CHAPTER 6

Regulating Artificial Intelligence in the European Union

Mireille M. Caruana and Roxanne Meilak Borg

Abstract

The EU AI Act marks a significant milestone as the world's first comprehensive legal framework for Artificial Intelligence. This pioneering legislation aims to regulate AI systems' development, deployment, and use to foster innovation while ensuring safety, transparency, and adherence to fundamental rights. The chapter will analyse the AI Act in light of the underlying goal of 'regulation for innovation', which seeks to enhance the adoption of AI within the Internal Market by bolstering trust in technological advancements. The narrative provides an overview of the AI Act, outlining its framework and key components. It delves into the Act's adoption of a risk-based regulatory approach, which precludes AI systems that pose an unacceptable risk, heavily regulates AI systems that pose high risks, and lightly regulates systems posing limited risks. The chapter further explores other essential aspects of the Act, including regulating general-purpose AI models, requirements and obligations like conducting Fundamental Rights Impact Assessments, the role of standards and benchmarks, regulatory sandboxes, and the functions of the European AI Office and AI Board. These elements underscore the Act's potential to harmonise AI regulation within the Internal Market and influence the future landscape of AI in Europe and beyond. The discussion refers to the achievements and challenges faced by the EU legislator in creating a unified regulatory environment in light of the potential impacts on the Digital Single Market and the EU's competitiveness on the global stage. It assesses how the Act navigates the balance between fostering innovation and safeguarding fundamental rights. This chapter offers an overview of the AI Act, focusing on pivotal aspects of shaping the Internal Market and influencing global AI regulation frameworks. It also discusses the protection of fundamental rights within these regulatory structures.

1 Introduction

The uptake of AI systems has the potential to bring several individual and societal benefits and foster innovation and global competitiveness. While the

 | DOI:10.1163/9789004712119_008

industry does a great job promoting these benefits, we focus on AI's risks and potential harms in this chapter. AI systems are posed to generate various significant challenges and negative consequences, particularly through their impact on the democratic fabric of society, fundamental (or 'human') rights and other significant risks, including safety hazards when integrated into products and services. Thus, unregulated AI deployment may have irreversible negative consequences for humans and humanity; philosophers, lawyers, and others must discuss and debate the immensely important subject of regulating AI.

This chapter considers conversations about regulating AI, focusing on the legislative initiatives that have been, or are in the process of being, undertaken within the European Union ('EU').[1] In April 2021, the European Commission published a proposal for an EU regulatory framework on Artificial Intelligence. It was intended as future-proof legislation with flexible mechanisms that would allow it to adapt and evolve with new regulatory challenges and concerns. This proposal represents the first-ever attempt to enact a horizontal regulation of AI within the EU. The Council of the EU adopted its common position on such proposed legislation in December 2022,[2] and the European Parliament ('EP') in June 2023.[3] In December 2023, the Council and the EP reached a final political agreement on the text of the AI Act, which the EP has now passed.[4]

1 The Council of Europe recently adopted the first-ever international treaty on artificial intelligence, which will not, however, be considered in this chapter. For more information, see: <https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence> accessed 20 May 2024.

2 The Council adopted its common position ('Proposed AI Act (Council General Approach') on the AI Act on 6 December 2022: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – EU Council General approach - Brussels, 25 November 2022 <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf> accessed 20 May 2024. See also Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights, Press release 6 December 2022, <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/> accessed 20 May 2024.

3 Amendments adopted by the EP on 14 June 2023 on the proposal for a regulation of the EP and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf> accessed 17 May 2024.

4 The text of the final draft (European Parliament 'Corrigendum' of 16th April 2024) is available here: <https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf> accessed 17 May 2024. The Regulation is expected to be finally adopted before the end of the legislature and also needs to be formally endorsed by the Council.

The stated purpose of the Act is to improve internal market functioning through a uniform legal framework for the development, placing on the market, putting into service and use of AI systems in the Union to promote the uptake of human-centric and trustworthy AI while ensuring a high level of protection of health, safety, fundamental rights as enshrined in the EU Charter of Fundamental Rights, to protect against the harmful effects of AI systems in the Union, and to support innovation. The Act ensures the free movement, cross-border, of AI-based goods and services. It prevents Member States from imposing restrictions on the development, marketing and use of AI systems unless explicitly authorised by it.[5]

The AI Act thus has a dual concern and relative function: free trade/elimination of trade barriers and individual and societal impact/protection of health, safety and fundamental rights. It is based on Article 114 of the Treaty on the Functioning of the European Union ('TFEU'), the 'internal market clause'. However, to the extent that it contains specific rules on the protection of individuals about the processing of personal data concerning restrictions of the use of AI systems for (i) remote biometric identification for law enforcement, (ii) risk assessments of natural persons for law enforcement and (iii) biometric categorisation for law enforcement, and in as far as those specific rules are concerned, the Act is based on Article 16 TFEU (the right to the protection of personal data).[6]

The Act will apply to private and public entities, regardless of their location within or outside the EU, as long as the AI system is available in the Union market or its use affects people in the EU. Among the Act's most important elements are the requirement of transparency and the need for rigorous safety checks before an AI system is released into the public domain. The broad shape of things to come in Europe has been set. However, this is continually being challenged and impacted by scientific and technological developments in the sector that the law sets out to regulate, such as the emergence during the legislative process of heightened debate and controversy surrounding regulating generative AI as a result of the public availability of 'large language models' ('LLMs') like ChatGPT.

This chapter will provide an overview and commentary on the AI Act's regulatory strategies and general structure. An article-by-article analysis of the Act is beyond the scope of this book chapter. Rather, we seek to shine a light on selected fundamental aspects of the legislation. The chapter proceeds as

5 Recital (1) AI Act.
6 Recital (3) AI Act.

follows: it first considers the subject matter that we are discussing regulating ('Artificial Intelligence Systems') insofar as necessary to make sense of the discussion that ensues. It then highlights the 'why' of regulating, or in other words, what fears and legal and/or other factors have driven the EU to regulate rather than allow market forces free rein. It clarifies why regulation is necessary to move on to a discussion of 'how' we should regulate. Here, the focus is on the 'risk-based' approach adopted by the European Union, which calibrates the regulatory burden to the gravity or likelihood of the (anticipated) risk or threat. Laws lay down rights and obligations; Product safety-type laws such as the AI Act mostly lay down obligations. Hence, it is necessary to consider 'who' is subject to these obligations (providers, deployers, importers, distributors, etc.) carefully. The discussion then proceeds to the substantive obligations imposed upon these actors. Given the technical nature of the field, the limits of law are evident. Thus, co-regulation plays a role, particularly for standards, the standardisation bodies that set them, and relative certification. The chapter outlines the governance and enforcement structures that the Act sets up and assesses their likelihood of effectiveness. It also considers the strategy and role of so-called 'AI regulatory sandboxes' and concludes with thoughts on the 'regulatory balance' that is sought between protection, technological development and innovation and the likelihood of the EU achieving this delicate balance.

2 Regulate What? Defining 'AI systems'

The AI Act regulates 'AI systems', defining an AI system as one that is 'machine-based', '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'.[7] The Council's General Approach on the Act narrowed the definition to systems developed through machine learning approaches and logic – and knowledge-based approaches;[8] this position was later abandoned,[9] and a technologically-neutral definition made it to the final draft text.

The debate on LLMs paved the way for the regulation by the Act of 'general-purpose' AI (GPAI) models and 'general-purpose' AI (GPAI) systems, which

7 Article 3(1) AI Act.

8 Article 3(1) Proposed AI Act (Council General Approach).

9 EP DRAFT Compromise Amendments 16 May 2023.

were not covered by the Commission's original proposal. The Act now includes a definition for both these terms, as follows:

> A 'general-purpose AI (GPAI) model' means "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";[10]
>
> A 'general-purpose AI (GPAI) system' means "an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems".[11]

The definition of a GPAI model is broad. It is so broad that it has been argued that it could "be interpreted in such a way as to potentially include a vast variety of technological cases".[12] Typical examples of GPAI models are generative AI models[13] such as LLMs. These models are a tool for processing and generating natural language and have many potential applications in language translation, content generation, and conversational AI. They use deep learning, which involves training a neural network on vast amounts of textual data, such as books, articles, and web pages. An LLM aims to learn the patterns and relationships between words and phrases in natural language to generate coherent and meaningful text. This involves analysing the syntax and semantics of language and understanding the context in which words are used. ChatGPT-4, developed by OpenAI and released on the public web in March 2023, is one of the most well-known examples of LLMs.

