large online platforms" (DSA) or "gatekeepers" (DMA), but regulates regarding contexts of use, such as AI systems for public services or law enforcement. The particular relationship of the AI Act towards other legal acts on the Union level has yet to be fully clarified, however, it is important to note that the Act will not replace the GDPR, but will have significant overlaps when AI systems process personal data. Art. 2(7) states that Union law on the protection of personal data, privacy, and the confidentiality of communications applies to personal data processed in connection with the rights and obligations laid down in the AI Act, which shall not affect the GDPR.

The AI Act follows a risk-based regulatory approach and the creation of a horizontal (as opposed to sectoral) legal framework. From this, AI systems are to be classified into four risk categories: unacceptable (Art. 5), high (Art. 6, 7, Annex III), low (Art. 50), and systemic (Art. 52) for the category

6 NLF refers to a revision and harmonisation of technical standards for the internal union market. It addresses market surveillance, accreditation, conformity assessments, and labelling (e.g., CE marking). After more than 20 years, the "New approach" was revised and updated, with the so-called NLF adopted in 2008. It came into force in January 2010 (European Commission, 2008).

7 cf. stemming from Latin *confer*, meaning "compare".

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

of general-purpose AI systems. Depending on the risk classification, different obligations for providers and deployers apply. On the one hand, very low risk systems, such as email spam filters, are not subject to regulation. On the other, unacceptable risk systems, such as manipulative AI, social scoring, and remote biometric identification are banned, the latter of which is subject to broad exemptions for judicial and law enforcement authorities (cf. Art. 5 AI Act). Practically speaking, high-risk systems represent the most important category, since the majority of the Act's provisions address them. The Commission assumes that 5–15% of the AI systems on the market will fall under the high-risk category (European Commission, 2021).

The AI Act has 13 chapters and follows the classical formation of a European regulation starting with general provisions (I), followed by the prohibited practices (II), standards for high-risk systems (III), transparency obligations (IV), general-purpose models (V), measures in support of innovation (VI), governance (VII), requirements for the EU database for high-risk systems (VIII), post market monitoring and market surveillance (IX), codes of conduct (X), delegation of power (XI), confidentiality and penalties (XII) and, lastly, final provisions (XIII).

3.2 Scope of application

The scope of application of the AI Act is divided into the territorial and material scope of application, following the requirements from article 2 of the Act.

3.2.1 Material scope of application

Firstly, the AI Act's material scope must apply. The material scope describes the subject matter of regulation, such as the regulatory objects (AI systems and models) and actions (putting an AI system on the market). It can be limited by exceptions. The material scope of application of the AI Act includes placing AI systems on the market or putting them in service. While AI as a regulatory object is disputed, the definition of an AI system in Art. 3(1) requires levels of autonomy and outputs that influence physical or digital environments (see the critical discussion of the term AI system in section C I). As such, this rather broad definition includes many AI systems based on machine learning (ML), or simpler algorithmic decision-making systems (ADMs).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

3.2.2 High-risk classification as the relevant regulatory definition

Considering the Act's overall structure, most of its provisions address high-risk systems. Perhaps counterintuitively, the relevant regulatory definition for the material scope is the high-risk classification (cf. Art. 6, 7, Annex III AI Act) or prohibition in Art. 5 and the general-purpose qualification (Art. 51) instead of the actual definition of the AI system. According to the Act, placing an AI system on the market involves first of all making the system or general-purpose AI model available on the Union market (Art. 3(9)). Here, a system is put into service for customers when it is supplied for first use directly to the deployer or for its own use in the Union for its intended purpose (Art. 3(11)).

AI systems can be classified as high-risk under Art. 6 in two ways: first, when they are products or safety components of products covered by the Union harmonisation legislation (detailed in Annex I), and, second, due to their relevance for possibly infringing on fundamental rights regarding the context of use (covered by Annex III). The reference to Union harmonisation legislation in the area of product safety law in Annex I itself is subdivided into Sections A and B. Section A refers to the NLF, while Section B refers mostly to vehicle and traffic provisions, such as the regulation on the approval and market surveillance of two- or three-wheel vehicles and quadricycles (Annex I B(14)). These harmonised rules are not part of the NLF, but part of the older Union legislation which follows the concept of detailed harmonisation, and can thus not be easily synchronised with the new AI Act. Most of the requirements of the AI Act do not apply to products under the old regulatory regime, as the old regime and the NLF follow fundamentally different approaches and metrics for product safety regulation – e.g., the old concepts established only government standards and the review of requirements by government agencies. This creates friction with the requirements of the AI Act, which is largely based on newly implemented standards established through private standardisation organisations, internal conformity assessment procedures, or procedures of a private notifying body (cf. Art. 43 et seq.). Art. 2(2) thus states that, for these systems under the old regime, only Art. 6(1), Art. 102–109, and Art. 112 apply. Art. 6 lays down the classification of high risk systems, while Art. 102 et seq. are final provisions amending other regulations and directives. Art. 57 sets the requirement to establish regulatory sandboxes for the testing of AI systems and applies only insofar as the requirements for

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

high-risk AI systems under this Regulation have been integrated in that Union harmonisation legislation.

3.2.3 Exceptions in the material scope

There are several exceptions in the material scope of the AI Act applications. Art. 2 names some of them: AI systems and models that are specifically developed and put into service for the sole purpose of scientific research and development are not covered by the regulation. In the EU rationale, this is because the aim of the AI Act is to foster innovation and support research. Recital 25 explicitly states that the AI Act shall not affect research or scientific freedom. The prerequisite for this exception is that the models are specifically developed and used for the sole purpose of research, which naturally leaves room for interpretation, given that many commercial start-ups in the AI sector stem from, or are connected to, university research. Furthermore, private funding for AI university research by Big Tech is especially prevalent in the Anglo-American context, but also increasingly in Europe, with Meta, for example, financing an AI ethics centre at the technical university of Munich (Kreiß, 2019). Moreover, training data for scientific research is often taken from the public rather than from research, such as with ChatGPT or other LLMs being trained on online content. As it stands, the private research departments of the Big Tech companies that aim at developing and improving products may not fall under the definition of solely research purposes, but how the AI Act applies in detail here remains to be seen in practice.

Beyond science, the AI Act does not apply to product-oriented research, testing and development activity regarding AI systems or models prior to those systems, and models being put into service or placed on the market (Art. 2(8)), except for testing under real-world conditions as part of the regulatory sandboxes of Chapter VI. Regulatory sandboxes are a testing environment for AI systems, such as finance apps and other applications, that can, for instance, affect customers. The AI Act defines regulatory sandboxes as controlled frameworks established by competent authorities which offer (prospective) providers of AI systems the possibility to develop, train, validate, and test innovative AI systems, where appropriate in real-world conditions, pursuant to a sandbox plan for a limited time under regulatory supervision (Art. 3(55)) (Ruschemeier, 2024b). Consequently, the training of AI systems and models does not fall within the scope of the AI Act. Additionally, the Act does not apply to obligations of deployers who are

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

natural persons (humans, not legal entities) using AI systems in the course of a purely personal non-professional activity, since these are understood as typically low risk, and thus not subject to regulation.

The AI Act excludes AI systems that are released under free and open-source licences unless they are placed on the market or put into service as high-risk AI systems or those which fall under Art. 5 or 50. Art. 5 regulates the forbidden AI systems that pose unacceptable risks and are therefore prohibited, while Art. 50 lays down transparency obligations for providers and users of certain AI systems and general-purpose AI models. The provisions on the latter have been implemented very late in the legislative process as a reaction to the rising popularity of generative models running chatbots, such as ChatGPT. Art. 3(63) defines a general-purpose AI system 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.

The exception for open-source systems is rightfully limited to those that are not prohibited or general-purpose AI, deepfakes, and those interacting with natural persons (Art. 50). Nevertheless, excluding open-source models from the legislation should not obscure the fact that these models can also harbour risks, e.g., when used in Annex III contexts (Mühlhoff and Ruschemeier, 2024d).

Finally, it is worth noting that the Act entirely excludes the military application of AI. This is a striking omission given the dual-use applicability of civil/military AI innovation and the research capabilities and use of AI in the military sector – as seen with the unhalted development of autonomous weapon systems (Bhuta, Beck and Liu, 2016). Especially in the US, state agencies cooperate with major technology corporations contributing to national military and intelligence imperatives. This is also the case with some European states (Germany, France, Spain), who cooperate with the private sector and heavily invest into military AI with the development of the European Future Combat Air System (FCAS), aiming to develop "combat clouds" with the implementation of communication hubs or real-time data analytics for synchronising their military forces (see Ernst, forthcoming).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Given the fragile current world political situation, military supremacy is trending high on many national geostrategic security agendas. The global regulatory debate on autonomous systems is being held at the UN Convention on Certain Conventional Weapons (CCW), where the compliance to International Humanitarian Law applies, but is currently gridlocked (Bächle and Bareis, 2022). EU Member States seemingly do not want to relinquish control of military AI use to the EU, thus leaving a significant loophole for unchecked AI development and use.[8]

3.3 Personal scope of application

The AI Act addresses different entities in the AI lifecycle (Art. 2(1)). Firstly, it applies to providers of AI systems that are placing them on the market or putting them into service (Art. 2(1a)). Secondly, it addresses deployers, providers, importers and distributors, product manufacturers, authorised representatives of providers, and affected persons (Art.22 (1) a–g). Thirdly, obligations also extend to importers and distributors (Art. 23–27) in a manner akin to the product safety regime, aiming to prevent dangerous products manufactured outside the EU from entering its market. Nonetheless, the primary actor upon whom these obligations are imposed is the provider (Edwards, 2022c).

3.4 Territorial scope of application

Akin to the GDPR, the AI Act follows the domestic-market principle (Koloßa, 2020), meaning that it applies to placing AI models on the EU market, regardless of whether the providers are established or located within the Union or in a third country (Art, 2(1a)). Furthermore, it is already sufficient that the output of the AI system is used in the Union when providers and deployers of systems are located in a third country for Art. 2(1c) to be

8 A detrimental use of current military AI can be witnessed in the Gaza strip, where the Israel Defense Forces (IDF) are using AI in the military operations in Gaza following Hamas's terrorist attack of 7 October, 2023. Investigations about the "Lavender" and "Habsora" scoring system show how target recommendation of "militant suspects" is automated by the IDF, and air strikes are largely conducted without a human in the loop (Abraham, 2024). This has caused gross human rights violations in the massive bombing of the Gaza strip. The case strikingly shows that AI recommender systems being largely applied in the public domain can also be used for military purposes.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

applicable. Following this, the relevant data (e.g., to train the AI system) can be processed outside the Union, as long as the results of the system are used within the single market. Additionally, the AI Act applies to deployers of AI systems established or registered within the Union (Art. 2(1b). Even if this wording is misleading, the scope of application with regard to users only refers to the spatial boundaries of the 27 Member States (Gless and Janal, 2023, p.30). The establishment refers to the deployers rather than to the AI systems, meaning that the former must be within the Union. Art. 3(4) defines a deployer as a "natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity". As such, this broad definition of the scope is convincing as AI is a digital technology whose impact does not stop at national borders.

