The Dawn of Regulated AI: Analyzing the European AI Act and its Global Impact

Kalojan Hoffmeister*

Contents

* Kalojan Hoffmeister, Dipl. Jur. (HU Berlin), Maitre en droit (Paris Panthéon Assas), LL.M (King's College London); Since April 2024 Policy Assistant at the European Commission's Directorate General for Economic and Financial Affairs (DG ECFIN); views expressed in this article are personal (manuscript was finalised in March 2024). Email: kalojanhoffmeister@gmail.com.

Abstract

The article provides an overview of the genesis of the EU AI Act, its economic and security context, and the intricacies of its international implications. It discusses the main elements of the Act, particularly some changes it underwent during the negotiation process between the EU Commission, Parliament, and Council. The AI Act is set against the backdrop of global economic and security landscapes, reflecting on the strategic implications of AI in the US-China geopolitical rivalry and the EU's positioning within it. In particular, the article critically highlights the prohibited practices under the AI Act, the introduction of a nuanced classification system for high-risk AI applications, the fundamental rights impact assessment obligation, the new provisions on General Purpose AI and the Act's governance structures. The article concludes with a forward-looking perspective on the EU's role in shaping global AI governance, indicating the Act's potential to serve as an international benchmark.

Keywords: Artificial Intelligence Act, AI-Act, European Union, Hiroshima Process, AI Safety, Regulation, Global Digital Compact, General Purpose AI, Fundamental Rights Impact Assessment, AI Office

A. Introduction

On March 13th 2024, the European Parliament voted in favor of the "Regulation on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act – AI Act)". This marks the end of a long legislative process and the beginning of Europe's regulatory attempts in the field. Given the EU's importance in global trade and technology, the AI Act is likely to become a reference point for many other

legislators in the world ("Brussels effect"). This article will first track down the international context (B.) and the road to the AI Act itself (C.). The ensuing section will then present the main elements of the AI Act with a particular emphasis on some changes that occurred during the negotiations between the Commission, the Parliament, and the Council (D.), before offering a conclusion (E.). As the AI Act will only be published in the Official Journal in May or June 2024, the references will refer to the version as adopted by the Parliament.[1]

B. The international context

I. Economic significance and security impact

It is common ground that AI will have a significant economic impact on global productivity. A recent 2023 research indicates that AI could add the equivalent of USD 2.6 trillion to USD 4.4 trillion annually to the global economic output around the world.[2] A previous study found that the improved productivity could contribute up to USD 15.7 trillion on the global economy in 2030.[3]

Besides the enormous economic potential, AI is also considered a game-changer technology that can give a strategic advantage in international security competition. In the United States, AI is seen as a central part of the US-China geopolitical competition and regarded as a matter of national security. For example, the recent US export controls[4] on advanced computing semiconductors and manufacturing equipment was based on national security and foreign policy concerns, in response to China's "military/civilian fusion and military modernization" strategy. In July 2017, the Chinese government outlined its ambitious goal of making China a global leader in AI by 2030. The US seeks to curb China's ability to follow suit in this race.[5] One such tool is regulation.

1 European Parliament legislative resolution of 13 March 2024 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)).

2 *Chui* et al., The economic potential of generative AI: The next productivity frontier (McKinsey Digital, 14 June 2023), available at: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier (22/3/2024).

3 *PWC*, Sizing the prize – What's the real value of AI for your business and how can you capitalise?, available at: https://www.pwc.com.au/government/pwc-ai-analysis-sizing-the-prize-report.pdf (22/3/2024).

4 *US Department of Commerce, Bureau of Industry and Security*, Commerce Strengthens Restrictions on Advanced Computing Semiconductors, Semiconductor Manufacturing Equipment, and Supercomputing Items to Countries of Concern, available at: https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3355-2023-10-17-bis-press-release-acs-and-sme-rules-final-js/file (22/3/2024).

5 See for example the analysis in *Bunde* et al., Munich Security Report 2024: Lose-Lose?, Munich Security Conference, 2024, p. 99.

During the summer of 2023, the US worked on developing their own national plan on how to deal with AI. Firstly, President *Biden* obtained a set of voluntary commitment to manage the risk posed by AI from 15 companies[6] working on frontier generative AI models. Secondly, a 110-page Presidential Executive Order on the Safe, Secure, and Trustworthy AI was issued on October 30th 2023.[7] Thirdly, together with the EU through the G7 Hiroshima Process, a voluntary Code of Conduct guiding organizations developing the most advanced AI systems was launched on October 30th 2023.[8]

Similarly, AI still remains high on the Chinese domestic agenda. Mid-October 2023, President *Xi* announced China's intentions to help shape international AI governance in form of a Global Artificial Intelligence Governance Initiative.[9] It seeks to establish China as a leader in what they describe as a "principled, cooperative development of AI worldwide". The timing of the Global AI Governance Initiative is notable, coming just a day after the US export controls on semiconductors and just before the UK's AI Safety Summit, which China eventually still joined. While the specifics of a new international structure for dealing with AI are still to be developed (at the UN), the 1-2 November 2023 UK AI Safety Summit was a first step in establishing a common understanding of the issues and a shared commitment to addressing them.[10] The fact that both the US and China participated is worth noting.

II. International cooperation

The initial international work on AI has been going on since 2019, where the OECD adopted its AI Principles.[11] Since then, AI has been subject to international cooperation in a number of organisations and fora. The EU's work has first and foremost taken place in the context of G7, the EU-US TTC and the Council of Eu-

6 Leading AI Companies that signed up include: Adobe, Amazon, Anthropic, Cohere, Google, IBM, Inflection, Meta, Microsoft, Nvidia, OpenAI, Palantir, Salesforce, Scale AI, and Stability.

7 *Biden,* Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, October 30th 2023.

8 *European Commission,* Press release, Commission welcomes G7 leaders' agreement on Guiding Principles and a Code of Conduct on Artificial Intelligence (30 October 2023), available at: https://digital-strategy.ec.europa.eu/en/news/commission-welcomes-g7-leaders-agreement-guiding-principles-and-code-conduct-artificial (5/5/2024).

9 *Wang/Yin,* China launches Global AI Governance Initiative, offering an open approach in contrast to US blockade (Global Times, 18 October 2023), available at: https://www.globaltimes.cn/page/202310/1300092.shtml (22/3/2024).

10 See in this respect the Bletchley Declaration by countries attending the AI Safety Summit, 1-2 November 2023, available at: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 (5/5/2024).

11 *OECD*, OECD Legal instruments, Recommendation of the Council on Artificial Intelligence (22/5/2019), available at: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (5/5/2024).

rope. The latter is set to conclude a process of developing a framework Convention on AI and Human Rights in May 2024, which will be the first of its kind, should it succeed.[12] The work on a global framework for AI is expected to commence in the context of the UN 2024 Summit for the Future, where a Global Digital Compact is likely to be enacted. AI is expected to be a significant focus of the Compact. The UN Secretary General recently established a 39-member High-Level Advisory Board on AI to provide guidance during the intergovernmental process. In addition, the UN TECH-envoy will prioritize AI matters until the Summit. There are regularly calls for an international body to be set up to address the risks of AI and establish the science behind it.

III. The position of the EU

Against the backdrop of the US-Chinese rivalry and the rather slow progress on the international level, the European Union positioned itself in the middle. It affirmed a need to grasp the benefits of AI, while not turning a blind eye to the eventual pitfalls in the technological race. Regulating the risks in a unilateral manner would enable the European economy to move forward without being accused of falling foul of European values.[13] Hence, from a European perspective, the AI Act employs a comprehensive, risk-based, human-centric approach to governing AI, balancing innovation and ethical principles. Before turning to them, the legislative history of the AI Act will be briefly recalled.

C. The road to the AI Act

I. Policy documents of the European institutions

The Commission published its first thoughts on regulating AI in a Communication in April 2018.[14] It also established a "High Level Expert Group on AI", which published ethical guidelines for trustworthy AI[15] and policy and investment recommen-

12 See in this respect, *Council of Europe*, Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law, available at: https://rm.coe.int/-1493-10-1b-committee-on-artificial-intelligence-cai-b-draft-framework/1680aee411 (5/5/2024).

13 Critical of this narrative and arguing that the EU's AI policy prioritises "jurisdictional independence over citizens sovereignty" when entering the global AI race: *Mügge*, JEPP 2024; equally critical that the AI Act is accompanied by a "side effect", limiting the spread of values and protection of fundamental rights worldwide: *Almada/Anca*, GLJ 2024.

14 *European Commission*, Communication From the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe 2018, C(2018) 237 final.