LLMs can be considered a broader category encompassing foundation models. This type serves as a starting point for building other, more specialised models and constitutes an example of a general-purpose AI system. The term

10 Article 3(63) AI Act.

11 Article 3(66) AI Act.

12 Innocenzi Genna, 'Foundation Models: how they are regulated in the AI ACT' (*radiobruxelleslibera*, 5 February 2024) <https://radiobruxelleslibera.com/2024/02/05/foundation-models-how-they-are-regulated-in-the-ai-act/#:~:text=Foundation%20Models%20are%20now%20defined,components%20into%20an%20AI%20system> accessed 20 May 2024.

13 Recital (99) AI Act.

'foundation model' refers to the fact that these models provide a foundation of knowledge and understanding of language that can be used to build other, more specialised models for specific natural language processing ('NLP') tasks. For example, a foundation model could be fine-tuned on a smaller text dataset for a specific task, such as sentiment analysis or question answering, and then used to make predictions or generate text in that domain. The European legislator has chosen not to limit the provisions of the Act to 'foundation models'; these fall within the broader notions of LLMs and GPAI models.

Emily M. Bender has coined the term 'stochastic parrots' for LLMs,[14] which is now commonly used when referring to such models because, while impressive in their ability to generate natural language, they ultimately do not understand the meaning of the language they are processing (they just 'parrot'). LLMs are incapable of true reasoning or understanding; they merely rely on statistical patterns in data to generate responses. As a result, they are prone to errors and biases, as they can perpetuate stereotypes and other problematic patterns in language. Furthermore, they are not always transparent about how they arrive at their responses.[15] The output of "hallucinations", or made-up answers, has indeed triggered a GDPR-based complaint by NOYB against OpenAI (the company that has developed ChatGPT) on account of the output of false information about people and the inability or unwillingness of OpenAI to comply with the data access request and in so doing be transparent about their training data.[16]

3 Why Regulate? Threats to Fundamental Rights and Values

While there is no doubt about the promise of AI to make things better, the drive for regulation stems from certain fears relating to the development and deployment of AI systems, combined with a perceived insufficiency of

14 Emily M. Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?' (2021) Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency 610 <https://doi.org/10.1145/3442188.3445922> accessed 20 May 2024.

15 Explained simply here: Muhammad Saad Uddin, 'Stochastic Parrots: A Novel Look at Large Language Models and Their Limitations' (*Towards AI*, 20 April 2023) <https://towardsai.net/p/machine-learning/stochastic-parrots-a-novel-look-at-large-language-models-and-their-limitations#:~:text=At%20its%20core%2C%20the%20term,the%20language%20they%20are%20processing> accessed 20 May 2024.

16 NOYB, 'ChatGPT provides false information about people, and OpenAI can't correct it' (*noyb,* 29 April 2024) <https://noyb.eu/en/chatgpt-provides-false-information-about-people-and-openai-cant-correct-it> accessed 20 May 2024.

current laws and ethical frameworks to protect against these feared harmful or otherwise negative consequences for individuals and/or for society. At the most extreme, concerns have been expressed that AI threatens to enslave or even exterminate us[17] or that future developments will mean that 'enslaving' AI will amount to nothing less than another form of modern-day slavery.[18]

A more nuanced viewpoint (embraced by these authors) holds that AI, a creation of humans, will be controlled by (some) humans, but its deployment may result in the enslavement of others. However, since (some) humans control its development, AI could be developed to protect fundamental or human rights and values 'by design'.

The AI Act refers to the "risk of harm to the health and safety, or an adverse impact on fundamental rights".[19] As expressed succinctly by the RAILS,

> AI systems have the potential to unpredictably harm people's life, health, and property. They can also affect fundamental values on which western societies are founded, leading to breaches of fundamental rights of people, including the rights to human dignity and self-determination, privacy and personal data protection, freedom of expression and assembly, non-discrimination, or the right to an effective judicial remedy and a fair trial, as well as consumer protection.[20]

Indeed, certain AI systems may be deemed to pose such a serious risk that they should not merely be regulated but banned outright. For example, the advocacy advisor on AI Regulation at Amnesty International has expressed the opinion that:

> The EU must ban the use of discriminatory AI systems which disproportionately affect people from marginalised communities, including

17 As reflected in science fiction in the Terminator series of films or in Isaac Asimov's I, Robot. See also: Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (OUP 2014); Chris Vallance, 'Artificial Intelligence could lead to extinction, experts warn' (*BBC News,* 31 May 2023) <https://www.bbc.com/news/uk-65746524> accessed 20 May 2024.

18 Andrew Murray, 'When machines become sentient, we will have to consider them an intelligent life form' (*LSE*, 10 August 2018) <https://blogs.lse.ac.uk/businessreview/2018/08/10/when-machines-become-sentient-we-will-have-to-consider-them-an-intelligent-life-form/> accessed 20 May 2024.

19 Article 7(1)(b) AI Act.

20 Martin Ebers and others, 'The European Commission's Proposal For An Artificial Intelligence Act—A Critical Assessment By Members Of The Robotics And AI Law Society (RAILS)' (2021) 4(4) Multidisciplinary Scientific Journal (2021) <https://www.mdpi.com/2571-8800/4/4/43> accessed 5 May 2023.

> migrants, refugees and asylum seekers. Such technologies profile people and communities, claiming to 'predict' crimes or 'identify' people who supposedly pose a security risk, even leading to them being denied the right to asylum. EU lawmakers must not miss this opportunity to prohibit the use of certain AI-based practices and protect the rights of migrants, refugees, and asylum seekers against harmful aspects of AI.
>
> Use of mass surveillance technologies, such as retrospective and live remote biometric identification tools must also be banned. The proposed law must also ban discriminatory social scoring systems that prevent people from accessing essential public and private services, such as child support benefits and education.[21]

In this vein, the AI Act does prohibit certain AI practices. It considers, among others, that "AI systems providing social scoring of natural persons by public or private actors may lead to discriminatory outcomes and the exclusion of certain groups. …"[22]

AI is broad enough to include environments many of us already interact with daily, such as technology-powered search engines or social networking sites. Manipulative and/or deceptive practices may significantly impact individuals and societies; for example, digital nudging techniques exploit human vulnerabilities and the methods used by social media to make them addictive and keep users glued to the screen for as long as possible.[23]

Apart from the addictive element, the potential manipulation of online platforms ('intentional manipulation of their service, including by inauthentic use or automated exploitation of the service, as well as the amplification and potentially rapid and wide dissemination of illegal content and of information that is incompatible with their terms and conditions, to use the language of

21 Mher Hakobyan, Advocacy Advisor on Artificial Intelligence Regulation at Amnesty International, quoted in 'EU: European Union must protect human rights in upcoming AI Act vote' 26 April 2023 <https://www.commondreams.org/newswire/eu-european-union-must-protect-human-rights-in-upcoming-ai-act-vote> accessed 27 April 2023.

22 Recital (31) AI Act.

23 Tina van der Linden, 'Regulating Artificial Intelligence: Please Apply Existing Regulation' (Amsterdam Law Forum 2021) <https://www.amsterdamlawforum.org/articles/abstract/432>. Action against addictive design is part of formal proceedings opened by the Commission against TikTok in February 2024: European Commission, 'Commission opens formal proceedings against TikTok under the Digital Services Act' (*European Commission,* 19 February 2024) <https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926> accessed 15 May 2024.

the Digital Services Act[24]), coupled with the practice of behaviourally targeted advertising, has various facets, not least their potential to influence elections and thus threaten an essential aspect of our democracies. A sustained focus on the detrimental impact of fake news and disinformation, coupled with harnessing the strengths of technology to target 'news' or political advertising in a particular way, started around the time of the Cambridge Analytica scandal, the British consulting firm that collected personal data belonging to millions of Facebook users without their consent, predominantly to be used for political advertising, and that was allegedly determinative of the election of Donald Trump as President of the United States, and the 'Brexit' referendum vote for the United Kingdom to leave the European Union.