4. Forbidden high-risk systems and systemic risks

The AI Act establishes different levels of regulatory measures according to the risk classification of the system. Art. (5) prohibits the use of certain systems (4.1), Art. 6 classifies high-risks systems (4.2), and Art. 50 et seq. establish specific provisions for general-purpose AI systems (4.3).

4.1 Prohibited AI practices

Art. 5 prohibits certain types of AI systems which can be classified into eight categories: 1) subliminal techniques, 2) exploitation of vulnerabilities, 3) social scoring, 4) person-based predictive policing, 5) the creation of facial recognition databases via untargeted scraping, 6) biometric categorisation systems, 7) the emotional recognition systems in the workplace, and 8) real-time biometric identification.

First, the putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness, or purposefully manipulative or deceptive techniques are forbidden. The term "subliminal techniques" is itself problematic, since there is no clear evidence or history of non-valid experiments in this field (Neuwirth, 2023). These techniques should include the objective or effect of materially distorting the behaviour of a person or group of persons by appreciably impairing their ability to make informed decisions, thereby causing them to take decisions they

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

would otherwise not have taken. This refers to the reasonable likelihood to cause that person, another person or group of persons, significant harm (Art. 5(1a). Recital 29 names audio, image, and video stimuli that persons cannot perceive (by being beyond human perception), or other manipulative or deceptive techniques that subvert or impair a person's autonomy, decision-making, or free choice in ways that people are not consciously aware of or, where they are aware of them, can still be deceived or are unable to control or resist them as examples for subliminal techniques. Concrete facilitation could be by machine-brain interfaces or virtual reality as they allow for a higher degree of control of what stimuli are presented to persons, insofar as they may materially distort their behaviour in a significantly harmful manner (Recital 29). Another concrete example of concerns resulting from subliminal and supraliminal messages in the field of cybersecurity are the so-called "social engineering attacks", such as phishing, that refer to means of "manipulating people into performing actions or divulging confidential information" (Neuwirth, 2023).

Secondly, systems that exploit any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability, or specific social or economic situations, with the objective, or effect, of materially distorting their behaviour in a manner that causes (or is reasonably likely to cause) significant harm are prohibited under Art. 5(1b). The "Unfair commercial practices directive" establishes a similar provision (art. 5 UPD; Directive 2005/29/EC). Regarding the AI Act, the specific characteristics exclude other characteristics, such as race, sex, religion, or ethnicity. Smuha et al (2021) suggested expanding these to all of the characteristics protected under EU equality law as laid down in Art. 21 of the EU Charter on Fundamental Rights. It is not yet clear which specific practical examples are included. The specific exploitation of vulnerability leading to a change in behaviour may already be the purchase of an overpriced product or, for example, in-app purchases of video games for children. In general, the secondary use of sensitive data, such as health or other data relating to the specific vulnerabilities for commercial purposes is highly problematic (Mühlhoff and Ruschemeier, 2024c, 2024d). In these cases, however, it is questionable whether, for example, financially disadvantageous purchases fall under the concept of significant harm, which may only be assumed in the area of criminal disproportionality.

Third, systems for social scoring are prohibited under Art. 5(1c). Social scoring systems are used to evaluate or classify natural persons or groups over a certain period of time based on their social behaviour or known, in-

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ferred, or predicted personal or personality characteristics. The social score leads to either or both of the following: (i) the detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts unrelated to the contexts in which the data were originally generated or collected; and (ii) the detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity. The practical reference is China's social scoring system, where camera surveillance, consumer data analytics, and geo-tracking are used to form a disciplining scoring system (Qian et al, 2022). Scoring systems with different characteristics are also used by other countries, such as in the UK's (UK Parliament, 2021) concept of digital identity. During the legislative process, the prohibition was extended to private actors. Risk-scoring practices by private actors are essentially ubiquitous, ranging from the calculation of healthcare insurance premiums to creditworthiness scoring (Citron and Pasquale, 2014). Here too, the regulatory hurdle is again the consequence of these practices, which on the one hand must be proven and on the other unjustified. The associated Recital 31 does not state any use cases or concrete examples.

The fourth prohibition addresses predictive policing techniques related to natural persons in order to assess or predict their risk of committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics. As per Art. 5(1d), 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.

As a reaction to the business practices of Clearview and other facial recognition databases not compliant with the GDPR (Pathak, 2022), but still hard to come by because of the structural enforcement deficits towards malicious actors, Art. 5(1d) prohibits systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage. Clearview and PimmEyes have illegally, and essentially secretly, scraped social media platforms and many other websites for images of faces to build huge databases for the private use of facial recognition. These databases can be used by any individual and by public authorities for a certain fee to identify almost every person whose picture can be found online – indeed, as of 2021, the Clearview database contained 10 billion pictures (Dul, 2022). Accordingly, these business practices aim at abolishing any privacy and personal integrity. Persons

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

can be easily identified with AI-powered facial recognition technology, where uploaded pictures of individuals show results within seconds, including links to the websites from which the pictures were scraped (Hill, 2022; Rezende, 2020).

The sixth prohibition includes the use of AI systems to infer a natural person's emotions in the workplace and educational institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons (Art. 5(f)). Emotion recognition systems are designed to measure, for example, whether content has been understood by students or whether employees are productive and satisfied. The reliability or even effectiveness of emotion recognition systems has yet to be scientifically demonstrated (Heaven, 2020). It is therefore welcome that the AI Act bans these systems, at least in the context of work and training – but their general use remains questionable. Human emotions should not be used for performance reviews, as their scoring depicts a strong risk of abuse (see above with the "Clearview" case).

The seventh prohibition includes the use of biometric categorisation systems that individually categorise natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, and sexual lives or orientation. This prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or the categorising of biometric data in the area of law enforcement (Art. 5(1g)).

Finally, the eighth prohibition includes the use of "real-time" remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement (Art. 5(1h)). The scope of the ban on biometric recognition systems was one of the most debated issues in the legislative process, and is beyond the scope of this paper (see, for example, Edwards, 2022a; Barkane, 2022; Veale and Borgesius, 2021). Biometric surveillance systems carry a high risk of mass surveillance, including those used for social scoring and predictive policing, as discussed above (Wendehorst and Duller, 2021). Art. 5 names a broad number of exceptions of the use of biometric systems in publicly accessible spaces for different objectives of law enforcement, which render the scope of application of the actual prohibition very narrow (Ebers et al, 2021). These exceptions include: (i) the targeted search for specific victims of abduction, trafficking, 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, present, or foreseeable

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

threat of a terrorist attack; and (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 by a custodial sentence or detention order for a maximum period of at least four years. Point (h) of the first subparagraph is without prejudice to Art. 9 of the GDPR for the processing of biometric data for purposes other than law enforcement.

4.2 High-risk systems

Art. 6 concerns the requirements for categorising high-risk systems and is thus a central requirement of the Regulation. The requirements for risk classification are of considerable practical significance, as many AI systems of relevance (will) fall into the category of high-risk systems. The standard is closely linked to the harmonisation provisions listed in Annex I, which largely determine the requirements for risk determination in the context of product safety law in accordance with the AI Act's first paragraph. In the Regulation's structure, Art. 6 follows the second section on prohibited practices of AI, which contains only one provision (Art. 5). The categorisation as a high-risk system under Art. 6 triggers the obligations under Art. 9 et seq., such as the requirements for human oversight (Art. 14) or data governance (Art. 10). The addressees of the AI Act (providers) are the same as those of the new legal framework for product manufacturers (Ruschemeier, forthcoming).

The first approach for classifying AI systems as high risk is established in Art. 6(1) with references to already existing product safety law. To be classified as high risk, the AI system must either be intended to be used as a safety component of a product or is itself a product, as covered by the Union harmonisation legislation listed in Annex I. A safety component of a product is defined in Art. 3(14) as a component of a product or of an AI system which fulfils a safety function for said product or system, or the failure or malfunctioning of which endangers the health and safety of persons or property. For example, an AI system used as a safety component could be an automatic detection of the need for lift maintenance. Additionally, the system as a product itself or as a safety component of a product must be required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of said

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

product pursuant to the Union harmonisation legislation listed in Annex I. Under product safety law, a third-party conformity assessment is required for products with a higher risk, while other products can be self-assessed by the provider. This first variety of high-risk classification is aligned with the system of European product safety law, and is thus not a new regulatory approach under the AI Act.

Nonetheless, the second approach for classifying AI systems as high risk establishes a new assessment of fundamental rights implications. Under Art. 6(2), systems are classified as high risk if they are used in the application contexts listed in Annex III. According to Paragraph 2, the systems to be covered are those which, by virtue of their purpose, pose a high risk of harming the health and safety or fundamental rights of persons, taking into account both the severity of the potential harm and its likelihood to occur. They have to fall within the scope of Annex III. This important annex lists eight different areas of applications for high-risk AI systems: 1) biometrics (which are not already prohibited under Art. 5); 2) critical infrastructure; 3) education and vocational training; 4) employment, workers management, and access to self-employment; 5) access to and enjoyment of essential private and public services and benefits; 6) law enforcement, insofar as their use is permitted under relevant Union or national law; 7) migration, asylum, and border control management, insofar as their use is permitted under relevant Union or national law; and 8) administration of justice and democratic processes. Biometric systems under Annex III no. 1 include remote biometric systems, which are: (a) systems intended to be used for biometric categorisation, according to sensitive or protected attributes or characteristics based on their inferences; (b) systems intended to be used for emotion recognition; and (c) those which go beyond the prohibition of the use of such systems in the workplace or educational institutions prohibited in Art. 5. Critical infrastructure under Annex III no. 2 includes critical digital infrastructure, road traffic, or in the supply of water, gas, heating, or electricity.

The area of education and vocational training classifies systems intended to be used to evaluate learning outcomes, including when said outcomes are used to steer the learning process of natural persons in all levels of educational and vocational training institutions, assessing the appropriate level of education that an individual will receive or be able to access, and for monitoring and detecting prohibited behaviour of students during tests in the context of, or within, educational and vocational training institutions at all levels. Furthermore, the fourth category refers to employment and

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

workplace systems, especially AI systems in recruitment and those that make decisions in work-related relationships, such as regarding promotions or performance evaluations.

Of key importance here is the access to essential private and public services under Annex III (5), including AI systems intended to be used by, or on behalf of, public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services, AI systems for credit scoring, risk assessment for life and health insurances, and the classification of emergency calls.