15 *European Commission*, Ethics Guidelines for Trustworthy AI | Shaping Europe's Digital Future, 8 April 2019, available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (18/4/2023).

dations a year later.[16] In December 2018, the European executive presented a Coordinated Plan for AI.[17] This was followed by a further Communication[18] (2019) and an Expert Group Assessment List.[19] The White Paper of February 2020[20] stimulated a broad multi-stakeholder discussion, the outcome of which was published in an advisory paper. The European Council, the Council, and the European Parliament (EP) were not passive either: In 2017,[21] 2019[22] and 2020[23] respectively, the European Council and the Council stressed the urgency of the issue and the importance of fundamental rights protection in the light of AI. The EP, in turn, called on the Commission to take legislative action in the field of AI as early as 2017 in a robotics resolution.[24] The EP adopted another resolution in June 2020, on AI and industrial policy,[25] and finally set up its own special committee on AI in June 2020.[26] This was followed by a series of resolutions in October 2020 on ethics,[27] liability,[28] and copy-

16 *European Commission*, Policy and Investment Recommendations for Trustworthy Artificial Intelligence | Shaping Europe's Digital Future, 26 June 2019, available at: https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence (18/4/2023).

17 *European Commission*, Coordinated Plan on Artificial Intelligence | Shaping Europe's Digital Future, 7 December 2018, available at: https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence (18/4/2023).

18 *European Commission*, Communication: Building Trust in Human Centric Artificial Intelligence | Shaping Europe's Digital Future, 8 April 2019, available at: https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence (18/4/2023).

19 *European Commission,* Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment | Shaping Europe's Digital Future, 17 July 2020, available at: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment (18/4/2023).

20 *European Commission*, White Paper on Artificial Intelligence: A European Approach to Excellence and Trust, 19 February 2020, available at: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (18/4/2023).

21 *European Council*, European Council Conclusions of 19 October 2017 (EUCO 14/17).

22 *Council of the European Union Permanent Representatives Committee*, Note of 11 February 2019 on Artificial Intelligence, b) Conclusions on the coordinated plan on ai adoption (Doc. 6177/19).

23 *European Council General Secretariat of the Council*, Note of 2 October 2020 on Special meeting of the European Council – Conclusions (EUCO 13/20).

24 *European Parliament*, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), 2018/C 252/25.

25 *European Parliament*, Resolution of 12 February 2019 on a comprehensive European industrial policy on artificial intelligence and robotics (2018/2088(INI)), 2020/C 449/06.

26 *European Parliament*, Decision of 18 June 2021 on setting up a special committee on artificial intelligence in a digital age, and defining its responsibilities, numerical strength and term of office (2020/2684(RSO), 2021/C 362/42).

27 *European Parliament*, Resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence robotics and related technologies 020 (2020/2012(INL)), 2021/C 404/4.

28 *European Parliament*, Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014 (INL), 2021/C 404/05.

right.[29] Further resolutions in the fields of law enforcement,[30] education, culture, and audio-visual[31] came along. In May 2022, the EP published a comprehensive resolution[32] consolidating its position on AI issues. It should be recalled, however, that the Council's conclusions and the Parliament's resolutions had no legal effect because only the Commission has the right of initiative for binding legislation according to Art. 17 TEU.

II. The Commission proposal of April 2021

The Commission changed the situation when it exercised its right of initiative under Article 17 TEU. Its proposal of 21 April 2021[33] then triggered a formal legislative process. The European Economic and Social Committee,[34] the European Committee of the Regions,[35] the European Data Protection Board (EDPB), the European Data Protection Supervisor (EDPS),[36] and the European Central Bank (ECB)[37] delivered their respective opinion in the second half of 2021.

III. The trilogues

Consultations in the Council started under the Portuguese Presidency (first half of 2021) and continued with the Slovenian (second half of 2021) and French Presidency (first half of 2022). The Council eventually adopted a general approach to the AI Act during one of the last meetings of the Czech presidency in December 2022. In

29 *European Parliament*, Resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies (20/2015(INI)), 2021/C 404/06.
30 *European Parliament*, Draft Report on Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters (2020/2016(INI)).
31 *European Parliament*, Draft Report on Artificial intelligence in education, culture and the audiovisual sector (2020/2017(INI)).
32 *European Parliament*, Resolution of 3 May 2022 on Artificial intelligence in a digital age (2020/2220 (INL)).
33 *European Commission*, Proposal for a Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, C(2021) 206 final.
34 *European Economic and Social Committee*, EESC Opinion on the Artificial Intelligence Act, available at: https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/regulation-artificial-intelligence (3/4/2024).
35 *European Committee of the Regions*, Opinion of the European Committee of the Regions — European Approach to Artificial Intelligence — Artificial Intelligence Act (Revised Opinion), available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021AR2682 (3/4/2024).
36 *European Data Protection Board and European Data Protection Supervisor*, 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, available at: https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en (16/4/2023).
37 *European Central Bank*, Opinion of the European Central Bank of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence, CON(2021)/40.

the European Parliament, the discussions were led by the Committee on Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home under a joint committee procedure. The Legal Affairs Committee (JURI), the Committee on Industry, Research and Energy (ITRE), and the Committee on Culture and Education (CULT) were associated with the legislative work on shared and/or exclusive competences. The Parliament adopted its position to the AI Act in mid-June 2023. Thereafter the trilogues began. In total, the co-legislators held five trilogues. A political agreement was struck in December 2023 after which co-legislators proceeded to carry out further technical work in January 2024 to align the text of the recitals with the text of the articles as agreed during the final trilogue. The text was approved by the Council in February 2024 and by Parliament on March 13th 2024. The AI Act will enter into force twenty days after its publication in the official Journal, and be fully applicable 24 months after its entry into force, except for bans on prohibited practices, which will apply six months after the entry into force date; codes of practice (nine months after entry into force); general-purpose AI rules including governance (12 months after entry into force); and obligations for high-risk systems (36 months).

D. Main elements of the AI Act

The AI Act contains twelve chapters. This section will briefly present them and focus on those parts, which were most debated during the legislative process.

I. General provisions (Chapter I)

Chapter I sets out the subject matter, scope, and definitions of the Act. These general provisions contain important principles, which elaborate on the interpretation and application of the entire Regulation.

1. Subject matter (Article 1)

Reflecting the two legal bases of the Act, namely Article 16 TFEU on data protection and Article 114 TFEU on the internal market, Article 1 (1) names as purpose of the AI Act a) to improve the functioning of the internal market and to uptake a human-centric approach and the trustworthiness of AI on the one hand, and b) the protection of health, safety, and human rights against its harmful effects, on the other hand. Importantly, the Parliament added the purposes of c) fighting risks against democracy and the rule of law, being aware of the role of AI in recent election interferences in the United States and other elections.[38] Finally, the provision also makes clear that the AI Act is not supposed to inhibit the development of AI as

38 *Adam/Hocquard,* Artificial Intelligence, democracy and elections, European Parliamentary Research Service, October 2023, available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751478/EPRS_BRI(2023)751478_EN.pdf (22/3/2024).

such, but mostly intended to regulate its potential negative use. In this respect, the addition of d) "support innovation" as the fourth purpose may be important when adopting future guidelines on specific AI applications.

The subject matter and the legal basis thus make apparent that the AI Act is a mixture between a regulation focusing on fundamental rights *and* a product safety regulation. This is interesting considering that the prevailing narrative pushed for by the Commission focused predominantly on safeguarding fundamental rights and the need for a human-centered AI. Nevertheless, the choice of Art. 114 TFEU as the internal market legal basis (and legal basis closer to product safety regulations than to fundamental rights) can easily be explained: As *Almada and Petit* rightly point out, the EU's AI Act takes a product safety approach, reflecting both the EU's limitations and strengths.[39] The Act's broad scope due to its horizontal nature (regulating across sectors), clashes with the EU's primarily sector-specific legislative competencies. To bridge this gap, the Commission chose to leverage the EU's competence for single market harmonization. This necessitates framing the regulations as market-focused to avoid lengthy and potentially highly problematic discussion on the EU's powers for the regulation of AI. Furthermore, this product safety approach also plays to the EU's strengths as regards global influence and domestic enforcement. Decades of established EU product safety law offers a robust foundation which influences regulations globally.[40] In terms of enforcement, the EU and its Member States can take advantage of existing knowledge and enforcement infrastructure in the area of product safety, avoiding the need for entirely new regulatory norms.

2. Scope (Article 2)

According to Article 2(1), the Act applies to those who bring AI application into the market, such as providers and deployers of AI, importers and distributors, or product manufacturers, which place their products on the market. Moreover, not only their customers can rely on the act, but all persons located in the Union which are affected by the employment of AI (lit. g). In essence, the AI Act is designed to have extraterritorial effects, meaning it applies regardless of where an AI system's provider or operator is based, as long as EU users are affected. Article 2 therefore tries to ensure that companies cannot circumvent the AIA's regulations by simply relocating to countries with looser laws if they want to take advantage of the European market. often referred to as the "Brussels Effect"[41] or even a "post Westphalian world order"[42]. At the same time, this large scope is limited by the enactment of several exceptions. Three of them merit particular attention.