Beyond the spread of disinformation, 'filter bubbles' – an algorithmic bias that skews or limits the information an individual user sees on the Internet – are also problematic for democratic discourse, as it means that an individual's online experience is not open to alternative, contrasting viewpoints to those already held.[25]

Thus, search engines and digital platforms such as social media platforms can have a deleterious impact at both the individual and the societal levels, threatening individual autonomy and dignity, as well as the democratic fabric of society and the rule of law. These developments in our online life experiences directly result from the profit-making target of the big businesses behind them.

It is important to protect fundamental rights in all situations. It has been argued that fundamental human rights can or should not be treated and evaluated equally with business (that is, profit) interests.[26] However, this is not to say that the freedom to conduct a business should not be considered.[27] The matter with AI is that, like other tools, it can be used for good or for bad. It can

24 Article 34(2) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) OJ L 277/1

25 Eli Pariser, *The filter bubble: How the new personalized web is changing what we read and how we think* (Penguin, 2011); Paula Johanson, *Online Filter Bubbles* (Greenhaven Publishing, 2017).

26 Fanny Hidvegi, Daniel Leafier and Estelle Massey, 'The EU should regulate AI on the basis of rights, not risks (*Access Now*, 17 February 2021) <https://www.accessnow.org/eu-regulation-ai-risk-based-approach/> accessed 22 May 2023.

27 Article 16 EU Charter; Peter Craddock Op-Ed: Who dares question the primacy of data protection? 7 May 2024 <https://www.linkedin.com/pulse/op-ed-who-dares-question-primacy-data-protection-peter-craddock-kfkxe/?trackingId=xYW%2BEJ2RRjyEbnNgn8Lfsw%3D%3D> accessed 17 May 2024.

play a role in tackling environmental challenges,[28] but it can also contribute to them;[29] it can be used to protect and uphold fundamental rights, including human dignity, or to undermine them.[30]

4 How to Regulate? A 'Risk-Based' Approach

The AI Act adopts a 'risk-based' approach and lays down a 'horizontal' (as opposed to sector-specific) legal framework for AI that aims to ensure legal certainty; this is desirable as regulating all AI systems in an equally heavy-handed manner would arguably suppress innovation and merely help to strengthen the already dominant market players.[31] However, the difficulty in applying such an approach lies in deciding how or according to the criteria 'high-risk' to be determined. It has been suggested that initial criteria must be developed for assessing what is 'high risk' – thus determining whether an obligation to proceed to a full impact assessment is triggered.[32]

The Act does not, however, include any such criteria. It merely precludes AI systems that pose an unacceptable risk, listing several prohibited practices. Thus, it identifies and heavily regulates AI systems with high risks and briefly mentions systems with limited risks. It also includes a new Chapter V on general-purpose AI ('GPAI') models, which speaks of GPAI models with

28 See for example UNEP, 'How Artificial Intelligence is helping tackle environmental challenges' (*UNEP*, 7 November 2022) <https://www.unep.org/news-and-stories/story/how-artificial-intelligence-helping-tackle-environmental-challenges> accessed 29 May 2023.

29 See for example Annette Ekin, 'AI can help us fight climate change. But it has an energy problem, too' (*Horizon,* 12 September 2019) <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/ai-can-help-us-fight-climate-change-it-has-energy-problem-too> accessed 29 May 2023.

30 European Union Agency for Fundamental Rights, Getting the Future Right: Artificial Intelligence and Fundamental Rights, 2020 <https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf> accessed 29 May 2023.

31 Cf. Thomas Höppner and Luke Streatfeild, 'ChatGPT, Bard & Co.: An Introduction to AI for Competition and Regulatory Lawyers' (2023) 9 Hausfeld Competition Bulletin (1/2023), Article 1, Available at SSRN <https://ssrn.com/abstract=4371681> or <http://dx.doi.org/10.2139/ssrn.4371681> Accessed 5 June 2023.

32 Lilian Edwards, Expert Opinion: Regulating AI in Europe – Four problems and four solutions, 31 March 2022 Ada Lovelace Institute <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf> accessed 31 May 2023.

'systemic risk' and ones "released under a free and open-source licence".[33] Each category mentioned in this paragraph is discussed in further detail below.

4.1 *Prohibited AI Practices*

The AI Act prohibits eight practices deemed to present unacceptable risks. These relate to manipulation, social scoring, the assessment of the risk of an individual committing a criminal offence, the compilation of facial recognition databases through untargeted data scraping, the inference of emotions in workplaces or education institutions, biometric categorisation systems and 'real-time' remote biometric identification ('RBI').[34]

Thus, the Act prohibits the placing on the market, putting into service, or use of an AI system:

1. That "deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm";[35]
2. That "exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm";[36]
3. For "the evaluation or classification of natural persons or groups of persons over a certain period based on their social behaviour or known, inferred or predicted personal or personality characteristics", with the social score leading to 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, and/or in social contexts that are unrelated to the contexts in which the data was originally generated or collected;[37]

33 Article 53(2) AI Act.
34 Article 5 AI Act.
35 Article 5(1)(a) AI Act.
36 Article 5(1)(b) AI Act.
37 Article 5(1)(c) AI Act.

4. For "making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics",[38] although this prohibition does not apply to 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";[39]
5. That "create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage";[40]
6. To "infer emotions of a natural person in the areas of workplace and education 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";[41] and
7. That "categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or political beliefs, sex life or sexual orientation"[42] ('biometric categorisation systems'); this prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorising of biometric data in the area of law enforcement.

The Act also prohibits the use of 'real-time' RBI systems in publicly accessible spaces for law enforcement unless and in so far as such use is strictly necessary for certain objectives as further specified by it.[43] In such latter cases, RBI systems must be deployed "only to confirm the identity of the specifically targeted individual", "be authorised by national law", and "comply with necessary and proportionate safeguards and conditions" about their use by such national law. Furthermore, the relevant law enforcement authority must have completed a fundamental rights impact assessment concerning the RBI system and registered it in the EU database.[44]

Each use of an RBI system for law enforcement requires a prior authorisation granted by a judicial authority or by an independent administrative authority "whose decision is binding of the Member State in which the use is to take

38 Article 5(1)(d) AI Act.
39 Ibid.
40 Article 5(1)(e) AI Act.
41 Article 5(1)(f) AI Act.
42 Article 5(1)(g) AI Act.
43 Article 5(1)(h)(i) – (iii) AI Act.
44 Article 5(2) AI Act.

place".[45] Such use must also be "notified to the relevant market surveillance authority ('MSA') and the national data protection authority",[46] which shall submit 'annual reports' on such use to the Commission.[47] The Commission's original proposal listed four prohibited AI practices (points 1, 2, 3 and 8 above). The other four practices were added further to the trilogue negotiation proceedings. As is evident from the law's wording, the prohibitions in points 1, 2, 3 and 5 are absolute, while some exceptions are permissible from those described in points 4, 6, 7 and 8 in the prescribed circumstances.

Apart from including some new prohibited practices, some changes have also been made to the wording of the original four provisions. For instance, the provision prohibiting the use of AI systems that exploit individuals' vulnerabilities has been extended to cover persons vulnerable due to their social or economic situation, where it previously only covered "age, physical or mental disability", and the prohibition of using AI for social scoring now also covers private actors, where it previously only applied to public authorities. Likely, the provision prohibiting the compilation of facial recognition databases through the untargeted scraping of online and CCTV footage was introduced due to the heightened awareness resulting from the Clearview AI and PimEyes cases.[48] The final draft text also saw an increase in the requirements to be fulfilled for the lawful use of RBI systems in publicly accessible spaces for law enforcement purposes, including having a national law authorising such use, conducting a fundamental rights impact assessment, registering such systems in an EU database, and notifying their use to additional competent authorities such as data protection authorities. This more stringent approach attempts to strike a balance between the exigencies of the law enforcement sector and the resistance of civil society, witnessed through the public consultation process,[49] as

45 Article 5(3) AI Act.

46 Article 5(4) AI Act.

47 Article 5(6) AI Act.

48 See for instance, Kashmir Hill, 'A Face Search Engine Anyone Can Use is Alarmingly Accurate' *The New York Times* (New York, 26 May 2022) <https://www.nytimes.com/2022/05/26/technology/pimeyes-facial-recognition-search.html> accessed 20 May 2024.

49 See for instance Amnesty International 'Europe: Proposed legislation too weak to protect us from dangerous AI systems' (21 April 2021) <https://www.amnesty.org/en/latest/press-release/2021/04/eu-legislation-to-ban-dangerous-ai-may-not-stop-law-enforcement-abuse/> accessed 20 May 2024 and SHERPA, 'Feedback to the European Commission on its Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence' (6 August 2020) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665582_en> accessed 20 May 2024. All documentation

well as the opinion expressed by the European Data Protection Board ('EDPB') and the European Data Protection Supervisor ('EDPS'), which jointly called for a general ban on any use of AI for automated recognition of human features in publicly accessible spaces – such as of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioural signals – in any context.[50] Finally, it is worth noting that the AI Act prohibits only the 'use' – and not the 'placing on the market' – of 'real-time' RBI systems. Thus, EU vendors can sell biometric systems that would be illegal to use in the EU for regimes in third countries.[51]

4.2 *High-Risk AI Systems*

'High-risk' AI systems are dealt with in Chapter III of the Act, which provides for the classification of and requirements for such systems, the obligations incumbent on providers and deployers, and standards for such systems. This section focuses on the classification of AI systems as 'high-risk'.

An AI system is classified as high-risk under the Act if: (i) it is itself a product or a safety component of a product covered by certain Union health and safety harmonisation legislation (listed in Annex I, such as toys, machinery, lifts or medical devices) and it is required to undergo a third-party conformity assessment in order to be placed on the market or put into service according to such legislation[52] or (ii) it falls within a category listed in Annex III.[53] The areas set out in this Annex are the following: biometrics; critical infrastructure; education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and essential public services and benefits; law enforcement; migration, asylum and border control management; administration of justice and democratic processes.

---

submitted to the Commission pursuant to the public consultation period are available at <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback_en?p_id=24212003> accessed 20 May 2024.

50 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) 18 June 2021.

51 Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22(4) Computer Law Review International 101.

52 Article 6(1) AI Act.

53 Article 6(2) AI Act.

Notably, an AI system that falls within any of these categories but "does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision-making" because it is intended to "perform a narrow procedural task", "improve the result of a previously-completed human activity", "detect decision-making patterns or deviations from prior decision-making patterns and not meant to replace or influence the previously completed human assessment, without proper human review" or "perform a preparatory task to an assessment relevant for the use cases listed in Annex III" is not considered to constitute a high-risk system.[54] An embryonic version of this provision exempting AI systems that are unlikely to cause significant risk to the health, safety or fundamental rights was first introduced by the Council in its General Approach.[55] On the other hand, AI systems that perform profiling of natural persons shall "always be considered to be high-risk".[56]

The Commission can amend the list in Annex III through delegated acts. The Commission may include sub-areas within the existing ones if the AI system poses risks similar to those of an existing in-scope AI system, but cannot add entirely new areas.[57] Under a provision introduced by the Council's General Approach,[58] the Commission may remove high-risk AI systems from the list, provided such systems no longer pose significant risks to fundamental rights, health or safety.[59]

4.3 *Limited Risk AI Systems*

'Limited risk' AI systems are subject to transparency obligations in certain circumstances for providers and users of certain AI systems (chatbots, biometric categorisation, emotion recognition system, 'deep fakes'), so users know they are interacting with a machine.[60] For example, Article 50 *inter alia* provides that:

> Providers shall ensure that AI systems intended to interact directly with natural persons are designed and developed in such a way that the natural persons concerned are informed that they are interacting with an AI

54 Article 6(3) AI Act.
55 Article 6(3) Proposed AI Act (Council General Approach).
56 Article 6(3) AI Act.
57 Article 7 AI Act.
58 Article 7(3) Proposed AI Act (Council General Approach).
59 Article 7(3) AI Act.
60 Article 50 AI Act.

> system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use. (...)[61]

4.4 *Minimal Risk AI Systems*

The AI Act allows other types of applications to be legally developed. This could include applications such as AI-enabled video games or spam filters. Article 95, on "codes of conduct for voluntary application of specific requirements", is the only relevant provision.

4.5 *General-Purpose AI Models*

General-purpose AI (GPAI) models are regulated in Chapter V of the Act, distinguishing between generic GPAI models, GPAI models "with systemic risk", and "free and open-source licence" GPAI models.

A GPAI model is classified as one with systemic risk[62] if (a) it has high-impact capabilities evaluated based on appropriate technical tools and methodologies, including indicators and benchmarks, *or* (b) based on a decision of the Commission, *ex officio* or following a qualified alert from the scientific panel, it has capabilities or an impact equivalent to those set out in point (a) having regard to the criteria set out in Annex XIII of the Act.[63]

GPAI models with systemic risk classified as such under (a) above must be notified to the Commission.[64] If the Commission becomes aware of a GPAI model with systemic risk of which it has not been notified, it may designate it as such.[65] It is also possible for the provider of a GPAI model that meets this condition to argue that, exceptionally, such a model does not, in fact, due to its specific characteristics, present systemic risks and, therefore, should not be classified as such. However, the Commission may reject these arguments.[66] A designation decision may be changed upon a reasoned request by a provider and if accepted by the Commission.[67] A list of GPAI models with systemic risk must be published and updated by the Commission (without prejudice

61 Article 50(1) AI Act.
62 Defined in Art. 3(65) AI Act.
63 Article 51(1) AI Act.
64 Article 52 AI Act.
65 Ibid.
66 Article 52(3) AI Act.
67 Article 52(5) AI Act.

to intellectual property rights and confidential business information or trade secrets).[68]

These provisions represent a change from the Council's position, where GPAI systems to be used as high-risk AI systems or components of such systems would have had to comply with the requirements for high-risk AI systems.[69] The final version of the Act has introduced the notion of GPAI models with systemic risk, and GPAI models must comply with the new Chapter V and all other relevant obligations within the Act.

5 Regulate Who? Providers v Deployers

When regulating Artificial Intelligence, a central question is who should be subjected to legal requirements, obligations and potential future liability. The developer – or, in the language of the AI Act, 'provider' – is the entity that will have the knowledge and practical control of the training sets, algorithms, etc. It may be argued that developers should carry legal responsibility, as the high-risk obligations mostly arise at the development stage. Such an argument is strengthened when the developer significantly profits from licensing their product or service. However, a counter-argument about GPAI notes that provider responsibility (and potential subsequent liability) may not include responsibility/liability for unforeseeable uses/risks. Nevertheless, it may not be practical (or even fair) to impose certain obligations on deployers because it may be impossible for them to fix or even audit issues of data quality, etc without access to the upstream source code, training datasets etc (which are often secret/proprietary, as with ChatGPT 4). Previous versions of the Act were also criticised for not reflecting the actors in the AI ecosystem accurately;[70] the final version has clarified the taxonomy of such entities – including, for instance, replacing the previously-employed term of 'user' with 'deployer' – and bestows responsibilities and obligations on various actors.

68 Article 52(6) AI Act.

69 Article 4b Proposed AI Act (Council General Approach).

70 See for instance Norberto Nuno Gomes de Andrade, Laura Galindo and Antonella Zarra, 'Artificial Intelligence Act: A Policy Prototyping Experiment. Revisiting the Taxonomy of AI Actors' (*Open Loop*, April 2023) <https://openloop.org/programs/open-loop-eu-ai-act-program/?utm_source=substack&utm_medium=email> accessed 8 May 2023.

The Act regulates both public and private actors. It applies to:

a. providers placing on the market or putting into service AI systems or placing on the market general-purpose AI models in the Union, irrespective of whether those providers are established or located within the Union or in a third country;
b. deployers of AI systems that have their place of establishment or are located within the Union;
c. providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the system is used in the Union;
d. importers and distributors of AI systems;
e. product manufacturers placing on the market or putting into service an AI system together with their product and under their name or trademark;
f. authorised representatives of providers, which are not established in the Union;
g. affected persons that are located in the Union.[71]

The entities listed above are defined as follows:

A 'provider' is "a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its name or trademark, whether for payment or free of charge".[72] A 'deployer' is "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 during a personal, non-professional activity".[73] An 'authorised representative' is "a natural or legal person located or established in the Union who has received and accepted a written mandate from a provider of an AI system or a general-purpose AI model to respectively, perform and carry out on its behalf the obligations and procedures established by this Regulation".[74] An 'importer' is "any natural or legal person located or established in the Union that places on the market an AI system that bears the name or trademark of a natural or legal person established in a third country".[75] A 'distributor' is "any natural or legal person in the supply chain, other

71 Article 2(1) AI Act.
72 Article 3(2) AI Act.
73 Article 3(4) AI Act.
74 Article 3(5) AI Act.
75 Article 3(6) AI Act.

than the provider or the importer, that makes an AI system available on the Union market".[76] An 'operator' is "a provider, product manufacturer, deployer, the authorised representative, importer or distributor".[77]

6 Regulate How? (Requirements and Obligations)

This section first describes the requirements for high-risk AI systems and the obligations of providers and deployers of such systems as laid out in Chapter III, Sections 2 and 3 and Chapter IX, Sections 1 and 2 of the AI Act. Then, it moves on to the obligations of providers of GPAIs listed in Chapter V, Sections 2 and 3. Additional specific transparency obligations, incumbent on providers and deployers and set out in Chapter IV of the Act, are discussed in a subsequent section below.

6.1 *Requirements and Obligations for High-Risk AI Systems*

High-risk AI systems require establishing and implementing a 'risk management system', understood as a "continuous iterative process planned and run throughout the lifecycle" of such systems.[78] Where these systems make use of techniques involving the training of AI models with data, the "training, validation and testing data sets" must be "relevant, sufficiently representative, and to the best extent possible, free of errors and complete given the intended purpose" and must have "the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons about whom the high-risk AI system is intended to be used".[79]

In this regard, the AI Act provides a specific exemption from Article 9 of GDPR, which lays down an exhaustive list of legal bases for processing special personal data categories. Article 10(5) of the AI Act thus permits the processing of such data – "subject to appropriate safeguards for the fundamental rights and freedoms of natural persons' and 'to the extent that it is strictly necessary'- to ensure bias detection and correction".[80]

High-risk AI systems must be accompanied by technical documentation containing at least "the elements set out in Annex IV", which prescribes "a general description of the AI system" and a detailed description of the elements of

76 Article 3(7) AI Act.
77 Article 3(8) AI Act.
78 Article 9(1) and (2) AI Act.
79 Article 10(3) AI Act.
80 Article 10(5) AI Act.

the AI system and the process for its development, "detailed information about the monitoring, functioning and control of the AI system", "a description of the appropriateness of the performance metrics for the specific AI system", "a detailed description of the risk management system", "a description of the relevant changes made by the provider to the system through its lifecycle and a list of harmonised standards applied in full or in part … or detailed description of the solutions adopted to meet the requirements, a copy of the EU declaration of conformity and a detailed description of the system in place to evaluate the AI system performance in the post-market phase". These are to be provided to organisations involved in the assessment of compliance of the AI system with the requirements set out in the legislation.

High-risk AI systems must also "technically allow for the automatic recording of events ('logs') over the system's lifetime"[81] and be "designed and developed in such a way … that natural persons can effectively oversee them during the period in which they are in use".[82] The *raison d'être* of this provision on human oversight is stated in the provision itself: "Human oversight shall aim to prevent or minimise the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under the conditions of reasonably foreseeable misuse".[83] It is also reflected in the obligation incumbent on deployers to "assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support"[84] and in the additional specific requirements relating to RBI systems, where no action/decision is to be "taken by the deployer based on the identification resulting from the system unless this has been separately verified and confirmed by at least two natural persons".[85]

Insofar as obligations are concerned, providers must *inter alia* ensure compliance of their systems with the requirements described above, indicate on the system or its packaging/accompanying documentation their name, registered trade name/mark and address at which they may be contacted, have a quality management system in place that ensures compliance with the AI Act and keep logs automatically generated by their systems when under their control.[86] They are bound by rules related to documentation keeping,[87] must

81 Article 12(1) AI Act.
82 Article 14 AI Act.
83 Article 14(2) AI Act.
84 Article 26(2) AI Act.
85 Article 14(5) AI Act.
86 Articles 16(1)(a)-(c) and (e), 17 and 19 AI Act.
87 Article 18 AI Act.

ensure that high-risk AI systems undergo the relevant conformity assessment procedure (referred to in Article 43) prior to their being placed on the market or put into service[88] and must affix the CE marking to such systems to indicate conformity with the Act (per Article 48).[89] This conformity assessment system, discussed in the next section, has been adapted from EU product safety law. Providers who have reason to consider that a high-risk AI system is not in conformity with the AI Act must "immediately take the necessary corrective actions to bring that system into conformity, to withdraw it, to disable it, or to recall it, as appropriate" and inform distributors, deployers, authorised representatives and importers accordingly.[90] When a provider becomes aware that a high-risk system risks an individual's "health, safety, or fundamental rights",[91] they must investigate the cause and inform the competent MSA and, where applicable, the notified body.[92] Providers must cooperate with competent authorities,[93] and where they are established in third countries, they must appoint an authorised representative in the EU by written mandate.[94] Providers must also "establish and document a post-market monitoring system in a manner that is proportionate to the risks of the high-risk AI system"[95] and report any serious incident to the MSA of the Member States where the incident occurred.[96]

Deployers must take "appropriate technical and organisational measures" to ensure they use high-risk AI systems "per the instructions for use".[97] As alluded to above, they must assign human oversight to natural persons[98] and monitor the operation of high-risk AI systems based on the instructions for use.[99] When they have reasons to consider that the use of the instructions may result in the system presenting a risk, they must, without undue delay, inform the provider or distributor and the relevant MSA and suspend the use of the system. They must also inform the provider first, the importer or distributor, and the relevant

88 Article 16(f) AI Act.
89 Article 16(h) AI Act.
90 Article 20(1) AI Act.
91 Article 79(1) AI Act.
92 Article 20(2) AI Act.
93 Article 21 AI Act.
94 Article 22 AI Act.
95 Article 72 AI Act.
96 Article 73 AI Act.
97 Article 26(1) AI Act.
98 Article 26(2) AI Act.
99 Article 26(5) AI Act.

MSA when they have identified a serious incident. If the deployer cannot reach the provider, it must report such an incident to the relevant MSA directly.

Similarly to providers, deployers must keep logs automatically generated by the relevant AI system "to the extent that such logs are under their control".[100] Deployers who are employers must "inform workers' representatives and affected workers that they will be subject to using a high-risk AI system".[101] Deployers of high-risk AI systems for post-remote biometric identification must generally request authorisation from a judicial or administrative authority,[102] and deployers of high-risk systems referred to in Annex III of the Act must inform "natural persons that they are subject to the use of the high-risk AI system".[103] Finally, similarly to providers, deployers must cooperate with the relevant competent authorities.[104]

6.2 *Specific Obligations of Providers of General-Purpose AI Models*

Providers of GPAI models must draw up and keep updated relevant technical documentation, including regarding such models' training and testing process and the results of their evaluation;[105] draw up and make available relevant information and documentation to providers of AI systems who intend to integrate the GPAI model into their systems;[106] put in place a policy to comply with Union law on copyright and related rights;[107] and draw up and make publicly available a sufficiently detailed summary about the content used for training the AI model.[108] Providers of free and open-sourced licence GPAI models are exempt from the first two above-listed obligations; however, providers of GPAI models with systemic risks are not.[109]

Moreover, providers of models with systemic risk are bound by additional obligations to those described above and must also: "perform model evaluation … including conducting and documenting adversarial testing of the model to identify and mitigate systemic risks";[110] "assess and mitigate possible

100 Article 26(6) AI Act.
101 Article 26(7) AI Act.
102 Article 26(10) AI Act.
103 Article 26(11) AI Act
104 Article 26(12) AI Act.
105 Article 53(1)(a) AI Act.
106 Article 53(1)(b) AI Act.
107 Article 53(1)(c) AI Act.
108 Article 53(1)(d) AI Act.
109 Article 53(2) AI Act.
110 Article 55(1)(a) AI Act.

systemic risks at Union level";[111] keep track of, document and report relevant information about serious incidents and possible corrective measures to address them without undue delay to the AI Office and as appropriate national authorities;[112] and ensure an adequate level of cybersecurity protection for the GPAI model with systemic risk and its physical infrastructure.[113]

The Act calls for the drawing up of codes of practice at the Union level relating to GPAI models and corresponding obligations;[114] providers of such models may rely on these codes to demonstrate compliance until harmonised standards are published.[115] Finally, it is also incumbent on providers of GPAI models established in third countries to appoint an authorised representative in the EU.[116]

6.3 *Fundamental Rights Impact Assessment*

The first use of 'impact assessments' concerning data-driven technologies was in the field of privacy and data protection. The terminology 'Privacy Impact Assessment'[117] was later replaced by 'Data Protection Impact Assessment' in the GDPR.[118] The terminology and the intended scope of such impact assessments have been extended to a '*fundamental rights* impact assessment'[119] to consider broader fundamental rights risks of AI systems. The impact assessment criteria would derive from the EU Charter of Fundamental Rights[120] and the European Convention on Human Rights.[121] An 'impact assessment'

111 Article 55(1)(b) AI Act.
112 Article 55(1)(c) AI Act.
113 Article 55(1)(d) AI Act.
114 Article 56 AI Act.
115 Article 55(2) AI Act.
116 Article 54 AI Act.
117 See for example the work of then Information and Privacy Commissioner/Ontario Anne Cavoukian, Privacy Impact Assessment Guidelines for the Ontario Personal Health Information Protection Act (October 2005) <https://www.ipc.on.ca/wp-content/uploads/resources/phipa_pia-e.pdf> accessed 31 May 2023; see also Roger Clarke, 'Privacy impact assessment: Its origins and development' (2009) 25(2) Computer Law & Security Review 123.
118 Article 35 GDPR.
119 Heleen Janssen, Michelle Seng Ah Lee, Jatinder Singh, 'Practical fundamental rights impact assessments' (2022) 30 International Journal of Law and Information Technology 200.
120 Charter of Fundamental Rights and Freedoms of the European Union [2012] OJ C 326/02.
121 Council of Europe Convention for the Protection of Human Rights and Fundamental Freedoms [1953]. On the relationship between these instruments, see Article 52(3) EU Charter: 'In so far as this Charter contains rights which correspond to rights guaranteed

presumes that it is carried out *ex-ante*, i.e. before the AI system is put on the market or released into society. However, *ex-post* or audit-type impact assessments are also essential given the fast-paced development of AI systems, their potential unpredictability, and the consequential significant risks posed to individuals or society.

The AI Act has previously been criticised for lacking a general fundamental rights risk assessment for all AI systems falling within its scope, not just high-risk ones.[122] Moreover, Edwards has criticised the restricted scope of such *ex-ante* assessments, noting, in particular, the lack of systematic concern for impacts on groups, particularly algorithmically constituted groups.[123] What should be legally mandated is a comprehensive *ex-ante* impact assessment, as well as regular *ex-post* audits for algorithmic systems released into society, to identify and mitigate the impact and all potential risks or harms that may result from the deployment of such systems. Moreover, such impact assessment should go beyond the traditional individual human rights approach and consider potential harms to groups and communities.

Today, however, the Act still only requires a fundamental rights impact assessment for high-risk AI systems in specific circumstances. It is only deployers of high-risk systems referred to in Annex III (except those relating to critical infrastructure) that are "bodies governed by public law" or "private entities providing public services", and deployers of high-risk systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score (except AI systems used to detect financial fraud)[124] and for risk assessment and pricing concerning natural persons in the case of life and health insurance[125] that must assess the impact on fundamental rights that the use of such systems may produce.[126]

---

by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention. This provision shall not prevent Union law providing more extensive protection'.

122 Edwards (n 32).

123 Ibid.

124 Annex III 5(b).

125 Annex III 5(c).

126 Article 27 AI Act.

7 Co-regulation, Standardisation and Certification

High-risk AI systems or GPAI models that follow harmonised standards enjoy a presumption of conformity with the requirements for such systems and models set out in Chapter III, Section 2 or Chapter V, Sections 2 and 3 of the Act.[127]

The Commission can mandate the European Committees for Standardisation ('CEN') and Electrotechnical Standardisation ('CENELEC') to develop harmonised standards.[128] The Commission may also adopt implementing acts establishing common specifications for the relevant requirements where it has requested a standardisation organisation to draft harmonised standards. However, the request has not been accepted, the standards have not been delivered within the deadline, fundamental rights concerns have not been addressed, or the request has not been complied with. No requirement reference has been published in the EU's Official Journal.[129] Systems and models following such specifications will also enjoy the same presumption of conformity described above to the extent that the common specifications cover the relevant requirements or obligations.[130]

Harmonised standards often function as a necessary point of reference for compliance. It is likely, however, that as the AI Act remains vague on how to implement the essential requirements for high-risk AI systems, the responsibility for determining the specifics of these requirements will be delegated to CEN and CENELEC. That standardisation will thus be where the real rule-making occurs. This practice of effectively privately outsourcing complex negotiations has been controversial.[131] For example, Ansari and Mardais argue that the lack of representation from human rights experts or civil society organisations in these bodies raises concerns about their ability to protect fundamental rights. They state that technical standards related to data governance, transparency, security, and human oversight will directly impact fundamental rights and that

127 Article 40 AI Act.

128 Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council OJ L 316/12. Amended by Directive (EU) 2015/1535 of the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services (codification) OJ L 241/1.

129 Article 41(1) AI Act.

130 Article 41(3) AI Act.

131 Veale and Borgesius (n 51) 105–106.

the lack of democratic scrutiny or legislative interpretation of these standards may weaken the implementation of the Act.[132]

In addition to ensuring adherence to harmonised standards or common specifications, providers of high-risk AI systems must undertake a conformity assessment.[133] The EDPS and EDPB have called for such assessment to be generally undertaken by third parties. However, the final draft of the Act permits a self-assessment in many cases.

Providers of high-risk AI systems falling within the Annex III category of biometrics that have applied relevant standards or specifications may opt for a self-assessment or an assessment involving a notified body. Such providers must perforce opt for the latter procedure if harmonised standards do not exist if they have not applied or have only applied part of existing standards, or if they have not applied existing specifications.[134] Notified bodies are independent technical organisations to be established under the national law of a Member State[135] and tasked with verifying the conformity of high-risk AI systems.[136] Providers can generally opt for a notified body of their choice unless the high-risk system is intended to be put into service by law enforcement, immigration or asylum authorities or by Union institutions, bodies, offices or agencies. In this latter case, as applicable, the MSA referred to in Article 74(8) or (9) must act as a notified body.[137] Providers of high-risk AI systems falling within the remaining Annex III categories shall conduct a self-assessment without needing third-party involvement,[138] and providers of high-risk AI systems covered by the Union legislation in Section A of Annex I must follow the relevant conformity procedure under those legal acts.[139]

The Act also contains provisions for a 'CE marking of conformity'[140] (or CE certification), an EU declaration of conformity. Digital CE markings shall be used for digitally provided high-risk AI systems.[141]

---

132 Mehwish Ansari and Vidushi Marda, 'Opinion. AI Act – leaving oversight to the techies will not protect rights' (*EU Observer*, 5 May 2023) <https://euobserver.com/opinion/156992#:~:text=In%20May%2C%20the%20European%20Parliament,regulate%20the%20use%20of%20AI.> accessed 8 May 2023.

133 Article 43 AI Act.

134 Article 43(1) AI Act.

135 Article 31(1) AI Act.

136 Article 34 AI Act.

137 Article 43(1) AI Act.

138 Article 43(2) AI Act.

139 Article 43(3) AI Act.

140 Article 48 AI Act.

141 Article 48(2) AI Act.

8 Governance and Enforcement

8.1 *Specific Transparency Obligations*

In addition to the obligations under Chapters III, V, and IX described above, the AI Act confers two specific transparency obligations on providers and four on deployers in Chapter IV.

Thus, providers must ensure that AI systems intended to interact with natural persons are designed and developed so that natural persons are informed that they are interacting with an AI system (unless this is obvious from the point of view of a natural person). This obligation does not apply to systems authorised by law to detect, prevent, investigate and prosecute criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties, unless such systems are available for the public to report a criminal offence.[142]

Additionally, providers of AI systems, including GPAI systems, generating synthetic audio, image, video or text content must ensure that the system outputs are marked in a machine-readable format and detectable as artificially generated or manipulated. This obligation is also not applicable to handling criminal offences, nor where such systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof.[143]

Deployers of biometric categorisation or emotion recognition systems must inform the natural persons exposed to such systems of their operation and process any personal data by the GDPR, even though it remains controversial under data protection law whether facial recognition technology, which processes personal data only transiently falls within the scope of the definition of "personal data" and thus, of the GDPR.[144] This obligation also does not apply to such systems permitted by law to detect, prevent, and investigate criminal offences, subject to appropriate safeguards for the rights and freedoms of third parties.[145]

Deployers of an AI system that generates or manipulates image, audio or video content constituting a deep fake must disclose that the content has been

142 Article 50(1) AI Act.

143 Article 50(2) AI Act.

144 Peter Alexander Earls Davis, 'Facial Detection and Smart Billboards: Analysing the 'Identified' Criterion of Personal Data in the GDPR' (2020) 6(3) European Data Protection Law Review 365; Damian George and Kento Reutimann, 'GDPR Bypass by Design? Transient Processing of Data under the GDPR' (2019) 9 International Data Privacy Law 14; Nadezhda Purtova, 'From knowing by name to targeting: the meaning of identification under the GDPR' (2022) 12(3) International Data Privacy Law 163.

145 Article 50(3) AI Act.

artificially generated or manipulated. Again, this obligation does not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or where the content is part of a creative, satirical, artistic or fictional work or programme subject to appropriate safeguards for the rights and freedoms of third parties.[146] In such cases, there could be two persons to protect: the one depicted in the deep fake (if it concerns a person) and the audience/recipient of the deep fake. This provision seems focused on the recipient: misplaced beliefs of authenticity may present dangers, particularly fake news and disinformation in media law.

As such, Chapter IV of the Act includes the following final obligation: Deployers of an AI system that generates or manipulates image, audio or video content published to inform the public on matters of public interest shall disclose that the said content has been artificially generated or manipulated. This is also not applicable where use is authorised by law to detect, prevent, investigate or prosecute criminal offences, where the AI-generated content has undergone a process of human review or editorial control and where a natural or a legal person holds editorial responsibility for the publication of the content.[147]

Further specific transparency obligations for providers of GPAI models are also found within the provisions on the obligations of such providers,[148] already referred to above. This is because GPAI models may form the basis for a range of downstream systems, often provided by downstream providers, defined in the Act as a provider of an AI system, including a general-purpose AI system, which integrates an AI model, regardless of whether the AI model is provided by themselves and vertically integrated or provided by another entity based on contractual relations.[149] This necessitates "a good understanding of the models and their capabilities to enable the integration of such models into their products and to fulfil their obligations under [the AI Act] or other regulations".[150]

8.2 *Market Monitoring and Market Surveillance*

Regarding post-marketing controls, the Act provides that each Member State must establish or designate at least one notifying authority and at least one MSA for its purposes as national competent authorities.[151] These authorities are to

146 Article 50(4) AI Act.
147 Article 50(4) AI Act.
148 Article 53(1)(b) AI Act.
149 Article 4(68) AI Act.
150 Recital (101).
151 Article 3(48) and Article 70 (1) AI Act.

exercise their powers independently, impartially and without bias to safeguard the objectivity of their activities and tasks and ensure the application and implementation of such Acts; as long as these principles are observed, such activities and tasks may be performed by one or several designated authorities, per the organisational needs of the Member State.[152]

MSAs should have all enforcement powers under the AI Act and Regulation (EU) 2019/1020.[153] The EDPS will be the competent authority for AI systems put into service or used by Union institutions, agencies, offices and bodies.[154] MSAs will receive information through a chain of notification obligations. For instance, as discussed earlier, deployers who have reason to consider that the use per the instructions of a high-risk AI system may result in that system presenting a risk to the health or safety or fundamental rights of persons must inform (amongst others) the relevant MSA and both providers and deployers of high-risk AI systems are under an obligation to inform MSAs about any serious incidents.[155]

While it is envisaged that investigatory powers will primarily be in the hands of the national authorities, the AI Act also establishes a European Artificial Intelligence Board (AI Board)[156] with specified tasks centrally focused on facilitating the "consistent and effective application" of the Act.[157] During the legislative procedure, the co-rapporteurs of the Act in the EP proposed replacing the AI Board with an EU AI Office, envisaged as an independent body with its legal personality, funding, and staff.[158] The AI Office has since been established by a Commission Decision,[159] with the mission "to develop Union expertise and capabilities in the field of AI and to contribute to the implementation of Union law on AI".[160] At the same time, the AI Board has been retained as an entity to be established by the AI Act.

152 Article 70 AI Act.
153 Recital (156) AI Act. This is also reflected in the definition of an MSA in Art. 3(26) AI Act.
154 Article 70(9) AI Act.
155 Articles 26(5) and 73 AI Act.
156 Article 65 AI Act.
157 Article 66 AI Act.
158 Title VI 'Governance', Chapter 1 'European Artificial Intelligence Office', Articles 56–58, Text adopted by the leading parliamentary Committees of the European Parliament on 11 May 2023 <https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf> accessed 10 June 2023.
159 Commission Decision of 24.1.2024 establishing the European Artificial Intelligence Office C(2024) 390.
160 Recital (148) AI Act.

The AI Act further establishes a new, central database managed by the Commission to register 'standalone' high-risk AI systems.[161]

> In order to facilitate the work of the Commission and the Member States in the Artificial Intelligence field as well as to increase the transparency towards the public, providers of high-risk AI systems (other than those related to products falling within the scope of relevant existing Union harmonisation legislation) should be required to register themselves and information about their high-risk AI system in a EU database, to be established and managed by the Commission.[162]

The obligation of registration in the EU database has also been extended to deployers of high-risk AI systems listed in Annex III (except systems relating to critical infrastructure) that are public authorities, agencies or bodies.[163]

9 Regulatory Sandboxes

A regulatory sandbox allows innovators to explore and experiment with new and innovative products, services or businesses under a regulator's supervision. It potentially benefits all stakeholders: it allows regulators to understand the technology better, provides innovators with incentives to test their innovations in a controlled environment, and fosters consumer choice in the long run. However, regulatory sandboxes also come with a risk of being misused or abused. To succeed, an appropriate legal framework is needed.[164]

An 'AI regulatory sandbox' is defined in the Act as "a controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real-world conditions, an innovative AI system, under a sandbox plan for a limited time under regulatory supervision".[165]

Member States' national authorities must set up at least one AI regulatory sandbox at the national level and put a framework for governance and

161 Article 71 AI Act.
162 Recital (131) AI Act.
163 Article 49(3) AI Act.
164 Tambiama Madiega with Anne Louise Van De Pol, European Parliament Briefing Artificial Intelligence Act and regulatory sandboxes, June 2022 <https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf> accessed 30 May 2023.
165 Article 3(55) AI Act.

supervision in place.[166] Importantly, participants remain liable under applicable Union and Member States' liability legislation for any damage caused in the course of their participation in an AI regulatory sandbox, although in such cases, provided that the relevant providers observe the plan and terms and conditions for their participation, no administrative fines shall be imposed on them for infringements of the Act.[167] The modalities and conditions for establishing and operating the AI regulatory sandboxes will be adopted through implementing acts.[168] These will include matters such as eligibility criteria, application and selection procedures, participant rights and obligations, and further procedures relating to participation and exit from regulatory sandboxes.

An Open Loop report exploring an earlier version of this provision concluded that it is, in fact, these implementing acts, rather than the provision itself, that could create the necessary conditions for successful regulatory sandboxes. This report warned, however, that Member States should avoid creating conditions that are too favourable for participants, as this could distort the level playing field for AI development in Europe.[169] National competent authorities that have established AI regulatory sandboxes will cooperate and coordinate their activities within the framework of the AI Board and the AI Office.[170]

SMEs, including start-ups, will be provided priority access to the AI regulatory sandboxes and, where appropriate, a dedicated channel for communication to provide advice and respond to queries about the implementation of the Regulation, particularly in AI regulatory sandboxes.[171] It is envisaged that these sandboxes will allow proportionate application of the rules to the SMEs, insofar as permitted under existing legislation, and thus provide a space for experimentation under the new rules and the existing legal framework.[172]

The Act also regulates the processing of personal data in AI regulatory sandboxes for "developing, training and testing certain AI systems", which it

166 Article 57(6) AI Act.

167 Article 57(12) AI Act.

168 Article 58 AI Act.

169 Norberto De Andrade, Laura Galindo and Antonella Parra, 'Artificial Intelligence Act: A Policy Prototyping Experiment. EU AI Regulatory Sandboxes' (Open Loop, April 2023) <https://openloop.org/programs/open-loop-eu-ai-act-program/?utm_source=substack&utm_medium=email> accessed 8 May 2023.

170 Article 57(15) AI Act.

171 Article 62(1)(a) and (c) AI Act.

172 European Commission Impact Assessment of the Regulation on Artificial Intelligence, 21 April 2021, SWD(2021) 84 final, p.71.

considers to be 'further' processing.[173] Such further processing is permitted only when the conditions laid out in Article 59(1)(a) to (j) of the Act are met.

10 Room for National Initiatives in the EU?

In EU law, a 'full harmonisation measure' means that it allows no scope for stricter regulation at the national level. It is unclear whether the AI Act is intended to be such a measure. A discussion of the complex relationship between EU law and national-level laws[174] and the room left for national-level regulation, in particular, if not 'pre-empted' by the AI Act, is too complex to be tackled in the space of this book chapter. Suffice it to say that while the Act is primarily concerned with regulating 'high-risk' AI systems, non-high-risk AI systems are barely regulated but still fall within its scope, thus arguably precluding national legislation on such systems.[175] Having said this, Article 26 concerning obligations of deployers of high-risk AI systems provides that the "obligations [relating to the use of high-risk AI systems and in particular the assignment of human oversight] are without prejudice to other deployer obligations under Union or national law and to the deployer's freedom to organise its resources and activities to implement the human oversight measures indicated by the provider"[176] thus specifically allowing room for national legislation.

11 Conclusion: Regulatory Balance

Regulation is necessary but tricky. It is important to get the regulatory balance right. Too little regulation and important rights and interests may be undermined; too much regulation and innovation, development and investment – and thus, the EU's competitiveness on the global stage – may be stifled. Excessive regulation also tends to help further entrench the market power of the incumbent big players. When discussing regulating AI, an appropriate legal framework should enable investment and innovation while upholding

173 Article 59 AI Act.

174 Discussed in: Stephen Weatherill, *The Fundamental Question of Minimum or Maximum Harmonisation*, in S Garben and I Govaere (eds), The Internal Market 2.0, Oxford: Hart Publishing, 2020, 261–284.

175 Veale and Borgesius (n 51) 110.

176 Article 26(3) AI Act.

and protecting fundamental rights and safety. This also falls within the broader picture of the Digital Single Market in the EU.

Strengths of the AI Act include the risk-based approach, the prohibition of certain AI systems, and the possibility for societal scrutiny via a public database. However, this chapter has also commented on potential weaknesses, such as the role of the standardisation bodies that lack the participation of human rights experts and/or civil society organisations and that are likely to write the real rules against which providers will self-assess.

The AI Act acknowledges the risks to individuals and society posed by AI technologies and seeks to prevent them without stifling innovation. Nevertheless, earlier versions of the Act were criticised for being "generally less encouraging of the development of new AI technologies" and not going far enough to protect fundamental rights.[177] It remains to be seen whether the Act will lead to a balanced model between fundamental rights and technological development or whether Europe's zealousness to regulate this space will mean that innovative businesses will base themselves elsewhere and choose not to offer their services within the EU. Beyond the internal market, the potential global implications of the EU AI Act should also not be ignored, considering the EU's capacity to set *de facto* international standards for AI regulation due to its substantial market size and regulatory influence. In this sense, the EU's influence may extend beyond hard law to a form of soft extraterritoriality—termed the "Brussels effect"—which may encourage non-EU countries and multinational corporations to align their AI practices with EU standards. This and other complex issues will undoubtedly be the subject of several further studies in future.

Bibliography

Ansari M, and V Marda, 'Opinion. AI Act – leaving oversight to the techies will not protect rights' (EUobserver, 5 May 2023) <https://euobserver.com/opinion/156992#:~:text=In%20May%2C%20the%20European%20Parliament,regulate%20the%20use%20of%20AI.> accessed 8 May 2023.

Bostrom N, Superintelligence: Paths, Dangers, Strategies (OUP 2014).

Earls Davis P A, 'Facial Detection and Smart Billboards: Analysing the 'Identified' Criterion of Personal Data in the GDPR' (2020) 6(3) European Data Protection Law Review 365.

177 See Edwards (n 32) and Veale and Borgesius (n 51).

Ebers M, Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act (August 6, 2021). The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics, Available at SSRN: https://ssrn.com/abstract=3900378 or http://dx.doi.org/10.2139/ssrn.3900378

George D and K Reutimann, 'GDPR Bypass by Design? Transient Processing of Data under the GDPR' (2019) 9 International Data Privacy Law 14.

Höppner T and L Streatfeild, 'ChatGPT, Bard & Co.: An Introduction to AI for Competition and Regulatory Lawyers' (2023) 9 Hausfeld Competition Bulletin (1/2023), Article 1, <http://dx.doi.org/10.2139/ssrn.4371681> accessed 20 May 2024.

Johanson P, *Online Filter Bubbles* (Greenhaven Publishing, 2017).

Pariser E, *The filter bubble: How the new personalised web is changing what we read and how we think* (Penguin, 2011).

Purtova N, 'From knowing by name to targeting: the meaning of identification under the GDPR' (2022) 12(3) International Data Privacy Law 163.

Raposo V L, 'Ex Machina: Preliminary Critical Assessment Of The European Draft Act On Artificial Intelligence' (2022) 30(1) International Journal of Law and Information Technology 88.

Truby J, R D Brown, I A Ibrahim and O C Parellada, 'A Sandbox Approach To Regulating High-Risk Artificial Intelligence Applications' (2022) 13(2) European Journal of Risk Regulation 270.

Veale M, and F Z Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (2021) 22(4) Computer Law Review International 97.

Weatherill S, 'The Fundamental Question of Minimum or Maximum Harmonisation', in S Garben and I Govaere (eds), *The Internal Market 2.0* (Hart Publishing 2020) 261–284.

Legislative Train

Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts COM/2021/206 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – EU Council General approach – Brussels, 25 November 2022 https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf

Artificial Intelligence Act: Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))

<https://assets-global.website-files.com/637e4725db842e4068de0899/6565f47fb10192c89ae028ad_PARLIAMENT%20POSITION.pdf >

CORRIGENDUM to the position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Regulation (EU) 2024/ ... of the European Parliament and of the Council 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), 19.4.2024 <https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf>

Legislation

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) OJ L 277/1

Official Publications

European Commission Impact Assessment of the Regulation on Artificial Intelligence, 21 April 2021, SWD(2021) 84 final

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) 18 June 2021 <https://edps.europa.eu/system/files/2021-06/2021-06-18-edpb-edps_joint_opinion_ai_regulation_en.pdf>