Categories 6 and 7 refer to the use of AI systems in law enforcement and border control. It is important to note that the AI Act only adds another regulatory layer here since these systems must be permitted under national or Union law. Examples include the assessment of the risk of a natural person becoming the victim of criminal offences, the use of polygraphs or similar tools, predictive policing, profiling, or assessments of such risks as those regarding security, irregular migration, or health by natural persons who intend to enter (or have done so) the territory of a Member State. Further areas are the assistance to competent public authorities for the examination of applications for asylum, visa, or residence permits, as well as for associated complaints regarding the eligibility to apply for a status, including related assessments of the reliability of evidence and for the purpose of detecting, recognising, or identifying natural persons, with the exception of the verification of travel documents.

High-risk systems in the fields of administration of justice and democratic processes include the assistance of a judicial authority in researching and interpreting facts and the law, and in applying the law to a concrete set of facts, or a similar use in alternative dispute resolution, and systems used for influencing the outcome of an election.

Art. 6(3) standardises exceptions to the risk classification of Paragraph 2, according to which it is assumed that, in the case of the areas of application listed in Annex III, the AI systems used present a high risk. By way of derogation, such AI systems shall not be considered high risk if they do not pose a significant risk "to the health, safety or fundamental rights of natural persons, even if they significantly influence the outcome or significantly the outcome of a decision".

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

4.3 Systemic risks for general-purpose AI

Further to the categories of prohibited practices, high-risk, and limited and low-risk systems, a third risk category was added in the final stages of negotiations on the AI Act: the systemic risk of general-purpose AI models. The central Art. 51 is, to some extent, the counterpart of Art. 6 in that it qualifies general-purpose AI systems under the category of "systemic risks". However, the concept of systemic risk in Art. 51 is fundamentally different from that of Art. 6(1–2), thereby introducing a further categorisation of risks. Systemic risks are defined under Art. 3(65) as "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". However, the systemic risks of Art. 51 tend not to be determined according to product safety law or the relevance of fundamental rights, but rather to their cause of action and the criteria set out in Annex XIII. If a general-purpose model exceeds the threshold of 10^25 FLOPs (Floating Point Operations per Second) in terms of the cumulative number of calculations used for training, it constitutes a systemic risk (Art. 51(2)). Overall, the rationale behind this technical threshold, implying that model power under it indicates less societal risk, remains unclear from the legislator (see Mühlhoff and Ruschemeier, 2024c). This calculation threshold has little in common with the risk categorisation of Art. 6, which relates to product safety law or the impact on fundamental rights, even if it can be assumed that larger and more powerful models and the number of end users (Annex XII) can be indicators of the relevance of fundamental rights. The relationship between Arts. 6 and 51 is not explicitly clarified by the legislator; the wording suggests that providers whose model is both a high-risk system under Art. 6 and carries systemic risk under Art. 51 must comply with both obligations cumulatively (Ruschemeier, forthcoming).

5. Oversight and governance

Chapter VII of the AI Act regulates the corresponding governance structures divided into the governance at the Union level (Arts. 64–69) and the national competent authorities (Art. 70). On the Union level, the new AI Office is established at the Commission (Art. 64). Art. 3(47) defines the

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

AI Office as the "Commission's function of contributing to the implementation, monitoring and supervision of AI systems and general-purpose AI models, and AI governance", provided for in the Commission Decision of 24 January, 2024. References in this Regulation to the AI Office shall be construed as references to the Commission. One should note that, although the AI Office has been formed, many details, practicalities, and tensions in law enforcement have, for the moment at least, been left open. As it stands, the AI Office shall have different tasks, such as monitoring general-purpose AI systems, establishing codes of practice, or assisting market surveillance authorities.

Additionally, Art. 65 establishes the European Artificial Intelligence Board (AI Board), which is composed of one representative per Member State and the European Data Protection Supervisor as an observer. The participation of the AI Office is required, but it will not vote. Furthermore, the AI Board establishes two standing sub-groups to provide a platform for cooperation and exchange among market surveillance authorities, and notify them of issues related to market surveillance and notified bodies, respectively. The aim of the AI Board is to ensure cooperation and coordination between the Member States and the relevant Union bodies. To this end, the AI Board shall advise and assist the Commission and the Member States in order to facilitate the consistent and effective application of the AI Act (Art. 66(1)). Article 66 establishes different detailed tasks, such as the collection and sharing of technical expertise (Section b), the contribution to the harmonisation of administrative practices in the Member States (Section d), or supporting the Commission in promoting AI literacy, and the public's awareness and comprehension of the benefits, risks, safeguards, and rights and obligations in relation to the use of AI systems (Section f). In addition to the AI board, an advisory forum shall be established (under Art. 67) to provide technical expertise, scientifically advise the Board and Commission, and contribute to their tasks. The members shall represent a balanced selection of stakeholders, including those of industry, start-ups, small and medium-sized enterprises (SMEs), civil society, and academia. The membership of the advisory forum shall be balanced in terms of commercial and non-commercial interests and, within the category of the former, regarding SMEs and other undertakings (Art. 67(2)). Members are appointed by the Commission. Moreover, the Fundamental Rights Agency, ENISA, the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC),

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

and the European Telecommunications Standards Institute (ETSI) are permanent members of the advisory board.

Besides the AI Office and Board, the Commission shall establish a scientific panel of independent experts to support the enforcement of activities under Art. 68 of the AI Act. This is implemented by the Commission following Art. 98's process on the committee procedure (2), and is thus not included in the AI Act itself. The goal of the scientific panel is to ensure independent scientific and technical expertise in the field of AI to support the AI Office, such as by alerting it to possible systemic risks or providing advice on the classification of various general-purpose AI models and systems (Art. 68(3)). Given the unclarity and open questions in these realms, such independent scientific expertise seems urgently needed, particularly in the still developing categories for the regulation of general-purpose AI. On the national level, the Member States can call upon the experts of the scientific panel to support their enforcement activities under Art. 69.

Furthermore, Art. 70 requires the designation of Member States' national competent authorities to enforce the Regulation's provisions. Art. 70 requires the establishment of one notifying authority responsible for establishing and undertaking the procedures necessary for assessing, designating, and notifying conformity assessment bodies. As mentioned above, these private conformity assessment bodies (e.g., equal to the TÜV in Germany for product safety assessment) are active for high-risk systems only. Their monitoring is laid down in Art. 28 et seq., foreseeing that one market surveillance authority supervises the other obligations of the AI Act on a national level.

On the execution level, the AI Act provides for various penalties and fines. Art. 99(1) states that Member States shall lay down the rules on penalties and other enforcement measures, which may also include warnings and non-monetary measures, applicable to infringements of this Regulation by operators, and shall take all measures necessary to ensure their proper and effective implementation. Furthermore, Art. 99(3) states that the non-compliance with the prohibition of the AI practices referred to in Art. 5 shall be subject to administrative fines of up to 35,000,000 EUR or, if the offender is an enterprise, up to 7% of its total worldwide annual turnover for the preceding financial year, whichever is higher. The non-compliance with obligations in Arts. 16, 22, 23, 24, 26, 31, 33(1, 3, 4), 34, and 50 is subject to administrative fines of up to 15,000,000 EUR or, if the offender is an enterprise, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher (Art. 99(4)). The supply of

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

incorrect, incomplete, or misleading information to notified bodies or national competent authorities in reply to a request shall be subject to administrative fines of up to 7,500,000 EUR or, if the offender is an enterprise, up to 1% of its total worldwide annual turnover for the preceding financial year, whichever is higher (Art. 99(5)). For SMEs and start-ups, the lower percentage or amount should be applied. The rules on administrative fines are imposed by the relevant competent authorities of the Member States, such as by courts or other bodies. In Germany, for example, the competent authority would be the national market surveillance authority.

Furthermore, Art. 100 lays down provisions on administrative fines on Union institutions, agencies, and bodies imposed by the European Data Protection Supervisor (EDPS). Finally, Art. 101 establishes fines for providers of general-purpose AI models not exceeding 3% of their annual total worldwide turnover in the preceding financial year, or 15,000,000 EUR, whichever is higher. Fined violations are, for example, such procedural failures as not complying to a request for documentation or information under Art. 91 or the material infringements with relevant provisions of the AI.

6. Part III: critical analysis

6.1 Definition of AI in the AI Act: inclusive but negating AI as a socio-technical phenomenon

Addressing AI directly as an object of regulation is complex due to the multitude of views on what AI *actually* is. In the modern field – stemming from computer science, cybernetic, and mathematical approaches of the 1940s – AI tends to be used as an umbrella term for different applications and has changed throughout the decades and hype circles. Given the complexity and unclarity in the academic field of AI, not every AI-related regulation directly names the technology (e.g., the DSA). The AI Regulation explicitly addresses "AI systems", but, in its first versions, defined them so broadly that practically any software was covered.

6.1.1 Towards the final AI definition

From a legal and regulatory perspective, the definition of the Regulation's subject matter is vital as it determines its scope. A concise instantiation

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

of the regulatory object is pivotal for avoiding legal loopholes. Moreover, the requirements of legal certainty, precision, and practicability must be met. However, due to the wide range of societal segments and sciences that are directly or indirectly affected by AI, each perspective leads to its own definition of what AI is and means for the respective area. Normative regulation and social sciences do not follow a purely technical understanding of AI, but have stressed that the context of use, the social phenomena it produces, and the protected goods and interests it affects are as important as the instantiating of the technical functionality (Bareis, 2024; Ruschemeier, 2023a). It can thus be expected from the legislator to narrow down a definition that, while not necessarily encompassing the complexity of the entire scientific debate, at least serves the regulatory purpose and does justice to the individual and societal harms present with AI.

In the subsequent legislative process of the draft Regulation, the AI definition was actually changed several times. Indeed, the European Parliament's proposal of June 2023 reads: "AI system means a machine-based system that is designed to operate with varying degrees of autonomy and that can generate outputs such as predictions, recommendations or decisions that influence physical or virtual environments for explicit or implicit goals". This definition also raises follow-up questions, such as what autonomy really entails, with its notions being contested due to always being situated (Suchman, 2023; Weber and Suchman, 2016). Instead, we argued that the risk profile of AI systems can only be determined from the interplay between the technical functionality *and* the application context (i.e., a social domain), thus pointing to a necessary revision of the Act's definition of AI.

The emphasis on the regulatory filter in the Regulation's draft was not adopted as the final definition. The EU arrived at the following final reading of AI (Art. 3(1) AI Act):

> "AI system" is a machine-based system 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.

Notably this rather broad definition seeks to cope with the AI field's rapid pace of technical innovation. The definition actually includes simpler ADMs that have "explicit objectives" (but which nowadays hardly carry the denotation of AI in the debate) as much as the latest ML-run applications that are highly data intensive and yield unexpected (even if deterministic)

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

results through statistical reasoning in unsupervised learning. This broad scope is, on the one hand, problematic, as unclarity may lead to legal loopholes. On the other, the definition can also be interpreted as welcomingly broad in encompassing algorithmic systems at large.