39 *Almada/Petit,* Robert Schuman Centre for Advanced Studies Research Paper No. 2023/59, pp. 11–12.

40 *Siegmann/Anderljung*, Centre for the Governance of AI 2022; critical however: *Almada/Anca*, German Law Journal 2024.

41 *Bradford.*

42 With further references to "post Westphalian word order", *Wörsdörfer,* p. 113.

a) National security exception

Article 2(3) excludes "activities which fall outside the scope of Union law, and in any event activities concerning military, defence or national security". This wording mirrors the position of the Council during the trilogue negotiations and is the result of a combined wording of similar provisions from the Data Act[43] and the Cybersecurity Act.[44] Even if one would concur with the (natural) wish of Member States to push for such a national security exception, the agreed upon wording in its final version is still problematic. First, the exception on defence or national security grounds is much wider than the initially proposed carveout for AI systems developed or used exclusively for military purposes. The current wording excludes not only national security activities but also "activities *concerning* military or defence security", which often, but not always, may fall under national security. It would have been advisable to opt for wording that is closer aligned with Article 4(2), third sentence TEU ("In particular, national security remains the sole responsibility of each Member State"). Second, the clause establishes a problematic relationship between the scope of application of the Act and the Member States' responsibility for national security. It suggests that measures adopted by Member States for the purpose of safeguarding national security are excluded from the application of EU law. This, however, is neither in line with Art. 4(2) TEU, nor corresponding case law by the European Court of Justice. In fact, according to the Court even measures adopted by the Member States for the purposes of safeguarding national security, defence and public security are not excluded from the application of EU law, with the consequence that Member States taking such measures must comply with that law.[45] Otherwise this might impair the binding nature of Union law and its uniform application.[46] Yet, the final version of Article 2(3) incorrectly suggests that EU law does not apply in these areas, which could create legal uncertainty.[47] For this reason, the

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

44 Regulation (EU) 2019/881 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJ L 151 of 7/6/2019.

45 See e.g. CJEU, case C-265/95, *Commission v. France*, ECLI:EU:C:1997:595, paras. 33–35; CJEU, case C-285/98, *Tanja Kreil v. Bundesrepublik Deutschland*, ECLI:EU:C:2000:2, paras. 16–17; CJEU, case C-186/01, *Alexander Dory v. Bundesrepublik Deutschland*, ECLI:EU:C:2003:146, para. 30; CJEU, case C-337/05, *Commission v. Italy*, ECLI:EU:C:2008:203, paras. 42–43; CJEU, case C-294/05, *Commission v. Sweden*, ECLI:EU:C:2009:779, paras. 43–46; and most recently CJEU, joined cases C- 511/18, C-512/18 and C-520/18, *La Quadrature du Net and Others v. Premier Ministre and others,* ECLI:EU:C:2020:791, para. 99.

46 CJEU, case C-285/98, *Tanja Kreil v. Bundesrepublik Deutschland*, ECLI:EU:C:2000:2, para. 16.

47 The same observations were made by the Commission with regard to the Data Act. See *European Commission,* Replies of the Commission to positions and resolutions adopted by the European Parliament – November I 2023 part-session of 6 December 2023, SP(2023) 632 final.

above-mentioned provisions may trigger in the future the need for an ECJ clarification.

On a positive note, however, it is to be welcomed that the legislator has clarified in recital 12a that so called "dual use goods" are covered by Union law. For example, if an AI system originally intended for military or national security is repurposed for civilian, humanitarian, law enforcement, or public security use, it must comply with the Regulation. Entities using the system for these non-military purposes must ensure it meets the Regulation's standards unless it is already compliant. AI systems designed for both military and non-military uses are subject to the Regulation and must meet its requirements. However, this does not prevent military, defense, or national security entities from using AI systems for their original purposes, which are exempt from the Regulation. Likewise, an AI system created for civilian or law enforcement but later used for military, defense, or national security does not fall under the Regulation.

b) Exceptions for scientific research and development

The Commission's initial text left room for ambiguity as to whether scientific research would fall under the scope of the Regulation and thus would have to abide by the obligations set out by it.[48] Article 2(6) now clarifies that the Regulation does not apply to AI systems and models, including their output, specifically developed and put into service for the sole purpose of scientific research and development. From a policy point of view this is laudable, as creating burdensome obligations on solely scientific research and development would slow down or even hinder innovation and scientific breakthroughs.

In this respect, however, the distinction between a model specifically developed for scientific and development purposes and an AI model that falls within the regulatory sandboxes is a very delicate one, with huge consequences: If an AI model was found to be solely for scientific purposes, the obligations would not apply to it. If, on the other hand, the AI model was not found to be developed or put into service specifically for scientific and development purposes, then such a model would both enjoy the administrative help of the competent authorities, and have to abide by the Regulation and its obligations set out for regulatory sandboxes in Chapter VI. These include, amongst others, the training, testing and validation requirements that will be set out by the terms and conditions in a Commission implementing act (Art. 58) and further obligations on the processing of personal data (Art. 59). Furthermore, national competent authorities will retain the power to suspend the testing process and participation in the sandbox, temporarily or permanently.

The main differentiating *criterium* seems to be the intention of the AI model providers on whether it is "developed or put into service specifically for scientific and development purposes", or, whether (as in the case of regulatory sandboxes) the intent is to place and make the AI model available on the market. Such a

48 *Hoffmeister*, WHI-Paper 2023/01; *Smuha and others,* LEADS Lab.

differentiation based on intent is inherently subjective and can be difficult to assess. This subjectivity can lead to inconsistencies in how the Regulation is applied. Also, the purpose of an AI model may evolve over time. A project that begins as a purely scientific endeavor could shift towards commercial application as it develops. This fluidity makes it problematic to establish a clear-cut moment when the model's intent changes from research to commercialization. Moreover, if intent is the only criterion, there may be a loophole for developers to simply claim scientific intent to circumvent the regulation, even when there is a clear potential for commercial application. In any case, it would be advisable for any AI science and research lab to apply for the regulatory sandboxes as if they were getting ready for potential market access, despite initially not having the intent to do so.

c) Workers' rights

At the wish of the European Parliament, the AI Act permits Member States or the Union to introduce more worker-friendly laws, regulations, or administrative measures, to safeguard workers' rights concerning the use of AI systems by employers, as well as to support or permit more advantageous collective agreements for workers. This "opening clause" in Article 2(11) aligns with the approach in Article 29(2), which indicates that the duties of users of high-risk AI systems are not definitive and can be expanded upon by Union or national legislation. This is also consistent with the overarching principles of Union law, particularly in view of Article 153(3) TFEU, which allows the EU to set only minimum standards in employment law, while Member States retain the discretion to enforce stricter measures that are in agreement with the Treaties. Yet, it should be noted that Article 153(3) TFEU pertains specifically to employment law. As such, any Member State that wishes to introduce new national provisions based on the opening clause of Article 2(11) AI Act must ensure that these provisions primarily pertain to employment law to maintain treaty compliance.

d) Open-source models

Finally, another controversial issue was whether open-source models should be regulated, as well. An open-source AI model refers to an AI system whose underlying code, algorithms, and possibly even datasets are freely available to the public. In concreto, an open source model's design and underlying code is accessible for anyone to use, modify, distribute, and even integrate into their own projects without having to pay for licensing fees or adhere to strict proprietary constraints. According to Article 2(12), open-source models are now being exempt from the obligations of the regulation, unless they are placed on the market, or put into service as high-risk AI systems, or an AI system that falls under Chapter II and IV.

The problem with an open-source exemption is that powerful open-source AI models are a "double edged sword".[49] As *Hacker* rightly points out, recently, France-based Mistral AI introduced their new model, the Mistral 8x7b, which features an innovative architecture and has been released shortly after the conclusion of the trilogue discussions.[50] This model demonstrates performance that matches or occasionally surpasses that of ChatGPT across various benchmarks. Remarkably, it is distributed as open-source, allowing free access to anyone, who contributes to the democratization of AI technology and serves as a counterbalance to monopolistic tendencies within the AI industry. The model is equipped with fundamental safety protocols which, however, can also be disabled. Such powerful unguarded models could fall in the hands of bad-faith actors and be misused for malicious activities.