6.1.2 Beyond technical AI: understanding AI as a socio-technical phenomenon

Despite this definition's breadth, it fails to grasp AI as a *social* phenomenon instead of a purely technical one. It is not that AI "decisions (...) can influence physical or virtual environments" only, but particularly *social* ones as well. There are two scandals connected to public agencies which effectively illustrate this point.

There are already very rudimentary algorithmic systems that can cause great societal damage. For example, the rather simple "Robodebt" scheme was installed in Australia to identify welfare fraud and overpayments in tax declarations. The scheme was not run by ML, but rather with a simple algorithm that cross-referenced payment data with annual income data provided by the Australian Tax Office (Murray, Cheong and Paterson, 2023). Robodebt was ruled unlawful and scrapped in 2020 because of the simple fact that it relied on imprecise income averaging and violated basic principles of procedural fairness and contestability, marking welfare recipients (i.e., structurally disadvantaged people) as potential cheaters.

Likewise, in the Netherlands, in the childcare benefits scandal ("*kinderopvangtoeslagenaffaire")* approximately 26,000 parents were wrongly accused by algorithmic flagging of fraudulent financial benefit applications and allowances had to be repaid to the Dutch financial ministry in full (see also Ruschemeier, 2024b). Some of the repayments totalled several tens of thousands of euros, which led to personal bankruptcies, the withdrawal of custody rights, and, ultimately, several suicides. The Dutch Data Protection authority investigated the tax and customs administration and ruled (Autoriteit Persoonsgegevens, 2020) that "the whole system was set up in a discriminatory way. [...] There was permanent and structural unnecessary negative attention for the nationality and dual citizenship of the applicants" (own translation). The scandal ultimately led to the resignation of the Rutte government and new elections in 2021.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

These two public agency scandals, based on rather simple algorithmic recommender systems, show that complex statistical inference[9] or chatbots based on the latest LLMs (what is currently referred to as AI in the public debate) are not necessarily needed to provoke massive individual and structural damage in societies. Powerful AI systems simply complicate the situation even further, as larger datasets, accelerating computing power, complex models, and server infrastructures owned and shielded by Big Tech companies can further aggravate the opacity of AI systems and distort political accountability if such errors as unrightful bias or privacy violations occur.

In their daily interactions, users never actually see code, databases, or backends of AI applications. As argued elsewhere (Bareis, 2024), AI is hardly perceived and approached as a clearly articulated, delimited, and external "thing", "model", or "tool" like the technical AI Act definition suggests. In essence, policymakers must consider that users are being presented with an AI end product that remains completely closed and opaque in its design process, operating mechanisms, and underlying normative choices. Rather than approaching AI as a self-standing entity that can be generalised (i.e., "AI is x"), recent sociological scholarship argues that AI is better understood as woven and negotiated in the everyday realities of users and society (Bodó, 2021; Suchman, 2023; Weber and Suchman 2016; Mackenzie, 2015), with its applications mediating human relationships, producing intimacies and alienations, social orders, and knowledge authorities. Here, the Australian Robodebt scheme and the Dutch childcare benefits scandal are highly indicative. AI systems (or simple ADMs) are increasingly penetrating into all spheres of society and are beginning to *mediate* and *rule* over social matters. They can enable social interaction on social media feeds with friends, but also execute physical violence (see the above-listed examples), as well as epistemic violence (derogatively portraying certain groups in society and damaging their reputations). The definition within the AI Act misses this *social* component identified by recent scholarship. Due to this technical reading, the AI Act also fails to clearly address and regulate some fundamental social risks caused by AI (see analysis in III). A less abstract and more empirical and hands-on approach understands

9 See, for example, the debate on the more complex US recidivism score system used in the US judiciary that uses the probability of criminals reoffending in its recommendations for or against parole, called the "Correctional Offender Management Profiling for Alternative Sanctions system" (COMPAS) (Angwin et al, 2022).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

AI not only as algorithmic performativity, but also includes the social phenomena it produces, and the *meaning* ascribed to them. Such a perspective would clearly make the EU AI regulatory framework more accessible and closer to every-day user experiences. Given that European standardisation bodies are currently trying to implement the AI Act on the Member State level (Gamito and Marsden, 2024), it remains to be determined how these social and epistemic risk dimensions can be entangled in a process of quantification the creation of risk scores (discussed in greater detail below).

6.2 Dualistic regulatory structure: the misfit of applying product safety law on fundamental rights protection[10]

The AI Regulation aims to not only improve the functioning of the internal market, but also to promote human-centred and trustworthy AI and ensure a high level of protection against harmful effects on health, safety, fundamental rights, democracy, the rule of law, and the environment – all while simultaneously promoting innovation (Recital 1).

The different duties the AI Act seeks to fulfil resemble an ambitious attempt to politically square the circle. It aims to satisfy various interests which are at odds with each other: the trustworthy and fundamental rights pillar to protect human-centred rights needs to accommodate the economic interest to which the vast profit potential of user data points – just to finally include all pillars in a harmonised but competitive free-market approach. Here, it remains to be seen whether these objectives – in particular, the protection of fundamental rights – can be achieved through a regulatory structure based on product safety law and risk-based governance. A growing number of scholars have rightly criticised this regulatory approach (Almada and Petit, 2023; Guijarro, 2023; Smuha et al, 2021; Veale and Borgesius, 2021). Although the adoption of an existing regulatory structure offers the advantage of established models and concepts, AI systems themselves are fundamentally different from the products on which the concept of product safety law and the tradition of risk-based governance are based. Such is also the understanding of risk in the protection of fundamental rights and product safety law.

10 The following critique (6.2-6.3) is an English version of the arguments made in Ruschemeier (forthcoming).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

6.2.1 Physicality and actors: AI systems are no fixed products

The regulation is characterised by the idea of a certain physicality of AI systems. Their purposes shall be determined ex ante and their changes accompanied by delegated acts. However, the extent to which the particularities of more and evolving complex systems can be captured is doubtful (Edwards, 2022b). This is because AI systems change as a result of new data creation and processing (see the current rise of synthetical data), steady model development (as with foundation models), or the growing platformisation and infrastructural integration of other possible systems (centralisation). The regulatory strand of product safety law, on the other hand, is based on assumptions that do not correspond to how AI systems function, even if software can be categorised as a product under the new legislation. AI systems are not products that are manufactured once and then placed on the market, and used only for fixed purposes in specific contexts. Instead, they are increasingly being used dynamically in different contexts with different effects on individuals and groups (Edwards, 2022b). The actors involved are also fundamentally different: product manufacturers tend to be experts in their production processes and are rightly the addressees of safety requirements. Moreover, the development of AI systems also differs, often involving different actors and institutions, with smaller developers in particular relying on building blocks, datasets, and other resources than larger companies in order to develop their own products, especially given the recent turn to the platformisation of AI. The *downstream* use of AI systems can therefore look very different from a system's original development.

6.3 Different understanding of risks and harms depict paradigms that are not compatible

In addition, product safety law is based on a specific understanding of risk that cannot be transferred to the socio-technical hazards and risks posed by AI systems. Therefore, it is unlikely for these risks to be adequately capture by a regulatory system based on the categories of product safety law. The understanding of risk in product safety law, as part of private law, is based on the reference to potential damage, which is then compensated through such claims as damages for injury to bodily integrity. Firstly, product safety law and the protection of fundamental rights are based on differ-

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ent concepts of risk. Furthermore, normative safeguards of freedom, such as human autonomy, cannot be measured exclusively in numerical terms and translated into metrics or standards, but always depend on a case-by-case assessment. The AI Act neither addresses the gaps between different regulations, such as data protection and discrimination law, nor clarifies important concepts, but could even exacerbate the problem (Adams-Prassl, 2022). In practice, the requirements of the AI Regulation itself are undermined by the presumption of compatibility under Articles 40 et seq.

This categorical tension in the regulatory approach stems from the dominance of a "risk-based" regulatory assessment paradigm that began to dominate the Anglo-Saxon world in the 1980s–1990s (Black, 2005; Hood, Rothstein and Baldwin, 2001) and has ever since influenced the EU in such areas as safety standardisation for the chemical and food industries, or in environmental impact assessments (Orwat et al, 2024; Paul, 2021). The paradigm of risk-based regulation resembles a shift away from a rather prescriptive approach based on formal legal statutes and normative principles. Instead, risk-based regulation promises empirically based and adaptive "cost-benefit" practices, requiring numerical assessments and classifications (Black, 2010). This paradigm not only implies that risks must be measurable (hence, "quantifiable"), but also that they can be managed and, to some degree, accepted: it is not about *avoiding* harms, but about their acceptable and bearable societal *handling*, ranging from acceptable to unacceptable harms, and deriving the appropriate levels of such regulatory measures as tests, benchmarking, approvals, requirements, bans, or moratoria. The ideal outcome is to find the right balance between over- and under-regulation. However, as argued elsewhere, risk-based regulation needs "sufficiently unambiguous and concrete criteria or principles for what constitutes relevant risks" (Orwat et al, 2024, p. 11). As such, finding the right risk scheme for AI is a particular challenge.

The problem with the EUC's reliance on this product safety risk-regulatory rationale with AI (for a critical reconstruction, see Paul, 2023) is that the nature and understanding of risk in the context of the protection of fundamental rights is by no means uniform. The risk to fundamental rights is not synonymous with potential harm from, for instance, chemicals in food or radiation, but lies in the potential violation of the fundamental right, which in turn does not necessarily presuppose harm. The understanding of constitutional protection of fundamental rights follows a precautionary principle, e.g., data protection is "protection beforehand" – that is, in advance of the actual danger. There has been an increase in the number of

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

proposals emerging which use risk regulatory metrics and thresholds to represent elusive values, such as "fairness", "justice", or "privacy" in order to make them manageable. However, given the strong contextualisation of anti-discrimination law, for example, the ability to translate normative values into numerical measures is limited from the outset (Ruschemeier, 2023b). See, for example, the AIEI report (Hallensleben and Hustedt, 2020) "From principles to practice", which exactly aims at establishing those metrics. However, it has been (somewhat problematically) suggested that rights and normative values can be quantified or even "cleared" with each other. Here, rights become labelled like washing machines, suggesting a legal clarity which is not the case. Factors of contextuality, residual risks, or intangible subjective harms, such as reputational damage, become completely neglected in this reductionist approach.