Although the final wording introduces the caveat for open-source AI models, which would be prohibited AI models in the sense of Chapter II, and for certain AI systems and GPAI models in the sense of Chapter V[51], the current framework does not seem to adequately address the potential risks associated with highly capable open-source AI models. It would be prudent to categorize such powerful software, specifically those exceeding certain computational thresholds as dual-use items. Dual-use goods, given their potential application in sensitive areas, should not be freely distributed but rather made available through a regulated and monitored platform. This would allow for appropriate oversight over the usage of these AI models. In fact, research from the Future Society Institute[52] suggests that the cost to meet regulatory standards for developing an AI model with the same level of capability as ChatGPT (which performs at 10^{24} FLOPS) is estimated to be around $60 million. Regulatory compliance accounts for about 1% of the investment. Therefore, for organizations investing in major open-source AI projects, the additional costs for compliance are relatively minor and should be feasible within their overall project budgets.

3. Definitions (Article 3)

In line with standard EU legislative practice, Article 3 contains a list of terms with their definitions. The most important in the list of 68 (!) definitions, is the first one on AI itself. Point 1 defines an AI system as a "machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after

49 *Bertuzzi,* AI Act's post-agreement commentary (Euroactiv, 15 December 2023), available at: https://www.euractiv.com/section/digital/podcast/ai-acts-post-agreement-commentary/ (22/3/2024).

50 *Hacker*, What's Missing from the EU AI Act: Addressing the Four Key Challenges of Large Language Models, VerfBlog, 13 December 2023, available at: https://verfassungsblog.de/whats-missing-from-the-eu-ai-act/, DOI: 10.59704/3f4921d4a3fbeeee; *Novelli* et al., p. 4.

51 See Art. 54 (5), Recital 103, 104.

52 *The Future Society*, EU AI Act Compliance Analysis: General-Purpose AI Models in Focus, December 2023, available at: https://thefuturesociety.org/wp-content/uploads/2023/12/EU-AI-Act-Compliance-Analysis.pdf (22/3/2024).

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".

To a large extent, this text is inspired by the definition of AI provided by the OECD.[53] Crucially, this definition of an AI system does not depend on techniques of approaches as defined in Annex I. Thus, the co-legislators moved away from the previously proposed enumerative approach that had granted the European Commission wide discretionary powers to further define which AI techniques would fall under the definition of AI. This leads to a more static approach, which critics could argue is not future proof. This is so because in an area such as AI, where new techniques and approaches are discovered constantly, and which is only expected to become more innovative, an enumerative approach giving the Commission some leeway to adapt to new situations could arguably have been better. Yet, especially in times where many are calling for AI regulation, the need for legal certainty is also important. Having a non-amendable AI definition warrants for more regulatory foresight, which companies and businesses have been calling for.

Turning to the definition itself, the notion of "autonomy" is key. Recital 6 clarifies that autonomy means having some degree of independence of actions from human involvement and of capabilities to operate without human intervention. Nevertheless, due to the vague term of "varying levels of autonomy", the primary difficulty of determining this degree of autonomy remains. Currently no widely accepted methodology exists for assessing autonomy in AI. Examples in guidelines would be helpful.

Most importantly, however, the chosen definition is to be welcomed because it aligns with international efforts in AI, thereby ensuring legal certainty beyond the EU, fostering global harmonization, and facilitating broad acceptance.

4. AI literacy (Article 4)

The final provision of the first chapter (Article 4) deals with the new concept of "AI literacy", introduced by the EP. It had proposed new measures requiring AI literacy for the general public and for staff working with AI. In particular, the Union and the Member States should promote measures for the development of a sufficient level of AI literacy, across sectors, and taking into account the different needs of groups of providers, deployers, and affected persons concerned, and their respective technical knowledge, experience, education, and training, and the context the AI systems are to be used in.

During the negotiations, the ambitions of the European Parliament (EP) were severely curtailed. According to the final version of Article 4, only providers and deployers of AI systems are under an obligation to provide AI literacy to their staff and "other persons dealing with the operation and use of AI systems on their be-

53 *OECD*, Recommendation of the Council on Artificial Intelligence of 22 May 2019 amended on 8 November 2023, OECD/LEGAL/0449.

half". Moreover, according to recital 20, the European Artificial Intelligence Board should support the Commission to promote AI literacy tools, public awareness and understanding of the benefits, risks, safeguards, rights, and obligations in relation to the use of AI systems.

This provision has good intentions but may be challenging to put into practice for both the European Commission and Member States, since the AI Act is not tied directly to any dedicated funding program. It might have been better to develop criteria for AI literacy in the relevant strategies like the Coordinated Plan on AI[54], or together with other funding options available through European, national, or regional programs. Additionally, this could include collaboration with public administrations on digital transformation and innovation, or align with the European Declaration on Digital Rights and Principles for the Digital Decade that supports digital education and skills.[55] In any case, regulators should watch to ensure that any requirements for providers and users do not create excessive pressure but are instead reasonable demands consistent with the responsibilities outlined in the AI Act, such as the necessary user training that providers must offer as per Article 9 (5)(c) for high-risk AI systems.

II. Prohibited AI practices (Chapter II)

1. The closed list technique

Chapter II consists of a single Article 5, which prohibits certain enumerated AI practices. The Article does not provide for any possibility to amend this closed list. In other words, the European legislator is stuck with the prohibited AI-applications they agreed on, unless they decide to change the legal text through the reopening of an ordinary legislative procedure.

Considering the fast-changing technological developments and potential detrimental risks for fundamental rights and European democracy one can question, whether using such a closed list technique is wise. Of course, such a rigid system provides for legal certainty, and a prohibition is the strongest possible intervention on the developers and companies' fundamental rights, such as the freedom to conduct business and the right to property. And yet, the difficulties that the legislator had encountered with the sudden appearance of ChatGPT and other Large Language Models (LLMs) provide a strong case for the necessity to retain some regulatory flexibility. For instance, some Aritificial Intelligence researchers have convincingly argued that GPT-4 could reasonably be considered already an

54 *European Commission*, Coordinated Plan On Artificial Intelligence, Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions Fostering A European Approach To Artificial Intelligence, C(2021) 205 final.

55 *European Parliament, Council and European Commission,* European Declaration on Digital Rights and Principles for the Digital Decade 2023/C 23/01, OJ C 23, 23 January2023, pp. 1–7.

early form of Artificial General Intelligence (AGI).[56] According to them, the model demonstrates "sparks of intelligence" by showcasing capabilities beyond simple language understanding. Notably, GPT-4 achieves human-level performance in diverse tasks such as mathematics, coding, vision, medicine, law, and psychology, all without any specialized prompting. In my view, it would have been better to have kept some regulatory flexibility by inserting an opening clause, according to which additional prohibitions could have been laid down in delegated acts.

2. The prohibited practices

Article 5(1)(a) prohibits AI systems from using subliminal techniques or manipulative or deceptive methods to distort behaviour and impair informed decision-making, leading to significant harm and AI systems exploiting vulnerabilities due to age, disability, or social or economic situations.[57] In this respect, nothing major has changed compared to the Commission's proposal. Importantly, though, the legislator clarified that a technique is not only prohibited if the deployer intends to manipulate the user, but also if the system has the effect of manipulation.[58]

A relatively straightforward prohibition concerns the exploitation of vulnerabilities (Article 5 (1)(b)) and the use of a social scoring system (Article 5 (1)(c)). As regards individual predictive policing (Article 5(1)(d)), the co-legislators agreed to prohibit AI systems from making risk assessments of natural persons to assess or predict the risk of a natural person to commit a criminal offence, based solely on the profiling of a natural person, or on assessing their personality traits and characteristics. Not prohibited are 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.

Untargeted scraping of the internet or CCTV footage for facial images to build or expand databases is prohibited under Article 5(1)(e), thus closing a loophole in the initial text.[59] The co-legislators also agreed on a definition of sensitive operational data (operational data related to activities of prevention, detection, investigation and prosecution of criminal offences, the disclosure of which can jeopardize the integrity of criminal proceedings), which should be exempted from sharing obligations.

Under Article 5 (1)(g), the co-legislators agreed to prohibit biometric categorization systems that categorize natural persons individually, based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sexual life, or sexual orientation. This

56 *Bubeck* et al.

57 Critical towards the notion of harm in the AI context: *Neuwirth/Migliorini.*

58 See previous critique relating to the original wording: *Palka*, The Phantom Menace: A Critique of the European Commission's Artificial Intelligence Act Proposal, p. 4, available at: https://law.yale.edu/yls-today/yale-law-school-events/phantom-menace-critique-european-commissions-artificial-intelligence-act-proposal-przemyslaw-palka (12/4/2023); *Hacker*, EULJ 2021.

59 *Hoffmeister*, WHI-Paper 2023/1, p. 16.

prohibition does not cover any labelling, or filtering, or lawfully acquired biometric datasets, such as images based on biometric data or categorizing of biometric data in the area of law enforcement. Moreover, biometric categorization based on sensitive characteristics would be added to the list of high-risk use cases.

As to real time biometric identification (Article 5(1)(h)), the Commission's initial proposal had included a prohibition for law enforcement purposes, with limited exceptions (accepted by the Council and slightly narrowed down). The EP initially wanted a full ban, not limited to law enforcement, and without any exception, but agreed to a compromise. Thus, the final compromise sets that real-time remote biometric identification for law enforcement purposes in publicly accessible spaces remains prohibited with exceptions limited to

i) the targeted search for specific victims of abduction, trafficking in human beings and sexual exploitation of human beings, and search for missing persons;
ii) prevention of threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;
iii) localisation or identification of a criminal suspect or offender, of one out of 16 categories of crimes[60] listed in an Annex.

The co-legislators also wanted to ensure that no decision that produces an adverse legal effect on a person may be taken by the judicial authority or an independent administrative authority whose decision is binding solely based on the output of the remote biometric identification system. The use would be subject to prior authorisation by a judicial or independent administrative authority whose decision is binding. In case of urgency, authorisation can be obtained within 24 hours; if the authorisation is rejected, all data and output need to be deleted. Their use would need to be preceded by a fundamental rights impact assessment and should be notified to the relevant market surveillance authority and data protection authority. In case of urgency, the use of the system may be commenced without registration. There is an annual reporting obligation, both for Member States and for the Commission (based on aggregated data provided by the Member States).

Private uses of real-time biometric identification remain largely prohibited by the General Data Protection Regulation (GDPR). Member States may introduce, in accordance with Union law, more restrictive laws on the use of remote biometric identification systems. Recital 19 clarifies the notion of publicly accessible space, outlining what is, and what is not covered by the intention of the legislator. Finally, post remote biometric identification remains in the high-risk category, and is

60 The list refers to terrorism, trafficking in human beings, sexual exploitation of children and child pornography, illicit trafficking in narcotic drugs and psychotropic substances, illicit trafficking in weapons, munitions and explosives, murder and grievous bodily injury, illicit trade in human organs and tissue, illicit trafficking in nuclear or radioactive materials, kidnapping, illegal restraint and hostage-taking, crimes within the jurisdiction of the International Criminal Court, unlawful seizure of aircraft/ships, rape, environmental crime, organised or armed robbery, sabotage, participation in a criminal organisation involved in one or more crimes listed above.

thus not prohibited, but instead subject to the respective high-risk conditions. For example, deployers would need a prior authorisation by a judicial authority, or an independent administrative authority when using the system for investigations of a person convicted or suspected of a serious criminal offence. Each use of post remote biometric identification would also be subject to a notification obligation to the data protection and market surveillance authorities. Here too, Member States would remain free to introduce more restrictive laws.

3. Assessment

Generally, Chapter II strikes a good balance between law enforcement prerogatives and individual liberties. Ultimately, these final prohibitions with their exceptions and caveats are the result of legitimate political compromises. One remaining concern shall, however, be mentioned: In times where we observe democratic backsliding in some Member States such as Hungary, Slovakia (and previously Poland, where the PiS government used the Pegasus system to spy on political opponents[61]), and where the independence of the judicial system and the rule of law is not necessarily always guaranteed, the oversight of the law enforcement actions should not solely be left to Member States. Instead, a stronger Union level oversight mechanism should have been considered.

As for post systems, the problems remain the same: real-time and ex-post identification systems can violate citizens' fundamental rights in equally substantive ways. The European Digital Rights society (EDRi),[62] along with over 200 civil groups, the European Data Protection Supervisor (EDPS), the European Data Protection Board (EDPB),[63] the European Parliament, and the UN High Commissioner for Human Rights, have all raised alarms about the risks these technologies pose to privacy, equality, freedom of expression, and other democratic principles. Both types of systems can create a "chilling effect", deterring people from exercising their rights due to fear of repercussions, leading to a culture of fear and self-censorship. This could, for example, discourage individuals from participating in future demonstrations, regardless of whether the identification happens in real-time or after the event. In concreto, it is still unclear when real-time identification ends and when post-identification starts. This delimitation is crucial, however, as the legal thresholds for both types of remote biometric identifications differ substantially.

61 *Politico*, Poland launches Pegasus spyware probe, 19 February 2024, available at: https://www.politico.eu/article/poland-pegasus-spyware-probe-law-and-justice-pis-jaroslaw-kaczynski/ (24/4/2024).

62 With references to civil society and others: *EDRi* et al., The EU's Artificial Intelligence Act: Civil Society Amendments, 3 May 2022, available at: https://edri.org/our-work/the-eus-artificial-intelligence-act-civil-society-amendments/ (13/4/2023).

63 *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, p. 11.

III. High-risk AI systems (Chapter III)

1. Classification rules for high-risk AI systems (Section 1)

Chapter III is dedicated to so-called high-risk systems. There are two ways an AI system is considered high-risk. First, under Article 6(1), an AI system is considered high risk if it is a safety component of a product, itself the product or required to undergo third party conformity assessment pursuant to union harmonization legislation enumerated in Annex II. For example, if an AI system is incorporated into a toy, a lift, a watercraft, forestry vehicles, or personal protective equipment (all of those and many more falling under harmonising secondary union legislation), then the product would be considered a high-risk AI system. Secondly, if an AI system falls under one of the systems enumerated in Annex III it is also considered a high-risk system under Article 6(2).

However, both the Council and the Parliament considered these initial Commission's classification rules to be too burdensome and rigid. Therefore, both institutions proposed introducing a so-called "filter", ensuring that not too many AI-systems automatically fall under the extensive high-risk requirements laid out in Section II, even if they are not likely to cause serious fundamental rights violations or other significant risks.

The compromise found on the "filter" questions provides in Article 6(3) that Annex III AI systems will not be considered as high-risk, if they do 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. This would be the case if one or more of the following criteria are fulfilled:

(a) the AI system is intended to perform a narrow procedural task;
(b) the AI system is intended to provide accessory input for a review or to improve the result of a previously completed human activity;
(c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or
(d) the AI system is intended to perform a preparatory task to an assessment relevant for the purpose of the use cases listed in Annex III. The corresponding recital 53 explains the reasoning behind this criterion: If the AI system is used only for a preparatory task, the possible impact of the output of the system is deemed to be very low in terms of representing a risk for the assessment to follow. The recital gives examples for such AI systems, notably smart solutions for file handling or AI systems used for translation. An AI system, however, would always be considered high-risk if the AI system performs profiling of natural persons.

The compromise also provides, in Article 6(4), that a provider who considers that an AI system referred to in Annex III is not high-risk must document its assessment before that system is placed on the market or put into service. Such provider would

be subject to a registration obligation set out in Article 51(2). Upon request of national competent authorities, the provider would provide the documentation of the assessment.

Under Article 6(6), the Commission is empowered to adopt delegated acts in accordance with Article 73, to amend the criteria laid down in points a) to d) mentioned above, by adding new criteria or modifying the existing ones. The empowerment is restricted to situations in which there is concrete and reliable evidence of the existence of AI systems that fall under the scope of Annex III, but that do not pose a significant risk of harm to the health, safety, and fundamental rights. Moreover, the Commission shall adopt delegated acts, deleting any of the criteria laid down in the first subparagraph of paragraph 2a where there is concrete and reliable evidence that this is necessary for the purpose of maintaining the level of protection of health, safety, and fundamental rights in the Union. In any case, however, the compromise specifies that any amendment to the criteria laid down in points a) to d) should not decrease the overall level of protection of health, safety, and fundamental rights in the Union.

Generally, it is crucial to emphasize the importance of preventing any misuse of such filter systems, and thus the circumvention of high-risk classifications, as well as the need to ensure uniform application among Member States. Positively, the Commission will provide guidelines specifying the practical implementation of the classification, completed by a comprehensive list of practical examples of high risk and non-high risk use cases on AI systems. This could indeed ensure uniform application across the Union. Additionally, the requirement for providers that conclude their AI system is not high-risk, namely to keep documentation of assessments, and provide them upon request by the national authority, is a step in the right direction to prevent the misuse. Crucially, however, unlike in the Parliaments proposal, there is no empowerment for national or Union authorities to challenge this self-assessment. The Parliament had proposed for the national supervisory authority to be empowered to review and reply to the notification of the providers, directly or via the AI Office, within three months if they deem the AI system to be misclassified (Art. 6 (2a) EP-AIA). Such a challenge of the provider's self-assessment has not found its way into the final AI Act. One could therefore wonder how the Commission, or any national authority, wants to ensure efficient and effective safeguards against any misuse of the filter system.

It's striking that the AI industry's reliance on self-assessment for ensuring product compliance stands in contrast to practices in other technological regulations, like the Cyber Security Act (CSA).[64] In the CSA framework, for products deemed above low risk, it is the certification authorities who are tasked with checking that the requirements of the certification scheme are met.[65] Self-assessment does offer the benefit of utilizing the AI provider's deep knowledge of their own product

64 Regulation (EU) 2019/881on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), OJ L 151 of 7/6/2019.

65 *Casarosa*, ICLR 2022/3, p. 128.

to certify its conformity.[66] However, the possibility of incomplete or incorrect evaluations is a clear risk. Third-party assessments could serve as an alternative to the self-assessment model, but they carry the burden of potential administrative logjams, as they might result in a significant backlog of conformity checks. Such a situation could particularly hinder the early stages of AI innovation, thereby risking Europe's strategic objective of leading the world in the development of innovative and reliable AI.[67] Consequently, third-party assessments might not be the most practical solution.[68] Another option could be to incorporate a fundamental rights impact assessment within the self-assessment process. This would proactively ensure that developers consider the potential effects of their AI systems on fundamental rights. Yet, as we will see later, the effectiveness and feasibility of this solution are also open to scrutiny.

2. Requirements for high-risk AI systems (Section 2)

Once an AI system is classified as high-risk, the general requirements under Section 2 kick in. They are subject to an extensive risk-management system (Article 9) and a strict regime on data governance, documentation, and record-keeping rules (Articles 10–12). Importantly, the AI Act also requires that such systems are capable to be subject to human oversight (Article 14) and must achieve an appropriate level of accuracy, robustness, and cybersecurity (Article 15). The latter is an important point, as cybersecurity threats, notably adversarial attacks on data integrity (data poisoning and projected gradient scent algorithms), are increasing.[69] Interestingly in this respect, some have argued that new technologies, such as blockchains and distributed ledgers, could mitigate those threats. Indeed, blockchain technology could with its tamper-proof ledger and cryptographic safeguards help meet the AI Act's requirements. It could restrict AI's access to critical infrastructure through tamper-proof decentralised infracstrucutres and enable secure and transparent data sharing mechanisms through decentralised storage, augmenting data integrity and immutability in AI.[70] It remains to be seen how emerging technological solutions could be adjusted to help meet the legal requirements set out in the act.

3. Obligations for Providers, Deployers and other parties (Section 3)

Articles 16–27 lay down very specific obligations for providers and deployers (i.e. users who deploy an AI system). In that context, the second major development

66 *Raposo*, IJLIT 2022/30, p. 94.
67 *European Commission*, A European Approach to Artificial Intelligence | Shaping Europe's Digital Future, available at: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence (3/4/2024).
68 Strongly in favour of an ex ante licensing mechanism including third parties see: *Malgieri/Pasquale*, Computer Law & Security Review 2024.
69 Very instructive *Kalodanis* et al., Information & Computer Security 2023.
70 Very instructive *Ramos/Ellul,* Int. Cybersecur. Law Rev. 2024.

in the area of high-risk systems is the introduction of a fundamental rights impact assessment (or short "FRIA") under Article 27, which was pushed for mostly by the European Parliament. In order to efficiently ensure that fundamental rights are protected, the deployer of high-risk AI systems should carry out a FRIA prior to putting the system into use, to assess the reasonably foreseeable impact on fundamental rights arising from the specific context of use. The impact assessment should be accompanied by a detailed plan, describing the identified measures or tools that will help mitigate the risks to fundamental rights. If such mitigating measures cannot be identified, the deployer should refrain from putting the system into use. This obligation only concerns entities such as public sector organizations, private entities engaged in the provision of public services, and financial institutions including banks and insurance companies utilizing AI systems deemed high-risk as per Annex III, sections 5(b) and (c). Besides, these entities are required to conduct a FRIA exclusively for areas not already addressed by existing legal mandates, for instance, the Data Protection Impact Assessment stipulated by the General Data Protection Regulation (GDPR).[71] This requirement aims to ensure a seamless integration with current procedures, thereby avoiding redundancy and unnecessary complications. Additionally, to ease the compliance process, the AI Office is responsible for creating a standardized questionnaire template. This tool is designed to assist deployers in fulfilling the necessary criteria without undue difficulty.

Yet, despite the efforts to make it easy to comply with, such an additional assessment must still be critically questioned. Firstly, while it is designed to preemptively evaluate and mitigate any adverse effects on fundamental rights, one must ponder whether this process will be pragmatic and actionable, or if it would merely result in a bureaucratic exercise that fails to produce tangible benefits. The reality is that the risk management mechanisms of Article 9 may already cover a broad spectrum of the issues a FRIA aims to address. Ensuring that these mechanisms are robust and sufficiently comprehensive might be a more efficient approach than instituting an additional fundamental rights assessment.

Moreover, the actual benefit of such FRIA crucially depends on whether potential breaches of fundamental rights can be efficiently addressed. Any natural or legal person having grounds to consider that there has been an infringement of the provisions of the AI Act has the right to submit a complaint to the relevant market surveillance authority (Article 85). However, affected persons may not have the sufficient resources or expertise to actually launch a complaint. The true potential for ensuring compliance with fundamental rights may thus lie in collective enforcement. The integration of the AI Act into the Directive on representative actions for the protection of collective consumer interests[72] could have provided a more effective

71 See Art. 35 of Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119/1 of 4/5/2016.

72 Directive (EU) 2020/1828 on representative actions for the protection of the collective interests of consumers and repealing Directive 2009/22/EC, OJ L 409/1 of 4/12/2020.

enforcement mechanism.[73] This is because consumer associations have established themselves as influential watchdogs, demonstrating considerable success in areas regulated by the GDPR. By representing the collective interest, they can overcome the limitations faced by individuals, such as limited resources and lack of expertise, thus serving as a powerful conduit for enforcing fundamental rights within the context of AI.

Finally, complex legal questions arise when private organizations are required to comply with fundamental rights. Under specific conditions, the European Court of Justice (ECJ) has established a direct effect of certain Charter rights on private interactions.[74] And although Article 27(1) limits the application of the FRIA to entities regulated by public law and those offering public services, this provision still encompasses sectors such as banking, insurance, education, healthcare, and housing. It is likely that the question of the horizontal direct effect of fundamental rights will become a battleground for future litigation in the field.

4. Notification (Section 4)

Articles 28–39 lay down the new obligations for Member States to establish "notifying bodies". They are responsible for setting up and carrying out the necessary procedures for the assessment, designation, and notification of conformity assessment bodies and for their monitoring.

5. Standards, conformity assessments, certificates, registration (Section 5)

Finally, high-risk AI systems need to be certified. Section 5 entrusts the Commission with the task to carry out the relevant standardization requests to the European standardisation organisations. Systems which are in line with relevant standards shall be presumed to be compatible with the requirements of the AI Act (Article 40). However, since standards especially in the domain of AI will have a big say on Algorithmic design and capabilities, and are rarely purely technical but can also absorb commercial interests, political prefere; nces or moral judgments[75] the usual debate has emerged on whether "delegating" the standard setting to Standard-Setting Organisations (SSOs) is compatible with democratic legitimacy.[76]

73 *Fokuhl, Klöckner, Bomke*, Künstliche Intelligenz: Was das KI-Gesetz der EU für Verbraucher bedeutet, Handelsblatt, 12 February 2024, available at: https://www.handelsblatt.com/politik/deutschland/kuenstliche-intelligenz-was-das-ki-gesetz-der-eu-fuer-verbraucher-bedeutet/100013160.html (5/5/2024).

74 CJEU, case C-414/16, *Vera Egenberger v. Evangelisches Werk für Diakonie und Entwicklung e.V.*, ECLI:EU:C:2017:851 Generally welcoming a direct horizontal effect: *Ciacchi*, European Constitutional Law Review 2019/2, pp. 294 et seqq.; slightly more cautious *Ruffert*, JuS 2020, p. 1.

75 *Laux* et al., Computer Law & Security Review 2024, p. 2.

76 For further insights into standardisation and legitimacy and references to the debate, see *Almada/Petit,* Robert Schuman Centre for Advanced Studies Research Paper No. 2023/59, p. 23.

6. EU database for high-risk systems (Chapter VIII)

Somewhat misplaced, Chapter VIII creates an EU database for the high-risk systems in Annex III. From a legal drafting point of view, the relevant Article 71 could also have been added as another Section 6 to Chapter III.

IV. Transparency requirements for certain providers and deployers of certain AI systems (Chapter IV)

Chapter IV, consisting of a single Article 50, lays down a number of transparency requirements for both providers and deployers. Importantly, probably sensitized by the rapid increase of "deep fakes", which can be easily construed by even less sophisticated deployers, the latter have the duty to inform that they manipulated relevant material (Article 50(4)). This device seems to be very important to suppress relevant deep fake material, which may be circulating on the net in order to undermine the credibility of public persons or politicians. In this respect, this chapter thus also contributes to the overall objective to protect the democratic process in the Union and its Member States. Yet, although at times transparency and especially "explainability" of an AI system can help mitigate the risiks associated with complex AI systems (commonly referred to "eXplainable AI" or XAI)[77], some research highlights limitations in its effectiveness.[78] Particularly concerning are post-hoc explanation methods, which struggle to provide a clear understanding of how these systems arrive at decisions, especially when someone might try to manipulate them. This lack of clarity stems from the inherent complexity of black-box models, where a single "true reason" for a decision may not even exist. Even in simpler scenarios with standard algorithms, post-hoc explanations can be ambiguous and unreliable. These limitations make them potentially misleading in adversarial situations, where they might create a false sense of security by suggesting a level of justification or objectivity that isn't present. Therefore, despite the different attempts and techniques to meet legal requirements relating to transparency and explianability such as those laid out for the risk management system and high-risk AIs (see recital 65 and Art. 13 (1) and (3) lit. b (iv.) and Annex XI), the limitations should always be kept in mind.

V. General purpose AI (Chapter V)

Although with different approaches, both Parliament and Council introduced rules to address concerns arising with the use of General Purpose AI (GPAI) systems (term introduced in the General Approach Council position) and foundation models (mentioned in the Parliament version). The latter expression has been used in the past in relation to existing models, based on machine learning and transformers, while in the future other technical approaches may emerge. Hence, in chapter V,

77 *Pavlidis*, Law, Innovation and Technology 2024/1.
78 *Bordt,* p. 65.

the AI Act uses the more appropriate expression of "General Purpose" AI.[79] On substance, the co-legislators came to an agreement on regulating GPAI models with a two-tier approach, distinguishing models with and without systemic risk at Union level.

GPAI models with no systemic risks (low tier) should only be subject to specific information and documentation requirements (Article 53). This includes a technical documentation at the hands of the AI Office for the purpose of governance, and a limited set of information for downstream actors that would like to integrate the model and, hence, require good understanding of the model and further information (e.g. on the datasets) to fulfil their obligations that may stem from this or other regulations. The technical documentation (for the Commission) should include, among other things, information about the energy consumption of the model. The Commission is empowered to adopt delegated acts, detailing calculation and measurement methodologies.

GPAI models with systemic risks (high tier) should be those that are developed with at least 10^{25} FLOP computing power used. The number of registered business users can be considered (in combination with other indicators) by the AI Office when identifying models with systemic risks. Providers of GPAI models with systemic risks need to fulfil all baseline GPAI obligations (from the low tier), along with additional obligations under Article 55, including (1) the need for state-of-the-art model evaluations, including adversarial testing/red teaming, (2) general assessment and mitigation systemic risks and their sources, e.g from development and putting into service, (3) documentation and reporting of serious incidents and corrective measures and lastly (4) cybersecurity protection. The list of items for the technical documentation directed towards the commission is extended by further items to reflect these additional requirements, and to enable a more in-depth understanding. Also, the Commission will have the task to publish, and keep up to date, a list of all GPAI models with systemic risks. Providers of GPAI models with systemic risks will be subject to an obligation to carry out adversarial testing, a process where testers are hired to deliberately "attack" a system using the same methods a hacker might use to find and exploit weaknesses.

As is apparent, the set threshold of 10^{25} FLOP computing power is key. Some like *Novelli*[80] reasonably argue that this threshold is too high. Some, like *Hacker*[81] for instance, already signalized that this threshold currently only encapsulates GPT-4 but leaves aside AI models such as GPT-3.5, Claude and Bard. Lowering this threshold would better address other GPAI systems, which bring similar systemic risks.

79 Some authors have even called to remove the label generative AI althogether, as they deem it unnecessary, see *Novelli* et al., p. 6.

80 *Novelli* et al., p. 4.

81 *Hacker*, What's Missing from the EU AI Act: Addressing the Four Key Challenges of Large Language Models, VerfBlog, 13 December 2023, available at: https://verfassungsblog.de/whats-missing-from-the-eu-ai-act/, DOI: 10.59704/3f4921d4a3fbeeee (6/5/2024).

Additionally, the lawmakers agreed that providers of general AI models must create a policy to comply with EU copyright laws. They need to recognize and honor any copyright claims according to Article 4(3) of the Copyright Directive[82], and prepare a detailed summary of the training content for the AI model to share with the public. As for the relationship across the "value chain", providers of GPAI systems, which are to be integrated in a high-risk AI system, shall give information to downstream providers of AI systems which is necessary to comply with the Regulation, thus, addressing the fundamental question of allocation of responsibilities across the value chain.[83] By mandating that upstream providers furnish necessary compliance-related information, it facilitates a smoother alignment with regulatory standards across different stages of AI development and deployment. Yet, one needs to keep in mind that the reliance on upstream providers for compliance information places a great deal of trust in their hands, potentially creating a single point of failure in the compliance chain. If the information provided is inaccurate or incomplete, it could jeopardize the entire system's compliance, affecting multiple stakeholders. Also, this model assumes a linear value chain that may not reflect the complex, interconnected ecosystems in which AI systems are developed and deployed. In such networks, determining the "upstream" and "downstream" providers can be challenging, complicating compliance efforts.

Crucially, compliance with the legal requirements for providers of GPAI models could be demonstrated through compliance with relevant Codes of Practice to be developed under Article 56. The Codes of Practice should cover all the obligations for the respective model tiers and would be developed under coordination of the AI Office and could be approved by the Commission through an implementing act, following the example of the GDPR, and given validity within the Union. The drawing up of the codes of practice is an open process to which interested stakeholders will be invited, such as companies, civil society, and academia.

This approach is positive, firstly because it promotes transparency and inclusivity by inviting a broad range of stakeholders to participate in the creation of these codes, ensuring diverse perspectives and expertise are considered. Secondly, the model draws from the successful example of the GDPR and the Digital Services Act (DSA)[84], suggesting a robust framework for data protection and privacy that could enhance trust in AI technologies. Furthermore, the formal approval of these codes by the Commission not only legitimizes them, but also harmonizes standards across the Union, fostering a consistent and secure AI ecosystem. This process potentially accelerates the adoption of ethical AI practices, contributing to a more responsible and innovation-friendly environment. Taking into consideration inter-

82 Directive (EU) No. 2019/790 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130/92 of 17/4/2019.

83 *Engler/Renda*, Reconciling the AI Value Chain with the EU's Artificial Intelligence Act (CEPS, 30 September 2022), p. 2, available at: https://www.ceps.eu/ceps-publications/reconciling-the-ai-value-chain-with-the-eus-artificial-intelligence-act (15/4/2023).

84 Regulation (EU) 2022/2065 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277/1 of 19/10/2022.

national approaches like the Hiroshima process under the G7 will also prove to be critical to align the European regulatory framework to global efforts of regulating AI.

VI. Measures to support innovation (Chapter VI)

In order to satisfy the overall objective of the AI Act to also support innovation, Chapter VI requires Member States to establish "AI regulatory sandboxes". According to the definition in Article 3 point 55, these are "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, pursuant to a sandbox plan for a limited time under regulatory supervision". Articles 58–63 then lay down detailed arrangements for their functioning and the protection of data in a testing environment, thereby establishing a sort of innovation framework for AI development at EU level.

VII. Governance (Chapter VII)

As for the governance structure, mainly the competent national authorities will supervise the implementation of the new rules at the national level. The enforcement lies primarily with the competent national authorities, including the market surveillance authorities. Natural persons will have the opportunity to complain to the authorities, and, via the market surveillance regulation, natural persons have the opportunity to complain about the authorities' decisions.

The competent national authorities will be gathered at the European Artificial Intelligence Board. An advisory forum to the Artificial Intelligence Board will be established to gather stakeholder feedback from civil society, SMEs, start-ups, academia, and industry representatives.

Following the introduction of the rules on the general-purpose AI models, the AI Act envisages an AI Office to be established within the Commission (Article 64). Its tasks will ensure coordination at the European level and supervise the implementation and enforcement of the new rules on these most advanced AI models.

A proper governance and implementation of the AI Act necessitates the recruitment of leading AI experts. It is critical to recruit leading technologists, scholars, and visionaries with a profound understanding of AI, as opposed to the conventional cadre of European bureaucrats. However, attracting such expertise poses a significant challenge due to the fierce international competition for AI talent from major tech corporations, as well as recent AI-focused governmental initiatives in the US and the UK. The AI Office's initial budget allocation of €46.5 million appears modest in comparison to the £100 million dedicated to the UK's AI safety institute.

Therefore, a substantial increase in funding in the forthcoming EU budget to ensure that the Office is well-equipped seems unavoidable.[85]

VIII. Post-market monitoring, information sharing, market surveillance (Chapter IX) and codes of conduct, guidelines, penalties and final provisions (Chapters X–XIII)

The final chapters focus on post market monitoring processes by providers and monitoring plans for high-risk AI systems, as well as the information sharing obligations of serious incidents for providers of high-risk AI systems. The Regulation outlines the enforcement framework for AI system surveillance and control within the EU, integrating existing EU legislation to ensure comprehensive market oversight, mandates annual reporting by market surveillance authorities, and delineates the responsibilities for supervising high-risk AI systems, including those in financial services and sensitive sectors like law enforcement.

The Regulation emphasizes coordination between national and EU bodies for effective compliance enforcement, allowing for joint activities and investigations across Member States. It grants authorities access to essential documentation and data, including source codes under specific conditions, to assess AI system compliance. Provisions for real-world testing, supervision, and mutual assistance in supervising general-purpose AI systems are established, alongside procedures for dealing with non-compliant or misclassified AI systems. Confidentiality obligations are strictly outlined to protect intellectual property and ensure the integrity of regulatory processes.

Finally, as for the penalties, violations involving forbidden AI systems can attract administrative penalties up to €35 million or 7% of the company's global annual revenue from the last financial year, whichever is greater. Breaches related to the obligations of high-risk AI systems may lead to fines up to €15 million or 3% of the company's global annual revenue from the previous year, depending on which amount is greater. Providing false, incomplete, or misleading details to notified bodies and competent national authorities upon their request, could result in fines up to €7.5 million or 1% of the company's worldwide revenue for the last financial year, with the higher amount being applicable. For small and medium-sized enterprises (SMEs), including startups, the fines outlined in the AI Act will be capped at the lower of the specified percentages or amounts.

85 *United Kingdom Government*, Press Release, Initial £100 million for expert taskforce to help UK build and adopt next generation of safe AI, 24 April 2023, available at: https://www.gov.uk/government/news/initial-100-million-for-expert-taskforce-to-help-uk-build-and-adopt-next-generation-of-safe-ai (22/3/2024).

E. Conclusion

The European Artificial Intelligence Act is the cornerstone of the EU regulatory framework on AI. Other regulations such as the GDPR, The Digital Services Act and Digital Markets Act supplement the EU AI Act on a number of areas. The Act establishes a uniform and horizontal legal framework for AI to ensure the protection of fundamental rights and user safety. The risk-based approach is generally flexible to address future challenges through a set of established principles although in particular the closed list technique for prohibited AI systems may turn out being too rigid. The broad scope of the AI Act tries to export the Union's regulatory approach beyond its borders. Yet, some questions relating to national security and open-source exceptions still remain and may prompt future court decisions. In a similar vein, it remains to be seen whether the built in "filters" and FLOP thresholds will prove to be the right response to high-risk AI applications and General Purpose AI. As part of its new governing structure, the Act also establishes an AI Office, which, together with the national competent authorities, will be the first body globally to enforce binding rules on AI. All in all, the AI Act will help to formulate EU positions when it comes to finding a global framework for AI. It serves the EU's objective well to ensure that the underlying principles in the Act become a blueprint for the global debate, and that the AI Act serves as an international reference point. It therefore strengthens the EU's position in the global AI race that is in full swing.

Bibliography

ALMADA, MARCO; PETIT, NICOLAS, *The EU AI Act: a medley of product safety and fundamental rights?*, Robert Schuman Centre for Advanced Studies, Research Paper No. 2023/59, 2023, http://dx.doi.org/10.2139/ssrn.4308072

ALMADA, MARCO; ANCA, RADU, *The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy*, German Law Journal, 2024, pp. 1–18

BORDT; SEBASTIAN; *Explainable Machine Learning and its Limitations*, PhD Thesis, 2023, Universität Tübigen

BUBECK, SÉBASTIEN, CHANDRASEKARAN, VARUN, ELDAN, RONEN et al., *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, 2023, arXiv:2303.12712 [cs.CL]

BUNDE, TOBIAS; EISENTRAUT, SOPHIE; SCHÜTTE, LEONARD (eds.), *Munich Security Report 2024: Lose-Lose?*, Munich Security Conference, February 2024, https://doi.org/10.47342/BMQK9457

BRADFORD, ANU, *The Brussels Effect: How the European Union Rules the World*, 2020, New York, Online Edition

CASARO, FEDERICA, *Cybersecurity Certification of Artificial Intelligence: A Missed Opportunity to Coordinate between the Artificial Intelligence Act and the Cybersecurity Act*, International Cybersecurity Law Review, 2022, Vol. 3(1), pp. 1–16

COLOMBI CIACCHI, AURELIA, *The Direct Horizontal Effect of EU Fundamental Rights: ECJ 17 April 2018, Case C-414/16, Vera Egenberger v Evangelisches Werk Für Diakonie Und Entwicklung e.V. and ECJ 11 September 2018, Case C-68/17, IR v JQ.* European Constitutional Law Review, 2019, Vol. 15(2), pp. 294–305

HACKER, PHILIPP, *Manipulation by Algorithms. Exploring the Triangle of Unfair Commercial Practice, Data Protection, and Privacy Law*, European Law Journal, 2021, Vol. 29(1–2), pp. 142–175

HOFFMEISTER, KALOJAN, *The European Artificial Intelligence Act – remaining challenges for the legislature in the trilogue*, Walter Hallstein Institute Policy Paper, 2023, Vol. 1, pp. 1–41

KALODANIS, KONSTANTINOS; PANAGIOTIS; RIZOMILIOTIS; DIMOSTHENIS, ANAGNOSTOPOULOS, *European Artificial Intelligence Act: an AI security approach*, Information & Computer Security, 2023

LAUX, JOHANN; WACHTER, SANDRA; MITTELSTADT, BRENT, *Three pathways for standardisation and ethical disclosure by default under the European Union Artificial Intelligence Act*, Computer Law & Security Review, 2024, Vol. 53, 105957

MALGIERI GIANCLAUDIO; PASQUALE FRANK, *Licensing high-risk artificial intelligence: Toward ex ante justification for a disruptive technology*, Computer Law & Security Review, 2024, Vol. 52, 105899

MÜGGE, DANIEL, *EU AI Sovereignty: For Whom, to What End, and to Whose Benefit?*, Journal of European Public Policy, 2024, pp. 1–26

NEUWIRTH, ROSTAM JOSEF; MIGLIORINI, SARA, *Unacceptable Risks in Human-AI Collaboration: Legal Prohibitions in Light of Cognition, Trust and Harm*, ETHAICS@IJCAI, 2023

NOVELLI, CLAUDIO; CASOLARI, FEDERICO; HACKER, PHILIPP; SPEDICATO, GIORGIO; FLORIDI, LUCIANO, *Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity*, 2024, https://ssrn.com/abstract=4694565 or http://dx.doi.org/10.2139/ssrn.4694565

PAVLIDIS, GEORGIOS, *Unlocking the Black Box: Analysing the EU Artificial Intelligence Act's Framework for Explainability in AI*, *Law, Innovation and Technology, 2024, Vol.* 16(1), pp. 293–308

RAPOSO, VERA LÚCIA, *Ex Machina: Preliminary Critical Assessment of the European Draft Act on Artificial Intelligence*, International Journal of Law and Information Technology, 2022, Vol. 30(1), pp. 88–109

RAMOS, SIMONA; ELLUL, JOSHUA, *Blockchain for Artificial Intelligence (AI): enhancing compliance with the EU AI Act through distributed ledger technology. A cybersecurity perspective*, International Cybersecurity Law Review, 2024, Vol. 5(1), pp. 1–20

RUFFERT, MATTHIAS, *Privatrechtswirkung der Grundrechte – Von Lüth zum Stadionverbot – und darüber hinaus?*, Juristische Schulung (JuS), 2020, Nr. 1, pp. 1–6

SIEGMANN, CHARLOTTE; ANDERLJUNG, MARKUS, *The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market*, Centre for the Governance of AI, 2022, ArXiv abs/2208.12645

SMUHA, NATALIE; AHMED-RENGERS, EMMA; HARKENS, ADAM; WENGLONG, LI, MACLAREN, JAMES; PISELLI, RICCARDO; YEUNG, KAREN, *How the EU Ca Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for Artificial Intelligence Act*, Leads Lab University of Birmingham, 2021, pp. 1–64

WÖRSDÖRFER, MANUEL, *Mitigating the adverse effects of ai with the European Union's artificial intelligence act: Hype or hope?*, Global Business and Organizational Excellence, 2023, Vol. 43(3), pp. 106–126