6.4 Watered down Fundamental Rights Impact Assessment

There are also concerns about the dual regulatory strategy's ability to strike a suitable balance between minimising risk and fostering innovation for applications in the public interest. AI systems used in the medical field (and subject to the MDR) will always be high-risk systems, while such lifestyle applications as smartwatches or fitness trackers will be subject to the requirements for high-risk systems, but may not even be subject to the general requirements of Art. 50. Such applications can pose significant risks, for example, with regard to the collection of health data. Health tracking has the potential to aggravate the individualisation of risk under the disguise of algorithmic and profit efficiency, thus undermining a system of public service and solidarity with weaker social-economic strata. There is a high likelihood that, under the logic of cost-efficiency, these strata will have to pay higher fees for premiums as the neighbourhoods in which they live provide aggregate health, education, or crime data. With the ongoing privatisation of the health and insurance sector in many countries, it is reasonable to assume that this will result in advantages for companies with considerable economic resources to challenge high-risk classifications, as expressly provided for in Art. 6 para. 3.

The AI Act establishes a Fundamental Rights Impact Assessment (FRIA) in Art. 27, which was included after an intervention of various academics in the legislative process (Mantelero, 2022; Liberties, ECF and ECNL, 2023). It reflects the impacts on fundamental rights to a certain degree. However,

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

the provision was watered down during the legislative process and now only applies to deployers that are bodies governed by public law, or private entities providing public services, and deployers of high-risk AI systems referred to in points 5(b) and (c) (credit and insurance scoring) of Annex III. This is unfortunate since all the other use cases listed in Annex III can have heavy implications and inferences with fundamental rights. Moreover, it does not seem particularly clear why the FRIA specifically addresses public entities (which are bound by fundamental rights anyway) and not private actors, who have no such binds (see also Mantelero, 2024). The insufficiency of the FRIA is one of the most significant misses of the AI Act.

7. Governance and the imbalance between private and public interest

7.1 Democratically unsupervised private standardisation procedures

In practice, the risk classification of Art. 6 and the subsequent obligations of the Regulation's third section are significantly influenced by the standardisation norms of Arts. 40 et seq. When harmonised standards are established, it is assumed that the corresponding AI systems comply with the requirements of Chapter 2 of Part 3 of the Regulation (Arts. 8–15). These requirements include, for example, obligations for risk management systems (Art. 9), data governance (Art. 10), technical documentation (Art. 11), and human oversight (Art. 14).

Standardisation procedures are well known and established in product safety law. However, the Regulation also stipulates that high-risk AI systems must meet certain mandatory requirements that align with the European interests of health, safety, and the protection of fundamental rights, such as risk and data management, transparency, and human oversight. It should then be possible to implement these requirements in harmonised standards developed by the European standardisation bodies. Regarding the relevance of systems to fundamental rights, there is no experience at the level of EU regulation of how these can be standardised. Standardisation focuses on areas where the state of the art is particularly relevant, and therefore the consideration of fundamental rights is not given prima facie. In this context, the development of standards cannot be purely technical (i.e., based on computer science and engineering). It must have a social dimension linked to considerations and findings from the humanities and social sciences, including law.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Moreover, this far-reaching power of definition, from which Member States can only deviate in individual cases by means of a single authorisation (according to Art. 47), does not correspond to democratic legitimation, but lies exclusively with private standardisation organisations. This standardisation is not subject to parliamentary debate, but is limited to the adoption of a consensus by the interest groups of each draft standard, clearly pointing to a democratic deficit. In practice, these interest groups are dominated by the leading international economic players most affected by the standard in question, mirroring an imbalance with the absent public interest actors. At least, the ECJ has now ruled that harmonised technical standards must be freely accessible and thus available free of charge (Public.Resource.Org, Inc. and Right to Know CLG v European Commission, 2024). However, the obligation to assess the impact on fundamental rights in Art. 27 does not change this state of affairs. This is because it does not lay down any requirements for the standardisation process, but solely obliges certain operators (Art. 3(4)) to conduct an impact assessment on how the system affects "fundamental rights" in certain cases. Given the Regulation's objective, this obligation would have been desirable in principle for *all* AI systems, but was considerably weakened in the legislative process. The obligation now only applies to public bodies or private operators providing public services and operators of systems under Annex III No. 5 lit. b) (credit scoring with the exception of financial fraud detection).

The different understandings of harm and risk by product safety law and fundamental rights protection are compounded by the enforcement and governance structures of the AI Regulation. The Regulation's lofty goals of protecting fundamental rights are largely dependent on private standardisation organisations (CEN and CENELEC) and procedures (see here Gamito and Marsden, 2024).[11] The product safety approach of technical standards, coupled with the presumption of conformity of Arts. 40 ff., is intended to both provide flexibility and avoid overburdening the supervisory authorities. This is convincing for the area of product safety law, where there is expertise and practical experience regarding standardisation and the implementation of safety in technical standards. However, when

11 Although they are not mentioned by name in the text of the Regulation, the standardisation organisations will have a crucial role to play. The Commission has already issued the first standardisation mandate (C(2023)3215) in support of Union policy on AI, which has been accepted by CEN and CENELEC (European Commission, no date). According to Art. 1 of the Implementing Decision, the standards shall be developed by 30 April, 2025.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

assessing the risk to fundamental rights, technical standardisation is highly problematic.

The classification for high-risk systems shall be based on a preliminary self-assessment, so the law is likely to exacerbate the problem of developers deliberately misclassifying their innovations so as to circumvent having to comply with the strict requirements. Suppliers who consider that their system is not high risk according to their own assessments (which falls under the use cases of Annex III para. 3), must first document this assessment before placing the system on the market (Ruschemeier, no date).

7.2 Missing participation of affected subjects

Moreover, the perspective of fundamental rights holders is not even considered in the AI Act; however, the relevance of fundamental rights cannot be examined in a supposedly technical vacuum, but only in relation to the affected subjects. It is unclear to what extent private standardisation organisations, which have neither the expertise nor the structures to assess fundamental rights, should be able to do this. It is doubtful whether fundamental rights implications can be adequately taken into account within this framework, despite all of the possibilities related to stakeholder participation. Collective dimensions, such as those that play a role in labour law, are not mentioned in the AI Act (Adams-Prassl, 2022). Nor does it contain a provision equivalent to Art. 88 of the GDPR, which would allow Member States to adopt more specific national provisions for the employment context, which would further limit their willingness to experiment with regulation.

7.3 The problem of algorithmic discrimination escaping the categories of anti-discrimination

The joint opinion of the European Data Protection Board and the European Data Protection Supervisor on the AI Act (European Data Protection Board, 2021) rightly points out that risks to groups of individuals or to society as a whole, such as group discrimination or the freedom of political expression, are not adequately addressed in the AI Act. This also applies to the specific risks of discrimination against individuals by data-intensive technologies. The AI Act mentions discrimination and social risks in sever-

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

al places and calls for studies on prohibited discrimination in the context of data governance (Art. 10 para. 2 lit. f). Further references can only be found in Art. 77 and Annex IV on supervision and technical documentation. The AI Act does not decide when further discrimination is undesirable or risky, which is highly relevant in terms of fundamental rights. The problem of algorithmic discrimination escaping the categories of anti-discrimination law remains unresolved (Wachter, 2023). In terms of supra-individual effects, Annex III categorises the areas of administration of justice and democratic processes as high-risk areas, not regarding expression, but in terms of systems intended to influence the outcome of an election or the voting behaviour of natural persons. While this is certainly welcome, it only addresses part of the problem.

7.4 The loophole of addressing recommendation systems on platforms as high-risk systems

The risk of influencing elections through political advertising should also be regulated. However, the proposed Regulation provides for the possibility of political targeting based on the consent of the data subject (Art. 18(1)(b)). However, this consent-instrument cannot consistently protect fundamental rights in the digital context as the large flood of information is simply not comprehensible or deliberately difficult to access in platform option settings (e.g., when seeking to obtain consent from hundreds of different data processors; Ruschemeier, 2022). The parliamentary proposal on the risk category of recommendation systems of very large online platforms and search engines under the DSA was deleted in the final version. As such, a large part of the AI systems that most people interact with on a daily basis are not covered by the AI Act as high-risk systems. Considering how much time citizens spend on social media – with global interactions averaging 2.31 hours per day (and up to 5.01 on smartphones) in 2023 – the impact of the consumed and widely disseminated content is not to be underestimated (Kemp, 2023). These numbers are all the more worrying for democratic processes in societies given that prior research has clearly pointed to a growing polarisation, political fragmentation, and self-reinforcing of political (often populist or extremist) opinion through echo chambers on social media platforms (Barberá, 2020; Fisher, 2022). Problematically, this has also affected how the DSA tackles misinformation (Arts. 14, 14 III, 19), as: "when polarization is high, misinformation quickly proliferates" (Cinelli et

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

al, 2021, p. 5). Considering the rate at which synthetically generated data is currently flooding the internet and social media, problematic and extremist content is likely to increase in scale and quality.

The AI Act does not address the dissemination and information power asymmetry of large platform companies, which contributes significantly to AI risks (Mühlhoff and Ruschemeier, 2024b). The deletion of the high-risk categorisation also prevents the important interaction between the DSA and the AI Act in the overall European regulatory framework, where it would have been informative to examine how the obligations under the DSA and the AI Act relate to very large providers.

7.5 Lobbying and the risk of tech-solutionism[12]

Some of the AI systems classified as high-risk under the AI Act are highly problematic. Indeed, certain systems which are not scientifically researched or validated, and have no clear or beneficial use for individuals or the public, can ultimately be mainstreamed or normalised. First, it is unclear why systems that are intended to be used to influence the outcome of an election (Annex III 8(b)) are even legal in the first place. Elections should be free and uninfluenced in order to be democratically legitimate, and AI systems which influence their results have no legitimate purpose in democratic states and should be banned. Systematically, it is not understandable why the use of AI by judicial authorities is not subject to national legal reservation, such as law enforcement, since both areas of use are highly influenced by national legislation. The (highly) scientifically questionable use of polygraphs in 1(c), 6(b), and 7(a) are legitimised as high-risk systems without any indication for their effectiveness (lie detectors have absolutely no scientific grounding, and can thus be termed pseudo-science). Many of the more restrictive takes on high-risk systems and general-purpose AI have been lowered throughout the legislative process. Consequently, the overall protection of fundamental rights throughout the AI Act has suffered substantially.

Reports by the Corporate Europe Observatory (Schyns, 2023) and Transparency International (Kergueno et al, 2021) have proven how Big Tech, corporate think tanks, and trade and business associations have been disproportionately active in blocking and watering down AI regulation in

12 Parts of this section are taken from Bareis (2023b, 2024).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Brussels. As discussed elsewhere on the final trilogue between the Commission, Parliament, and Council in late 2023, Big Tech efforts on the AI Act have been substantial (Bareis, 2023b). In 2023 alone, industry lobbyists had by far the most meetings with the EU commission on the AI Act, with 86% (73 out of 98) of all behind-closed-door meetings, and were most active in agenda and standard setting (Corporate Europe Observatory, 2023; Kergueno et al, 2021). For the AI Act, "tech companies have reduced safety obligations, sidelined human rights and anti-discrimination concerns" (Schyns, 2023, p. 3). Leaked documents strikingly show how companies have tried to pressure policy makers with their deregulatory agendas by staging such narratives as "Big tech is 'irreplaceable' when it comes to problem solving", "we're just defending SMEs and consumers", or "Europe wins the tech race against China, or it falls back into the Stone Age" (Bank et al, 2021, p. 27). This tech-solutionist take on AI is converting AI into an inevitability, catering to a narrative that suggests "only advancement in AI technology can assure that the current level of living can be maintained and future prosperity secured" (Bareis and Katzenbach, 2022, p. 868). With such an AI hype and the argumentative force of the TINA (there-is-not-alternative) mindset, politics becomes pressured towards an unreflective and unchecked uptake of AI across society. Instead, politics should act like a critical watchdog given the public's mandate, and clearly and effectively address the chances and risks of this multifaceted technology for the benefit of all.

In the final round of discussions on the AI Act, lobbying efforts have been directed against the designation of general-purpose AI as a "high risk" category, with industry representatives fearing that it would overburden and stifle innovation with strict conformity assessments. Such European startups as Mistral and Aleph Alpha joined forces with US Big Tech companies and derailed, with direct ties to political executives in France or Germany, the policy-making process in the last metres. Industry managed to water down the binding fundamental right assessment proposed by the European Parliament on general-purpose AI into mere transparency rules (Corporate Europe Observatory, 2023; Hartmann, 2023).

8. Conclusion and outlook

Despite all the criticism, the adoption of the AI Act is a milestone in digital regulation at the European level. It is important that the EU legislator

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

has recognised and regulated many problematic practices, such as the ban on indiscriminate scraping to create facial recognition databases, emotion recognition systems, and the risks of insurance and credit scoring.

However, we argue that the AI Act also has major caveats to effectively regulate AI in the service of the public interest of European citizens. The Regulation's enforcement is currently underway in the 27 EU Member States and transfers a great deal of power to private standard-setting organisations. As we have argued, this is problematic from the perspective of democratic legitimacy, as private organisations are given too much discretion in deciding upon sensitive rights and trade-offs of privileges and burdens in our society with respect to AI. Adding to the perspective of democratic inclusion, a stronger participation of affected subjects, a deeper understanding of anti-discrimination, and a more hands-on definition of AI, doing justice to the *social* phenomena it produces, would significantly contribute to the overall acceptance of the Regulation and help close its current loopholes.

While some of these points could be potentially revised in the aftermath of the AI Act's implementation, there are some decisions on the overall structure and design of the Regulation that seem unsuited to its overall purpose. The AI Act applies product safety law for the sake of fundamental rights protection. However, such a legal framework is ill-equipped to cover the socio-technical hazards and risks posed by AI systems. These systems are fundamentally different from the products on which the concept of product safety law and the tradition of risk governance are based. Risk regulation originates from safety standard setting of clearly measurable physical harms, such as those from chemicals or radiation. However, normative safeguards, rights, and political threats to democracy cannot be measured exclusively in numerical terms and translated into metrics or standards. The next few years will show to what extent the ambitious approach of combining product safety law with the protection of fundamental rights can be effectively implemented in practice.

It thus remains, seemingly by design, why recommendation systems on platforms are not marked as high-risk systems, given the very individual and structural damage they can inflict on reputations, cause democratic polarisation, and further exacerbate the power of Big Tech companies. These companies are currently some of the world's most profitable and have, time and again, proven that they aim for big profit, and not for the public good.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

All of this shows that the (European) discussion about AI regulation cannot end with the AI Act. The aim of our contribution is to further stimulate the discussion about the social risks and sensible applications in order to revise and improve the AI legal policy frameworks currently implemented around the world. Law, acting as a powerful instrument to distribute the benefits and burdens of this technology for the greater social good, must not lag behind Big Tech's consistently questionable endeavours. It must be socially leading.

References

Abraham, Y. (2024) '"Lavender": the AI machine directing Israel's bombing spree in Gaza'. *+972 Magazine*, 3 April [Online]. Available at: https://www.972-mag.com/lavender-ai-israeli-army-gaza/ (Accessed: 28 January 2025).

Adams-Prassl, J. (2022). 'Regulating algorithms at work: lessons for a "European approach to artificial intelligence"', *European Labour Law Journal*, 13(1), pp. 30–50.

AI HLEG (2019) *Ethics guidelines for trustworthy AI*. European Commission [Online]. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Accessed: 28 January 2025).

AI HLEG (2020) *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. European Commission [Online]. Available at: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment (Accessed: 28 January 2025).

Almada, M. and Petit, N. (2023) *The EU AI Act : a medley of product safety and fundamental rights?* Working Paper. European University Institute [Online]. Available at: https://cadmus.eui.eu/handle/1814/75982 (Accessed: 28 January 2025).

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2022) 'Machine bias' in Martin, K. (ed.) *Ethics of data and analytics. Concepts and Cases.* New York: Auerbach Publications, pp. 254-264.

Arias-Cabarcos, P., Khalili, S. and Strufe, T. (2023) '"Surprised, shocked, worried": user reactions to Facebook data collection from third parties', in *Proceedings on Privacy Enhancing Technologies* 2023(1), pp. 384–399.

Autoriteit Persoonsgegevens (2020) *Werkwijze Belastingdienst in Strijd Met de Wet En Discriminerend*. Den Haag [Online]. Available at: https://web.archive.org/web/20200719043135/https://autoriteitpersoonsgegevens.nl/nl/nieuws/werkwijze-belastingdienst-strijd-met-de-wet-en-discriminerend (Accessed: 28 January 2025).

Bächle, T. C. and Bareis, J. (Eds.). (2025). *The Realities of Autonomous Weapons*. Bristol University Press.

Bächle, T.C. and Bareis, J. (2022) '"Autonomous weapons" as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies', *European Journal of Futures Research*, 10(20), pp. 1-18.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Bank, M., Duffy, F., Leyendecker, V. and Silva, M. (2021) *The lobby network: Big Tech's web of influence in the EU*. Brussels and Cologne: Corporate Europe Observatory and LobbyControl.

Barberá, P. (2020) *Social media, echo chambers, and political polarization*. Cambridge: Cambridge University Press.

Bareis, J. (2024). *Ask Me Anything!* 🦦 *How ChatGPT Got Hyped Into Being* (preprint). Center for Open Science [Online]. Available at: https://doi.org/10.31235/osf.io/jzde2 (Accessed: 28 January 2025).

Bareis, J. (2023a) 'BigTech's efforts to derail the AI Act', *Verfassungsblog* [Online]. Available at: https://verfassungsblog.de/bigtechs-efforts-to-derail-the-ai-act/ (Accessed: 28 January 2025).

Bareis, J. (2023b) 'Die EU und Big Tech riskieren eine Krise des Wissens', *Der Standard*, 27 December [Online]. Available at: https://www.derstandard.at/story/3000000200822/die-eu-und-bigtech-riskieren-eine-krise-des-wissens (Accessed: 28 January 2025).

Bareis, J. (2024) 'The trustification of AI. Disclosing the bridging pillars that tie trust and AI together', *Big Data and Society*, 11(2), [Online]. Available at: https://doi.org/10.1177/20539517241249430 (Accessed: 28 January 2025).

Bareis, J. and Katzenbach, C. (2022) 'Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics', *Science, Technology, and Human Values*, 47(5), pp. 855–881.

Barkane, I. (2022) 'Questioning the EU proposal for an Artificial Intelligence Act: the need for prohibitions and a stricter approach to biometric surveillance', *Information Polity*, 27(2), pp. 147–162.

Bhuta, N., Beck, S. and Liu, H.-Y. (2016) *Autonomous weapons systems: law, ethics, policy*. Cambridge: Cambridge University Press.

Black, J. (2005) The emergence of risk-based regulation and the new public risk management in the United Kingdom. *Public Law*, Autumn, pp. 510–546.

Black, J. (2010) "Risk-Based Regulation: Choices, Practices and Lessons Being Learnt." Paris: OECD [Online]. Available at: https://doi.org/10.1787/9789264082939-11-en (Accessed: 28 January 2025).

Bock, K., Kühne, C.R., Mühlhoff, R., Ost, R.M., Pohle, J. and Rehak, R. (2020) 'Data protection impact assessment for the Corona App'. *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.3588172 (Accessed: 28 January 2025).

Bodó, B. (2021) 'Mediated trust: a theoretical framework to address the trustworthiness of technological trust mediators', *New Media and Society*, 23(9), pp. 2668–2690.

Broeders, D., Cristiano, F. and Kaminska, M. (2023). 'In search of digital sovereignty and strategic autonomy: normative power Europe to the test of its geopolitical ambitions', *Journal of Common Market Studies*, 61(5), pp. 1261–1280.

Burkhardt, S. and Rieder, B. (2024) 'Foundation models are platform models: prompting and the political economy of AI', *Big Data and Society*, 11(2), pp. 1–15.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W. and Starnini, M. (2021). 'The echo chamber effect on social media', *Proceedings of the National Academy of Sciences*, 118(9) [Online]. Available at: https://doi.org/10.1073/pnas.2023301118 (Accessed: 28 January 2025).

Citron, D. and Pasquale, F. (2014) 'The scored society: due process for automated predictions', *Washington Law Review*, 89(1), pp. 1–33.

Corporate Europe Observatory. (2023) *Byte by byte*. Corporate Europe Observatory [Online], 17 November. Available at: https://corporateeurope.org/en/2023/11/byte-byte (Accessed: 28 January 2025).

Council of the EU. (2023) *New rules for machinery: Council gives its final approval*. Council of the EU [Press release], 22 May [Online]. Available at: https://www.consilium.europa.eu/en/press/press-releases/2023/05/22/new-rules-for-machinery-council-gives-its-final-approval/ (Accessed: 28 January 2025).

Datenethikkommission (2019) *Opinion of the Data Ethics Commission*. Datenethikkommission [Online]. Available at: https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/datenethikkommission-abschlussgutachten-kurz.pdf;jsessionid=7D05795C3957AB73E7DAE2D912B49757.live861?__blob=publicationFileandv=3 (Accessed: 28 January 2025).

'Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive')' (2005) *Official Journal* L 149, 11.6.2005, pp. 22–39 [Online]. Available at: http://data.europa.eu/eli/dir/2005/29/oj (Accessed: 29 January 2025).

'Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys' (2009) *Official Journal L 170, 30 June, pp. 1–37* [Online]. Available at: http://data.europa.eu/eli/dir/2009/48/oj (Accessed: 28 January 2025).

Doezema, T. and Frahm, N. (2023) 'The law isn't lagging behind AI. It's leading it', *The New Atlantis*. Available at: https://www.thenewatlantis.com/publications/how-the-state-built-this-ai-moment (Accessed: 28 January 2025).

Dul, C. (2022) 'Facial recognition technology vs privacy: the case of Clearview AI', *Queen Mary Law Journal*, 3, pp. 1–24.

Ebers, M., Hoch, V.R.S., Rosenkranz, F., Ruschemeier, H. and Steinrötter, B. (2021) 'The European Commission's proposal for an Artificial Intelligence Act – a critical assessment by members of the Robotics and AI Law Society (RAILS)', *J*, 4(4), pp. 589–603.

EDPS (2024, May 24) *European Commission's use of Microsoft 365 infringes data protection law for EU institutions and bodies*. EDPS [Online]. Available at: https://www.edps.europa.eu/press-publications/press-news/press-releases/2024/european-commissions-use-microsoft-365-infringes-data-protection-law-eu-institutions-and-bodies (Accessed: 28 January 2025).

Edwards, L. (2022a) *Expert explainer: the EU AI Act proposal*. Ada Lovelace Institute, 8 April [Online]. Available at: https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer/ (Accessed: 28 January 2025).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Edwards, L. (2022b) *Expert opinion: regulating AI in Europe*. Ada Lovelace Institute, 31 March [Online]. Available at: https://www.adalovelaceinstitute.org/report/regulating-ai-in-europe/ (Accessed: 28 January 2025).

Edwards, L. (2022c) *The EU AI Act: a summary of its significance and scope*. Ada Lovelace Institute [Online]. Available at: https://www.adalovelaceinstitute.org/ (Accessed: 28 January 2025).

Elmer, G. (2003) *Profiling machines*. *Mapping the Personal Information Economy.* Cambridge (MA): The MIT Press.

Ernst, C. (forthcoming) In *The Realities of Autonomous Weapon Systems*, in Bächle, T.C. and Bareis, J. (eds.) Bristol: Bristol University Press.

European Commission (2008) *New legislative framework*. European Commission [Online]. Available at: https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en (Accessed: 28 January 2025).

European Commission (2020a) *Op-ed by Commission President von Der Leyen*. European Commission [Online], 19 February. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ac_20_260 (Accessed: 28 January 2025).

European Commission (2020b) Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Available at: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment (Accessed: 28 January 2025).

European Commission (2020c) *White paper on artificial intelligence: a European approach to excellence and trust*. European Commission [Online]. Available at: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (Accessed: 28 January 2025).

European Commission (2021) *Commission staff working document proposal for a Regulation of the European Parliament and of the Council, SWD/2021/84 Final*. European Commission [Online]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0084 (Accessed: 28 January 2025).

European Commission (2023) *European Chips Act*. European Commission, 21 September [Online]. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-chips-act_en (Accessed: 28 January 2025).

European Commission (2024, April 21) *Shaping Europe's digital future*. European Commission [Press release] [Online]. Available at: https://digital-strategy.ec.europa.eu/en (Accessed: 28 January 2025).

European Commission (no date) *C(2023)3215 – Standardisation request M/593* [Online] Available at: https://ec.europa.eu/growth/tools-databases/enorm/mandate/593_en (Accessed: 28 January 2025).

European Data Protection Board (2021) *EDPB-EDPS joint opinion 5/2021 on the proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* European Data Protection Board, 18 June [Online]. Available at: https://www.edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en (Accessed: 28 January 2025).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

European Group on Ethics in Science and New Technologies (2018) *Artificial intelligence, robotics and 'autonomous' systems*. European Commission [Online]. Available at: https://lefis.unizar.es/wp-content/uploads/EGE_Artificial-Intelligence_Statement_2018.pdf (Accessed: 28 January 2025).

Fisher, M. (2022) *The chaos machine: the inside story of how social media rewired our minds and our world*. Boston and New York: Little Brown and Company.

Frahm, N. and Schiølin, K. (2023) 'Toward an "ever closer union": the making of AI-ethics in the EU', *STS Encounters*, 15(2) [Online]. Available at: https://doi.org/10.7146/stse.v15i2.139808 (Accessed: 28 January 2025).

Gamito, M.C. and Marsden, C. T. (2024) Artificial intelligence co-regulation? The role of standards in the EU AI Act. *International Journal of Law and Information Technology*, *32*(1) [Online]. Available at: https://doi.org/10.1093/ijlit/eaae011 (Accessed: 28 January 2025).

Gless, S. and Janal, R. (2023) '§ 2 Anwendungsbereich und Adressaten' in E. Hilgendorf and D. Roth-Isigkeit (eds.) *Die neue Verordnung der EU zur Künstlichen Intelligenz: Rechtsfragen und Compliance*. Munich: C.H. Beck, pp. 15-33.

Goh, H.-H. and Vinuesa, R. (2021) 'Regulating artificial-intelligence applications to achieve the sustainable development goals', *Discover Sustainability*, 2(52) [Online]. Available at: https://doi.org/10.1007/s43621-021-00064-5 (Accessed: 28 January 2025).

Guijarro Santos, V. (2023) 'Nicht Besser als Nichts. Ein Kommentar zum KI Verordnungsentwurf', *Zeitschrift Für Digitalisierung Und Recht*, 1, pp. 23–42.

Hacker, P. (2018) 'Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law', *Common Market Law Review*, 55(4), pp. 1143–1185.

Hallensleben, S. and Hustedt, S. (2020) *From principles to practice. An interdisciplinary framework to operationalise AI ethics*. Bertelsmann Stiftung [Online]. Available at: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf (Accessed: 28 January 2025).

Hartmann, T. (2023) *AI Act: French government accused of being influenced by lobbyist with conflict of interests*. Euractiv, 21 December [Online]. Available at: https://www.euractiv.com/section/artificial-intelligence/news/ai-act-french-government-accused-of-being-influenced-by-lobbyist-with-conflict-of-interests/ (Accessed: 28 January 2025).

Heaven, D. (2020) 'Why faces don't always tell the truth about feelings', *Nature*, 578(7796), pp. 502–504.

Hill, K. (2022) 'The secretive company that might end privacy as we know it' in Martin, K. (ed.) *Ethics of Data and Analytics. Concepts and Cases.* New York: Auerbach Publications, pp. 170-178.

Hood, C., Rothstein, H. and Baldwin, R. (2001) *The government of risk: understanding risk regulation regimes*. Oxford: Oxford University Press.

Hong, S.-H. (2020) *Technologies of speculation*. New York: New York University Press.

Kello, L. (2017) *The virtual weapon and international order*. New Haven: Yale University Press.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Kemp, S. (2023) *Digital 2023: Global overview report* [Online]. Available at: https://datareportal.com/reports/digital-2023-global-overview-report (Accessed: 28 January 2025).

Kergueno, R., Aiossa, R., Pearson, L., Corser, N.S., Teixeira, V. and van Hulten, M. (2021b) *Deep pockets, open doors*. Transparency International EU [Online]. Available at: https://transparency.eu/wp-content/uploads/2024/10/Deep_pockets_open_doors_report.pdf (Accessed: 28 January 2025).

Koloßa, S. (2020) 'The GDPR's extra-territorial scope. Data protection in the context of international law and human rights law', *ZaöRV*, 4, pp. 791–818.

Krarup, T. and Horst, M. (2023) 'European artificial intelligence policy as digital single market making', *Big Data and Society*, 10(1) [Online]. Available at: https://doi.org/10.1177/20539517231153811 (Accessed: 28 January 2025).

Kreiß, C. (2019) 'Ethik-Institut an der TU München: Ein vielsagender geheimer Vertrag mit Facebook', *Der Tagesspiegel Online*, 19 December [Online]. Available at: https://www.tagesspiegel.de/wissen/ein-vielsagender-geheimer-vertrag-mit-facebook-4129213.html (Accessed: 28 January 2025).

Krügel, S., Ostermaier, A. and Uhl, M. (2022) 'Zombies in the loop? Humans trust untrustworthy AI-advisors for ethical decisions', *Philosophy and Technology*, 35(1), pp. 1–37.

Laux, S., Wachter, J. and Mittelstadt, B. (2023) 'Trustworthy artificial intelligence and the European Union AI Act: on the conflation of trustworthiness and acceptability of risk', *Regulation and Governance*, 18(1), pp. 3–32.

Von der Leyen, U. (2019) *Speech by President-elect Ursula von der Leyen at the 2019 Paris Peace Forum* [Online]. Available at: https://ec.europa.eu/commission/presscorner/detail/en/speech_19_6270 (Accessed: 28 January 2025).

Liberties, ECF and ECNL (2023) *Open Letter. The AI Act must protect the rule of law* [Online]. Available at: https://ecnl.org/sites/default/files/2023-09/AI_and_RoL_Open_Letter_final_27092023.pdf (Accessed: 28 January 2025).

Mackenzie, A. (2015) 'The production of prediction: what does machine learning want?' *European Journal of Cultural Studies*, 18(4–5), pp. 429–45.

Mager, A., Norocel, O.C. and Rogers, R. (2023) 'Advancing search engine studies: the evolution of Google critique and intervention', *Big Data and Society*, 10(2) [Online]. Available at: https://doi.org/10.1177/20539517231191528 (Accessed: 28 January 2025).

Malgieri, G. and Pasquale, F. (2024) 'Licensing high-risk artificial intelligence: toward ex ante justification for a disruptive technology', *Computer Law and Security Review*, 52(105899) [Online]. Available at: https://doi.org/10.1016/j.clsr.2023.105899 (Accessed: 28 January 2025).

Mantelero, A. (2024) 'The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template', *Computer Law & Security Review*, *54*(106020) [Online]. Available at: https://doi.org/10.1016/j.clsr.2024.106020 (Accessed: 28 January 2025).

Mantelero, A. (2022) 'Fundamental rights impact assessments in the DSA', *Verfassungsblog*, November [Online]. Available at: https://doi.org/10.17176/20221101-220006-0 (Accessed: 28 January 2025).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

'Meta Platforms Inc and Others v Bundeskartellamt' (2023) Judgment of the Court (Grand Chamber) of 4 July 2023. Case C-252/21. [Online]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62021CJ0252 (Accessed: 29 January 2025).

Metz, C. (2023) 'Chatbots may "hallucinate" more often than many realize', *The New York Times*, 6 November [Online]. Available at: https://www.nytimes.com/2023/11/06/technology/chatbots-hallucination-rates.html (Accessed: 28 January 2025).

Mittelstadt, B. (2019) 'Principles alone cannot guarantee ethical AI', *Nature Machine Intelligence*, 1(11), pp. 501–507.

Mogherini, F., Timmermans, F. and Domecq, J. (2016) *Implementation plan on security and defence. Note 14392/16*. Council of the European Union [Online]. Available at: https://www.consilium.europa.eu/media/22460/eugs-implementation-plan-st14392en16.pdf (Accessed: 28 January 2025).

Mühlhoff, R. (2023) 'Predictive privacy: collective data protection in the context of artificial intelligence and big data', *Big Data and Society*, 10(1), [Online]. Available at: https://doi.org/10.1177/20539517231166886. (Accessed: 28 January 2025).

Mühlhoff, R. and Ruschemeier, H. (2024a) 'Predictive analytics and the collective dimensions of data protection', *Law, Innovation and Technology*, 16(1), pp. 261–292.

Mühlhoff, R. and Ruschemeier, H. (2024b) 'Regulating AI with purpose limitation for models', *Journal of AI Law and Regulation*, 1(1), pp. 24–39.

Mühlhoff, R. and Ruschemeier, H. (2024c) 'Updating purpose limitation for AI: a normative approach from law and philosophy'. *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.4711621 (Accessed: 28 January 2025).

Mühlhoff, R. and Ruschemeier, H. (2024d). 'KI-Regulierung durch Zweckbindung für Modelle', ZfDR (4), pp. 337-364.

Murray, T., Cheong, M. and Paterson, J. (2023, July 10) *The flawed algorithm at the heart of Robodebt*. Pursuit [Online]. Available at: https://pursuit.unimelb.edu.au/articles/the-flawed-algorithm-at-the-heart-of-robodebt (Accessed: 28 January 2025).

Neuwirth, R.J. (2023) 'Prohibited artificial intelligence practices in the proposed EU Artificial Intelligence Act (AIA)', *Computer Law and Security Review*, 48(April), [Online]. Available at: https://doi.org/10.1016/j.clsr.2023.105798 (Accessed: 28 January 2025).

Orwat, C., Bareis, J., Folberth, A., Jahnel, J. and Wadephul, C. (2024). Normative challenges of risk regulation of artificial intelligence. *NanoEthics*, *18*(11) [Online]. Available at: https://doi.org/10.1007/s11569-024-00454-9 (Accessed: 28 January 2025).

OECD (2019) *AI principles overview*. OECD [Online]. Available at: https://oecd.ai/en/principles (Accessed: 28 January 2025).

Pathak, G. (2022) 'Manifestly made public: Clearview and GDPR', *European Data Protection Law Review (EDPL)*, 8(3), pp. 419-422.

Paul, R. (2021) *Varieties of risk analysis in public administrations: Problem-solving and polity policies in Europe*. New York: Routledge.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Paul, R. (2023) 'European artificial intelligence "trusted throughout the world": risk-based regulation and the fashioning of a competitive common AI market', *Regulation and Governance*, 18(4), pp. 1065–1082.

'Public.Resource.Org, Inc. and Right to Know CLG v European Commission' (2024) Judgment of the Court (Grand Chamber) of 5 March 2024. Case C-588/21 P. [Online]. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62021CJ0588 (Accessed: 29 January 2025).

Qian, I., Xiao, M., Mozur, P. and Cardia, A. (2022) 'Four takeaways from a *Times* investigation into China's expanding surveillance state', *The New York Times*, 21 June [Online]. Available at: https://www.nytimes.com/2022/06/21/world/asia/china-surveillance-investigation.html (Accessed: 28 January 2025).

'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)' (2022) *Official Journal* L 277, 27 October [Online]. Available at: http://data.europa.eu/eli/reg/2022/2065/oj (Accessed: 29 January 2025).

'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)' (2024) *Official Journal* L, 2024/1689, 12 July [Online]. Available at: http://data.europa.eu/eli/reg/2024/1689/oj (Accessed: 29 January 2025).

Rezende, I.N. (2020) 'Facial recognition in police hands: assessing the "Clearview Case" from a European perspective', *New Journal of European Criminal Law*, 11(3), pp. 375–389.

Ridgway, R. (2023) 'Deleterious consequences: how Google's original sociotechnical affordances ultimately shaped "trusted users" in surveillance capitalism', *Big Data and Society*, 10(1) [Online]. Available at: https://doi.org/10.1177/20539517231171058 (Accessed: 28 January 2025).

Rodríguez Codesal, P. (2024) '*"De-risking". Approach to the concept and study of the possible consequences of a strategically autonomous policy of the European Union*. Repositoria Comillas [Online]. Available at: https://repositorio.comillas.edu/xmlui/handle/11531/79540 (Accessed: 28 January 2025).

Ruschemeier, H. (2022) 'Privacy als Paradox? Rechtliche Implikationen verhaltenspsychologischer Erkenntnisse' in Friedewald, M. et al. (eds), *Künstliche Intelligenz, Demokratie und Privatheit*. Baden-Baden: Nomos, pp. 211–238.

Ruschemeier, H. (2023a) 'AI as a challenge for legal regulation – the scope of application of the Artificial Intelligence Act proposal', *ERA Forum*, 23(3), pp. 361–376.

Ruschemeier, H. (2023b) *Regulierung von KI*. Bundeszentrale für politische Bildung [Online]. Available at: https://www.bpb.de/shop/zeitschriften/apuz/kuenstliche-intelligenz-2023/541498/regulierung-von-ki/ (Accessed: 28 January 2025).

Ruschemeier, H. (2023d) 'Squaring the circle: ChatGPT and data protection', *Verfassungsblog*, 7 April [Online]. Available at: https://verfassungsblog.de/squaring-the-circle/ (Accessed: 28 January 2025).

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Ruschemeier, H. (2023e) 'The problems of the automation bias in the public sector: a legal perspective', *Weizenbaum Conference Proceedings 2023. AI, Big Data, Social Media, and People on the Move*, Weizenbaum Institute, pp. 59–69 [Online]. Available at: https://doi.org/10.34669/wi.cp/5.6 (Accessed: 28 January 2025).

Ruschemeier, H. (2024a) 'Generative AI and data protection'. *SSRN* [Online]. Available at: https://papers.ssrn.com/abstract=4814999 (Accessed: 28 January 2025).

Ruschemeier, H. (2024b) 'Prediction power as a challenge for the rule of law'. *SSRN* [Online]. Available at: https://ssrn.com/abstract=4888087 (Accessed: 28 January 2025).

Ruschemeier, H. (2024c) 'Thinking Outside the Box?' in Steffen, B. (ed) *Bridging the Gap Between AI and Reality. First International Conference, AISoLA 2023, Crete, Greece, October 23–28, 2023, Proceedings*. Cham: Springer, pp. 318-332.

Ruschemeier, H. and Hondrich, L. (2024) 'Automation bias in public administration – an interdisciplinary perspective from law and psychology'. *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.4736646 (Accessed: 29 January 2025).

Ruschemeier, H. and Mühlhoff, R. (2023) 'Daten, Werte Und Der AI Act: Warum Wir Mehr Ethik Für Bessere KI-Regulierung Brauchen', *Verfassungsblog*, 15 December [Online]. Available at: https://verfassungsblog.de/daten-werte-und-der-ai-act/ (Accessed: 29 January 2025).

Sandin, P. (1999) 'Dimensions of the precautionary principle', *Human and Ecological Risk Assessment: An International Journal*, 5(5), pp. 889–907.

Schyns, C. (2023) *The lobbying ghost in the machine: Big Tech's covert defanging of Europe's AI Act*. Brussels: Corporate Europe Observatory.

Siegmann, C. and Anderljung, M. (2022) *The Brussels effect and artificial intelligence: How EU regulation will impact the global AI market*. Centre for the Governance of AI [Online]. Available at: http://arxiv.org/abs/2208.12645 (Accessed: 29 January 2025).

Smuha, N.A. (2021) 'From a "race to AI" to a "race to AI regulation": regulatory competition for artificial intelligence', *Law, Innovation and Technology*, 13(1), pp. 57–84.

Smuha, N.A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R. and Yeung, K. (2021) 'How the EU can achieve legally trustworthy AI: a response to the European Commission's proposal for an Artificial Intelligence Act', *SSRN* [Online]. Available at: https://doi.org/10.2139/ssrn.3899991 (Accessed: 29 January 2025).

Stahl, B.C., Rodrigues, R., Santiago, N. and Macnish, K. (2022) 'A European agency for artificial intelligence: Protecting fundamental rights and ethical values', *Computer Law and Security Review*, 45(July) 105661 [Online]. Available at: https://doi.org/10.1016/j.clsr.2022.105661 (Accessed: 29 January 2025).

Suchman, L. (2023) 'The uncontroversial "thingness" of AI', *Big Data and Society*, 10(2) [Online]. Available at: https://doi.org/10.1177/20539517231206794 (Accessed: 29 January 2025).

Summers, R.S. (1998) 'Principles of the rule of law', *Notre Dame Law Review*, 74, pp. 1691–1712.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

'Treaty on the Functioning of the European Union' (2012) *Official Journal* C 326, 26 October, pp. 47-390 [Online]. Available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:12012E/TXT:en:PDF (Accessed: 29 January 2025).

UK Government (2021) *How to score attributes*. Gov.uk [Online]. Available at: https://www.gov.uk/government/publications/attributes-in-the-uk-digital-identity-and-attributes-trust-framework/how-to-score-attributes (Accessed: 29 January 2025).

United Nations (2023) *Interim report: governing AI for humanity*. United Nations [Online]. Available at: https://www.un.org/en/ai-advisory-body (Accessed: 29 January 2025).

Veale, M. and Borgesius, F.Z. (2021) 'Demystifying the draft EU Artificial Intelligence Act – analysing the good, the bad, and the unclear elements of the proposed approach', *Computer Law Review International*, 22(4), pp. 97–112.

Veale, M., Matus, K. and Gorwa, R. (2023) 'AI and global governance: modalities, rationales, tensions', *Annual Review of Law and Social Science*, 19(October), pp. 255–275.

Van der Vlist, F., Helmond, A. and Ferrari, F. (2024) 'Big AI: cloud infrastructure dependence and the industrialisation of artificial intelligence', *Big Data and Society*, 11(1), pp. 1–16.

Wachter, S. (2023) 'The theory of artificial immutability: protecting algorithmic groups under anti-discrimination law', *Tulane Law Review*, 97(2) [Online]. Available at: https://www.tulanelawreview.org/pub/artificial-immutability (Accessed: 29 January 2025).

Weber, J. and Suchman, L. (2016) 'Human–machine autonomies' in Bhuta, N., Beck, S., Geiβ, R., Liu, H.-Y. and Kreβ, C. (eds.) *Autonomous weapons systems*. Cambridge: Cambridge University Press, pp. 75–102.

Wendehorst, C. and Duller, Y. (2021) 'Biometric recognition and behavioral detection'. *SSRN* [Online]. Available at: https://papers.ssrn.com/abstract=4087455 (Accessed: 29 January 2025).

Whittaker, M. (2021) 'The steep cost of capture', *Interactions*, 28(6), pp. 50–55.

Wong, P.-H. (2020) 'Democratizing algorithmic fairness', *Philosophy and Technology*, 33(2), pp. 225–244.

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

https://doi.org/10.5771/9783748943990-41 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY