European Journal of Risk Regulation (2025), **16**, 684–703
doi:10.1017/err.2024.78

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act

Martin Ebers

University of Tartu, Tallinn, Estonia
Email: ebers@ai-laws.org

Abstract

The Artificial Intelligence Act (AI Act) of the European Union (EU) claims to be based on a risk-based approach to avoid over-regulation and to respect the principle of legislative proportionality. This paper argues that risk-based regulation is indeed the right approach to AI regulation. At the same time, however, the paper shows that important provisions of the AI Act do not follow a truly risk-based approach. Yet, this is nothing that cannot be fixed. The AI Act provides for sufficient tools to support future-proof legislation and to implement it in line with a genuine risk-based approach. Against this background, the paper analyses how the AI Act should be applied and implemented according to its original intention of a risk-based approach, and what lessons legislators around the world can learn from the AI Act in regulating AI.

Keywords: AI Act; AIA; artificial intelligence; EU; regulation; risk-based regulation

I. Introduction

The AI Act[1] is the world's first attempt at comprehensively regulating AI. As is well known, the regulation is claimed to follow a risk-based approach – one that tailors the choice and design of regulatory instruments based on the level of risk, according to the rule: "the higher the risk, the stricter the rules." To this end, the AI Act distinguishes four risk categories (unacceptable, high, limited, and minimal), defining regulatory requirements based on the risks posed by AI systems.[2]

Martin Ebers is President of the Robotics & AI Law Society (RAILS) and Professor of IT Law at the University of Tartu, Estonia. The author gratefully acknowledges the support of the Computer & Communications Industry Association (CCIA Europe) for this research paper, as well as the valuable help and insights provided by the Robotics & AI Law Society (RAILS) and the Munich Institute of Robotics and Machine Intelligence (MIRMI) at the Technical University of Munich (TUM). The views expressed in this paper reflect the author's own expertise and are based on independent academic research. All internet sources referred to were last accessed on September 21, 2024.

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

[2] In addition, during the negotiation of the AI Act, the co-legislators added the category of "general purpose AI models". However, as will be shown below (Section III.5.), this new category is inconsistent with a truly risk-based approach.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

Recital (26) AI Act points out that:

> "In order to introduce a *proportionate* and *effective* set of binding rules for AI systems, a clearly defined *risk-based approach* should be followed. That approach should *tailor the type and content* of such rules to the *intensity* and *scope of the risks* that AI systems can generate. It is therefore necessary to prohibit certain unacceptable AI practices, to lay down requirements for high-risk AI systems and obligations for the relevant operators, and to lay down transparency obligations for certain AI systems."[3]

Hence, the underlying objective of the AI Act's risk-based approach is to strike an optimal (or proportionate) balance between innovation and the benefits of AI systems on the one hand, and the protection of fundamental values such as safety, health and fundamental rights on the other.[4]

Recital (26) of the AI Act refers, particularly, to the principle of (legislative) proportionality enshrined in Article 5(4) TEU. According to this article, the "content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties." Further, the article requires institutions of the Union to "apply the principle of proportionality as laid down in the Protocol on the application of the principles of subsidiarity and proportionality." This protocol, in turn, states that draft legislative acts shall be justified with regard to the principles of subsidiarity and proportionality, taking into account "any burden, whether financial or administrative, falling upon the Union, national governments, regional or local authorities, economic operators and citizens, to be minimised and commensurate with the objective to be achieved."[5]

Against this backdrop, the AI Act's risk-based approach can be seen as a legislative technique for promoting a proportionate system of duties and obligations.

While a risk-based regulation is indeed the right approach to AI regulation, this analysis shows that important provisions of the AI Act *do not follow a truly risk-based approach.* Yet, this is nothing that cannot be fixed. The AI Act provides for sufficient tools to support future-proof legislation and to implement it in line with a genuine risk-based approach.

To further elaborate this, the following sections are structured as follows: Section II outlines the key elements of risk-based regulation – discussing the notion of "risk," the distinction between AI risk assessment, impact assessment, and risk management and the key elements of risk-based regulation. Section III criticizes the AI Act, arguing that some of its main provisions are not truly risk-based, leading to over-regulation in some areas and under-regulation in others. In particular, it analyses several problems with the AI Act, such as the lack of a risk-*benefit* analysis, limited reliance on empirical evidence, and lack of case-by-case risk classification. Section IV examines how the AI Act can be adjusted to better portray elements of a risk-based approach. To this end, the paper analyses the relevant instruments to implement the AI Act, such as guidelines, delegated and implementing acts, codes of practice, and harmonised standards. Finally, Section V draws conclusions on what lawmakers outside the EU can learn from the AI Act in regulating AI.

[3] Emphases added.

[4] G De Gregorio and P Dunn, "The European Risk-based Approaches: Connecting Constitutional Dots in the Digital Age" (2022) 59 Common Market Law Review 473, 499.

[5] Consolidated version of the Treaty on the Functioning of the European Union – Protocol (No 2) on the application of the principles of subsidiarity and proportionality, Art. 5, OJ C 115, 9.5.2008, p 1.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

II. Key elements of risk-based regulation

1. The notion of risk

The emergence of risk-based approaches to regulating AI systems in the EU and elsewhere in the world[6] raises, first and foremost, the question as to what policymakers (and the EU in particular) mean when they talk about risk. Generally speaking, "risk" is the likelihood that a source of hazard will turn into actual harm.[7] Based on this understanding, Article 3 No. 2 AI Act defines risk as "the combination of the probability of an occurrence of harm and the severity of that harm." As such, risks are usually distinguished from uncertainties.[8] While risks are "known knowns" with statistical probabilities and quantifiable effects, uncertainties are "known unknowns" that cannot be quantified because we do not know what the effects of a particular technology might be. In addition, there are "unknown unknowns," where we are not even aware that things or activities may have adverse effects at all.[9]

One of the most pressing issues for any risk-based approach is how to reach consensus on what risks to watch out for and how serious they are considered to be (either in terms of probability or impact, or both).[10] There was a lively debate, during the negotiations on the AI Act, about the criteria for determining what AI practices should be banned in the EU, as well as which AI systems should be classified as "high-risk" and thus allowed on the market if certain safeguards are put in place.[11] This illustrates how difficult this endeavour can be; especially if this assessment is not sufficiently based on empirical evidence and a sound methodology.[12]

2. The elements of risk-based regulation: risk assessment and categorisation, impact assessment and risk management

Typically, risk-based approaches to (AI) regulation consist of various elements or phases, namely – risk assessment and categorisation, impact assessment and risk management.[13]

At a high level, we can first distinguish between (i) the assessment of risks arising from the use of AI and (ii) the classification of AI systems or applications by risks.[14] The first type assesses the risks posed by the use of AI, which may include risks to safety and health, bias and discrimination, lack of fairness, lack of transparency, invasion of privacy and data protection rights, or other protected interests. In the second type, the assessor looks at the risks associated with the use of AI in order to classify the system into a category of risk.

[6] Risk-based regulation has become a dominant strategy for policymakers on AI – not only in the EU, but globally – both at the international and national levels, and in the work of (international) standard-setting bodies. For an overview cf. EY, Trilateral Research, "A survey of artificial intelligence risk assessment methodologies: The global state of play and leading practices identified" (2022) https://www.trilateralresearch.com/wp-content/uploads/2022/01/A-survey-of-AI-Risk-Assessment-Methodologies-full-report.pdf.

[7] Frank Knight, Risk, Uncertainty, and Profit 19–20, 233 (1921); Baruch Fischhoff, Sarah Watson and Chris Hope, Defining risk (1984) 17 *Policy Science* 123–139 <https://doi.org/10.1007/BF00146924>.

[8] Knight, ibid.

[9] Julia Black, "The role of risk in regulatory processes" in Robert Baldwin, Martin Cave and Martin Lodge (eds), *The Oxford Handbook of Regulation* (Oxford University Press 2010) 302–348, 310.

[10] Ibid, 311.

[11] Luca Bertuzzi, "AI Act: EU Parliament's crunch time on high-risk categorisation, prohibited practices" (2023) Euractiv <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-parliaments-crunch-time-on-high-risk-categorisation-prohibited-practices/>.

[12] On the limited reliance on empirical evidence cf. below, Section III.3.

[13] Claudio Novelli, Federico Casolari, Antonino Rotolo et al. AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act (2024) 3 *DISO* 13. https://doi.org/10.1007/s44206-024-00095-1; EY, Trilateral Research, supra n 6.

[14] Cf. EY, Trilateral Research, supra n 6, p 9 et seq.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

This classification process helps stakeholders prioritise their efforts and allocate resources more effectively, contributing to the determination of obligations in line with the extent of risk posed by AI systems.

An impact assessment, on the other hand, goes further than a risk assessment. While a risk assessment is about the identification, analysis and evaluation of AI-related risks, an impact assessment seeks to evaluate the wider impact of AI systems on several stakeholders – including users, society, and the environment, going beyond the mere discovery and analysis of risks. This usually entails a review of governance, performance, communication, threats to safety and security, and other protected interests.

Based on such an impact assessment, risk-based approaches to AI regulation usually also contain requirements for a risk management, which includes the determination, evaluation, and ranking of risks related to AI as well as putting policies in place to reduce, track, and manage the possibility of unforeseen events.

In the AI Act, all of the above-mentioned elements of risk-based regulation are present.[15] The AI Act *assesses* the risks posed by the use of AI and *categorises* them (mostly on a top-down basis[16]) as unacceptable, high risk, limited risk, and minimal (or no) risk. Moreover, Article 9 AI Act requires providers of high-risk AI systems to establish a *risk management system* which includes an *impact assessment* to identify and analyse foreseeable risks that the high-risk AI system may pose to health, safety, or fundamental rights, in order to take appropriate and targeted risk management measures designed to address the identified risks.

3. *Essential features of risk-based regulation*

Risk-based approaches to AI regulation can take different forms, depending on the ultimate goal of the regulator. As *Coglianese*[17] points out, such a regulation may aim to:

- Eliminate all risk (the zero-risk approach)
- Reduce risk to an acceptable level (the acceptable risk approach)
- Reduce risk until costs become unbearable (the feasibility approach) or
- Strike a balance between risk reduction and costs of regulation (the proportionate or efficiency approach).

As discussed above,[18] the AI Act is based on the idea of proportionate regulation – aimed at striking an optimal (or proportionate) balance between reducing the risks posed by the use of AI systems on the one hand, and innovation and the benefits of AI systems (or the costs of regulation) on the other.

Accordingly, the question arises as to what key elements a legislator must consider in adopting such an approach. Arguably, these elements include:

[15] Cf. also Tobias Mahler, "Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal" in Luca Colonna and Rolf Greenstein (eds), Nordic Yearbook of Law and Informatics 2020–2021: Law in the Era of Artificial Intelligence (The Swedish Law and Informatics Research Institute 2022) 249 et seq. https://irilaw.files.wordpress.com/2022/02/law-in-the-era-of-artificial-intelligence.pdf.

[16] In the AI Act, risk categorisation is mostly top-down, as it is the AI Act itself (and to a certain extent the European Commission, cf. Art. 7 AI Act) that decides into which risk category a particular AI system falls. However, with the so-called "additional layer" provided for in Art 6(3)–(4) AI Act, providers of systems listed in Annex III have the possibility to demonstrate (and document) that their AI system is not high-risk.

[17] Cary Coglianese, "The Law and Economics of Risk Regulation" (2020) University of Pennsylvania, Institute for Law & Economics Research Paper No. 20-18, 9 https://scholarship.law.upenn.edu/faculty_scholarship/2157/.

[18] See I.1.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

- **Risk-*benefit* analysis:** When assessing the risks of AI, it is necessary to look, beyond the possible harms that AI systems can cause, at their innovative economic and social benefits. After all, "risk" is something we take in the name of benefit; we don't typically choose to be harmed. Instead, we – as a society – choose to take certain risks in the name of current and potential societal gains.[19] Therefore, in order to assess which risks are acceptable and which risks present the possibility of unacceptable harm, a consistent application of the risk-based approach requires a thorough consideration of, not only the negative consequences, but also the positive contributions that AI brings to individuals and society. Such a risk-benefit analysis must include, in particular, the (opportunity) costs of underuse. As the European Parliament already pointed out in 2020, this underuse can also be considered a "major threat." Missed opportunities for the EU to use AI systems "could mean poor implementation of major programmes, such as the EU Green Deal, losing competitive advantage towards other parts of the world, economic stagnation and poorer possibilities for people."[20]
- **Technology Neutrality**: A true risk-based approach regulates the risks of applications, not the technology itself. This principle of "technology neutrality" has been recognised by many regulators around the world,[21] including the EU,[22] as an overarching principle for ICT regulation. The main aim of this principle is to ensure an equal treatment of technologies with equivalent effects, and to future-proof the law, i.e. to draft legislation in a way that is flexible enough not to impede future technological development and to avoid the need for constant legislative revision.[23]
- **Evidence-based Risk Assessment and Categorisation**: Another important feature of risk-based regulation is that the assessment and classification of risks requires sufficient empirical evidence and a clear methodology. As AI-related risks are being used to justify governmental regulation, there needs to be a common way how to assess and classify these risks. Accordingly, regulators, standard bodies and other stakeholders have been working for many years on risk assessment frameworks and science-based methodologies. Arguably, risk-based regulation works best on quantifiable problems. However, many harms are either not quantifiable at all, or represent a mixture of quantifiable issues with hidden policy choices.[24] In such cases, the question arises as to whether a risk-based approach is appropriate at all or whether other regulatory techniques (such as a rights-based approach) should be adopted instead.[25]

[19] Margot Kaminski, "Regulating the Risks of AI" (2023) 103 Boston University Law Review 1347, https://ssrn.com/abstract=4195066.

[20] European Parliament, "Artificial intelligence: threats and opportunities" (23 September 2020, last updated 20 June 2023) https://www.europarl.europa.eu/topics/en/article/20200918STO87404/artificial-intelligence-threats-and-opportunities. See also Ugo Pagallo and others, "The underuse of AI in the health sector: Opportunity costs, success stories, risks and recommendations" (2024) 14 Health and Technology 1 https://doi.org/10.1007/s12553-023-00806-7.

[21] As to the origins of this principle, see Anne Veerpalu, "Regulatory challenges to the use of distributed ledger technology: Analysis of the compliance of existing regulation with the principles of technology neutrality and functional equivalence" (PhD thesis, University of Tartu 2021) 30 https://dspace.ut.ee/bitstreams/12ad2896-93f2-4d23-ac81-c28d50c9f25e/download.

[22] See, for example, recital (15) GDPR; recital (10) Digital Content and Services Directive 2019/770.

[23] Bert-Jaap Koops, "Should ICT Regulation be Technology-Neutral" in Bert-Jaap Koops and others (eds), *Starting Points for ICT Regulation: deconstructing prevalent policy one-liners* (2006); Chris Reed, "Taking Sides on Technology Neutrality" (2007) 4(3) SCRIPTed 263; Brad Greenberg, "Rethinking Technology Neutrality" (2016) 100 Minnesota Law Review 1495 https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1206&context=mlr.

[24] Kaminski, supra n 19, 32.

[25] This issue is particularly relevant in the context of fundamental rights, see below, at III1.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

- **Proportionate Regulatory Burden**: Ideally, obligations and other regulatory burdens should be proportionate to the risks posed by AI applications to ensure that regulatory requirements are aligned with the potential harm and impact of AI systems. Risk-based regulation, therefore, seeks to create a legal framework in which legal obligations are tailored to the specific risks posed by the use of a particular AI system for a given purpose, in order to avoid overburdening of the regulated actors.
- **Flexibility and adaptability**: Risk-based regulation must also be flexible enough to adjust retrospectively if it turns out that the original risk assessment or categorisation was wrong. As Black puts it, "[r]esponding to risks and attempting to manage them necessarily involves anticipating the future" which by its very nature is unknown.[26] Especially with new technologies such as AI, it is impossible to reliably assess the risks and benefits of AI systems deployed in given areas ex ante. For this reason, a truly risk-based regulation must require the legislator, the regulator and those who manage and mitigate risks to monitor the performance of AI systems throughout their lifetime, periodically re-evaluating risks and implementing the necessary corrections.

Clearly, some caveats are necessary. First, it is important to note that the criteria listed above are by no means exhaustive. Also, it is essential to remember that risk-based approaches to regulation have their difficulties – such as how to specify, aggregate and quantify risks, how to reconcile conflicting values and how to set levels of acceptable risk – as well as limitations – such as an overreliance on quantification, how to deal with unquantifiable and, in particular, unknown risks, and how to take into account the risk of harm to individuals.[27] As a result, risk-based regulation usually needs to be complemented by additional set of rules.

These legitimate concerns should not, however, be used as an overall argument against this type of regulation. In fact, risk-based regulation has a number of strengths when properly implemented.[28] First, it rationalises government intervention by setting clear priorities and objectives. Second, it facilitates the effective use of scarce resources and allows regulators – if implemented according to a true risk-based approach – to focus compliance efforts on products/services and/or systems that pose the greatest risks. Last, but not least, risk-based regulation can be an effective tool for striking the right balance between the benefits and risks of a particular technology.

III. Is the AI act a truly risk-based regulation?

While risk-based regulation is indeed the right approach to AI regulation, important elements of the AI Act do not follow a truly risk-based approach.

1. Protecting fundamental rights with a risk-based approach?

One fundamental problem is that the AI Act with its risk-based approach seeks to protect – not only health and safety, but also – the fundamental rights of citizens. This is troubling for a number of reasons.

First, the European Union has no general competence to harmonise Member State's laws to protect human rights. As a result, the AI Act "shoehorns"[29] the protection of

[26] Black, supra n 9, 317.

[27] Kaminski, supra n 19, 32.

[28] Claudio Novelli et al, "AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act" (2024) Digital Society 8, https://doi.org/10.1007/s44206-024-00095-1.

[29] Marco Almada and Anca Radu, "The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy" (2024) *German Law Journal* 1-18, 3, https://doi.org/10.1017/glj.2023.108.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

fundamental rights into the scope of Article 114 TFEU, which gives the EU the competence to remove barriers to trade in the internal market. However, such an approach is very likely to fail, because it does not have the protection of rights as its primary goal, but rather the opening and shaping of markets.[30]

Moreover, the EU's decision to protect human rights in the AI Act primarily through a risk-based approach, rather than of a rights-based approach,[31] is generally ill-suited. Most importantly, such an approach neglects the minimum and non-negotiable nature of human rights. Instead, the AI Act, with its risk-based approach and its fundamental rights impact assessment, implies that fundamental rights violations can be quantified and measured in degrees. This is, however, not the case. As *Yeung* & *Bygrave* point out, while it is possible to speak of different levels of culpability, scale, and magnitude when talking about fundamental rights, "these variations do not imply that fundamental rights violations can be, without problems, ranked on a sliding scale from trivial to serious."[32] Instead, fundamental rights follow a binary logic in that an activity is either legal or illegal.[33]

Finally, risk is typically assessed at the level of the collective/society and not for the individual. Rather than preventing individual harm, risk thinking assesses harm at a social level or a society-wide scale. One of the consequences of this aggregate nature of risk is that individual differences are typically ironed out, as risk analysis often determines acceptable risks by looking at the average citizen.[34] Another consequence is that risk-based regulation often involves society-wide trade-offs (e.g. between fairness and efficiency), with the result that even immense individual harms may be dismissed.[35]

For all these reasons, a risk-based approach is difficult to reconcile with the protection of fundamental rights.

2. *Missing risk*-benefit *analysis*

Another problem with the AI Act is that it lacks a risk-*benefit* analysis – a fundamental component of a truly risk-based approach to regulation.

As outlined above,[36] a risk-*benefit* analysis involves assessing the potential risks and benefits of a particular action or technology to determine whether the benefits outweigh the risks. In addition to determining whether a system has the potential to cause harm and the severity of the likely harm, the analysis must also consider the benefits and (likely) positive outcomes of using a system – such as advancing scientific discovery.[37] Otherwise, there is no appropriate framework for a proportionate and balanced regulatory regime.

In sharp contrast to this, the AI Act does not consider the potential benefits of AI

[30] Hans-W Micklitz and Dennis Patterson, "From the Nation State to the Market: The Evolution of EU Private Law" (1 June 2012) EUI Working Papers LAW No 2012/15, https://ssrn.com/abstract=2115463; Almada and Radu (n 29).

[31] The AI Act contains only rudimentary individual rights, namely (i) a right to lodge a complaint with a market surveillance authority (Art. 85 AI Act), and (ii) a right to explanation of individual decision-making (Art. 86 AI Act).

[32] Karen Yeung and Lee Bygrave, "Demystifying the modernized European data protection regime: Cross-disciplinary insights from legal and regulatory governance scholarship" (2022) 16 Regulation & Governance 137–55, 146, https://doi.org/10.1111/rego.12401.

[33] Raphael Gellert, "We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection" (2016) EDPL 481–92, 483.

[34] Kaminski, supra n 19, 1392.

[35] Kaminski, supra n 19.

[36] Section II.3.

[37] London Borough of Waltham Forest, *Risk Assessment & Risk-Benefit Analysis* (London: LBWF Early Years, Childcare & Business Development Service, 2019) 5.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

systems alongside the risks they pose. Instead, it focuses primarily on preventing risks and threats to health, safety and fundamental rights, without considering the potential positive impacts of AI systems.

Besides undermining the idea of truly risk-based regulation, this approach ignores the positive contributions of technology and may also result in missed opportunities for societal progress and innovation. In other words, by not considering the positive aspects of the technology in the regulatory framework, the AI Act fails to harness the potential of AI systems to improve the common good.

This absence of a risk-*benefit* analysis in the AI Act also makes it difficult to strike an appropriate balance between potential risks and benefits in dilemma situations. This is illustrated by an example from health care,[38] where it is currently unclear whether AI-based medical devices should have a minimum level of transparency before they can be released to the market. Some data scientists argue that regulators should only allow inherently interpretable algorithmic models while banning AI systems with algorithmic opacity that cannot be technically resolved.[39] However, studies show that some opaque AI systems (e.g. deep neural networks) have a much higher degree of accuracy and efficiency than transparent systems (e.g. deductive and rule-based systems).[40] In such a situation, a trade-off between AI's accuracy and transparency must be made.[41]

The Medical Device Regulation (MDR)[42] provides for exactly such balance, allowing certain risks to be recognised as acceptable if they are outweighed by the corresponding benefits (Annex I No 4 MDR). Thus, the inherent algorithmic opacity of a medical device may be considered an acceptable risk, if the manufacturer can demonstrate that the benefits of using such a device outweigh the risks.

Whether such a trade-off is also possible under the AI Act – which (in the case of AI based medical devices requiring a third party conformity assessment) applies simultaneously to the MDR – is unclear, given that the AI Act considers only possible risks and their prevention. Therefore, the AI Act does not provide a clear answer to the question of whether a certain degree of algorithmic opacity can be considered an acceptable risk in light of the benefits of using AI. Obviously, the intention of the EU legislator was not to eliminate all risks (the zero-risk approach), but to strike a balance between risk reduction and costs of regulation (the proportionate or efficiency approach), as discussed in Section I. Therefore, a truly risk-based implementation of the AI Act requires striking a balance between algorithmic transparency and efficiency/accuracy.

[38] See also Anastasiya Kiseleva, Dimitris Kotzinos and Paul De Hert, "Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations" (2022) 5 Frontiers in Artificial Intelligence 1, 16 <www.frontiersin.org/articles/10.3389/frai.2022.879603/full>, 11.

[39] Cynthia Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead" (2019) 1(5) Nature Machine Intelligence 206 <https://doi.org/10.1038/s42256-019-0048-x>.

[40] Rich Caruana and others, "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission" (2015) *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1721 <http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf>; Bernhard Waltl and Roland Vogl, "Explainable Artificial Intelligence – The New Frontier in Legal Informatics" Jusletter IT (22 February 2018).

[41] Martin Ebers, 'Regulating AI and Robotics' in Martin Ebers and Susana Navas (eds), *Algorithms and Law* (Cambridge 2020) 37-99, 49 et seq.

[42] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices [2017] OJ L 117/1.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

3. Limited reliance on empirical evidence

Another reason why the AI Act is not truly risk-based is its limited reliance on empirical evidence for the design of the different risk categories. As scholars have pointed out,[43] the AI Act does not establish criteria for when AI poses an unacceptable risk to society and individuals. Instead, it merely provides a set list of categories of AI systems that are considered to pose "unacceptable risks" and are therefore banned in the EU.

For high-risk AI systems, Article 7(2) AI Act sets out the criteria to be taken into account by the European Commission when amending the list of Annex III high-risk AI systems. However, neither the recitals of the AI Act nor any accompanying EU document explain and justify how these criteria were applied to identify the areas listed in Annex III in the first place. As *Grozdanovski and De Cooman* conclude, when choosing which risks to address, "regulators were generally disinterested in statistical evidence on the possibly harmful features of various systems."[44]

This applies both to the Commission's original proposal which failed to gather empirical evidence for the design of the AI Act, and to the subsequent legislative process.

Instead of conducting its own practical studies in concrete use cases of AI, the European Commission relied heavily on public consultations[45] for its proposal, despite acknowledging that "robust and representative evidence for harms inflicted by the use of AI is scarce due to the lack of data and mechanisms to monitor Al as a set of emerging technology."[46] Inconsistencies in reported participation numbers as well as the methods employed during the consultation – in particular, the use of close-ended questions and pre-suggested answers – cast further doubt on the accuracy and representativeness of the data collected.[47] Moreover, there is a discrepancy between the results of the consultation and the final proposal. In certain areas, the proposal deviates from the consensus expressed by the respondents.[48]

The political nature of the definition of risk categories was also evident in the subsequent trilogue negotiations. The Council and the European Parliament proposed new prohibited AI practices and new areas for high-risk AI systems, but with little or no justification as to why these were chosen.[49]

All these lead to the conclusion that the supposedly "risk-based" nature of the Act is neither based on practical evidence nor justified by externally verifiable criteria, but is the result of a political compromise at a particular point in time and is therefore largely arbitrary.[50]

4. Pre-defined, closed risk categories

A closely related problem concerns the framing of the risk categories themselves.

At first glance, many of the pre-defined categories do not make sense. For instance, the AI Act does not apply to the most dangerous applications – such as killer robots (Article 2(3) AI Act); AI systems developed in the EU to provide support to dictators or hackers

[43] Lilian Edwards, *Regulating AI in Europe: Four Problems and Four Solutions* (Ada Lovelace Institute, March 2022) 11 https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf.

[44] Ljupcho Grozdanovski and Jerome De Cooman, "Forget the Facts, Aim for the Rights! On the Obsolescence of Empirical Knowledge in Defining the Risk/Rights-Based Approach to AI Regulation in the European Union" (2023) 49 Rutgers Computer & Tech LJ 207.

[45] For details see Grozdanovski and De Cooman, supra n 44, 236 et seq.

[46] European Commission, *Impact Assessment*, SWD(2021) 84 final, Part 1/2.

[47] Grozdanovski and De Cooman, supra n 44, 239.

[48] Grozdanovski and De Cooman, supra n 44, 240.

[49] Edwards, supra n 43, 11.

[50] Edwards, supra n 43, 11; Grozdanovski and De Cooman, supra n 44.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

outside the EU (Article 2(1)(c) AI Act);[51] and autonomous vehicles, drones/airplanes and vessels (Article 2(2) in conjunction with Annex I.B. AI Act).

On the other hand, many applications qualify as high-risk AI systems under Annex III, simply because they are used in a particular sector, even though they do not pose a serious risk of harm, such as tools to detect duplicates in datasets or tools to improve language.[52] While it is true, that in both of these cases providers have the possibility under Article 6(3) *sub* 2(a)-(b) AI Act to demonstrate that their systems do not qualify as high-risk, it does not change the fact that these tools can be covered by Annex III and only exceptionally exempted, which places the burden of proof (and documentation, Article 6(4) AI Act) on the provider.

At a more fundamental level, the examples discussed point to the real problem. The AI Act provides a broad and rather abstract classification of high-risk systems under Annex III. Instead of providing a risk classification on a case-by-case basis, it uses a pre-defined (closed) list of typical high-risk applications. Whether an AI system used in a specific sector for specific purposes poses a high risk to health, safety and/or fundamental rights, is not assessed for the *concrete* risk, but is pre-defined for typical cases in Annex III. Accordingly, the risk management system required by Article 9 AI Act is only obligatory in situations that are already classified by the AI Act as high-risk cases. As a result, the risk management obligations of providers under the AI Act consist mainly of risk mitigation rather than risk assessment.[53]

The choice of such a top-down regulation raises several issues. First, this approach leads to over-regulation where, for instance, an AI system falls into one of the eight categories listed in Annex III, but in reality does not pose a significant risk of harm. Second, the list of typical high-risk AI systems (albeit with broad definitions and open to updating) may not be easy for the European Commission to keep up to date in a timely manner, given how rapidly AI technology is evolving.[54] Moreover, the decision to delegate (to the Commission) the power to amend Annex III by adding, modifying and removing high-risk AI systems (Article 7 AI Act) raises concerns in terms of power allocation.[55]

Finally, the focus on a pre-defined list of high-risk AI systems also creates a sharp rift between this category and other lower-risk categories that are largely unregulated (with the exception of transparency requirements, Article 50 AI Act). In particular, such a rigid distinction is not justified in cases where an AI system is used in a specific sector (e.g. healthcare sector) but does not qualify as high-risk (e.g. because the system does not qualify as a medical device according to Article 6(1), Annex I.A.11 AI Act and the MDR),[56] but nevertheless poses numerous risks (e.g. to patients and care recipients due to its direct and indirect effects on the human body and mental health).[57]

[51] However, cf. also Martin Ebers and others, "The European Commission's Proposal for an Artificial Intelligence Act," Journal "J" (2021) 4, 589–603, 591, https://doi.org/10.3390/j4040043: Exclusion is justified in light of the fact that the AI Act is based on the internal market clause (Art 114 TFEU), because it is difficult to imagine how the AI Act could contribute to the internal market if an AI system is only developed in the EU, but never put into operation there.

[52] Recital (53) of the AI Act lists these cases as examples of cases in which an AI system could be classified as high risk under Annex III of the AI Act, but could be exempted under Art. 6(3) AI Act.

[53] Alessandro Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*, (Springer Nature 2022) 169 https://link.springer.com/content/pdf/10.1007/978-94-6265-531-7.pdf.

[54] Ibid, 170.

[55] Ibid.

[56] For example, robots and AI systems used in care for daily communication with the elderly, and applications which provide instructions for workouts, give tips on nutrition, or store the user's weight or pulse.

[57] Martin Ebers, "AI Robotics in Healthcare between the EU Medical Device Regulation and the Artificial Intelligence Act" (2024) 11(1) Oslo Law Review 1–12.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

5. Regulation of GPAI models as a contradiction to the risk-based approach?

Moreover, the specific obligations of systemic risk GPAI model providers to conduct model evaluations, assess and mitigate potential systemic risks, monitor and report serious incidents, take corrective action, and ensure an appropriate level of cybersecurity measures (Article 55 AI Act), are not consistent with a genuine risk-based approach.

First, by definition, a GPAI model is not limited to specific use cases or applications. Instead, such a model "displays significant generality and is capable of competently performing a wide range of distinct tasks" so that it "can be integrated into a variety of downstream systems or applications" (Article 3 No. 63 AI Act). Given that GPAI models can be used widely across a variety of industries, it is difficult (if not even impossible) to formulate precise standards for classifying risks as systemic. Therefore, the AI Act does not specify which risks are systemic; instead, Article 3 No. 65 AI Act refers (generally) to negative effects on public health, safety, public security and fundamental rights. Thus, providers have no guidance on what constitutes a systemic risk and how to mitigate it. Arguably, providers can demonstrate compliance through codes of practice which will be facilitated by the AI Office until a harmonised standard is published (Article 55(2)(1) AI Act). However, this does not change the fact that "systemic" risks cannot be quantified and specified in the same way as other risks regulated by the AI Act, because they do not concern the probability of the occurrence of a certain harm (cf. Article 3 No. 2 AI Act), but rather the impact on the "union market" or the "society as a whole" (cf. Article 3 No. 65 AI Act).

Such "risk" regulation has nothing in common with the type of risk-based regulation described above in Section II.

In addition, the AI Act assumes that a particularly high amount of computation used to train GPAI models also increases their risk (Recital 111 AI Act). Therefore, Article 51(2) AI Act states that a GPAI model is presumed to have high-impact capabilities – such that qualifies it as a model with "systemic risk" under Article 51(1)(a) AI Act – if the model's training involves more than 10^25 floating-point operations (FLOPs). However, such a threshold is questionable for at least three reasons:

- First, the (systemic) risk of GPAI models depends, not only on the quantity of computational resources used, but also on a number of other factors – such as the context of the application, the model architecture, and the quality of the training.[58]
- Second, research shows that the 10^25 FLOPs threshold is questionable, since LLMs with fewer FLOPs can be just as risky and even outperform larger models with more parameters.[59]
- And third, the FLOPs threshold was set primarily for political reasons: in order to strengthen the European economy with its two start-ups Mistral (from France) and Aleph Alpha (from Germany), France and Germany in particular successfully lobbied during the negotiations to keep both companies below the threshold.[60]

This displays, once again, how arbitrarily the AI Act defines "systemic" risks.

[58] Claudio Novelli and others, "Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity" (2024) https://ssrn.com/abstract=4821952, 4.

[59] Cornelia Kutterer, "Regulating Foundation Models in the AI Act: From "High" to "Systemic" Risk" (AI-Regulation Papers 24-01-1, 12 January 2024) 6 https://ai-regulation.com/wp-content/uploads/2024/01/C-Kutterer-Regulating-Foundation-Models-in-the-AI.pdf.

[60] Sandra Wachter, "Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond" (2024) 26(3) Yale Journal of Law & Technology 671–718, 695, 698.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

6. Overly broad AI definition

Another major setback from the perspective of risk-based regulation is the overly broad definition of AI.

Initially, the European Commission justified its Proposal with the specific characteristics of (unpredictable) software systems based on machine learning.[61] Accordingly, many of the AI Act's mandatory requirements and obligations for high-risk AI systems attempt to mitigate mainly the risks of AI systems based on machine learning, while such far-reaching obligations are not strictly necessary for other software systems.

From a truly risk-based approach as well as from the principle of technology neutrality, the AI Act should have, therefore, imposed different regulatory burdens on different designs, because predictable AI systems do not pose the same risks as unpredictable systems based on machine learning.[62]

However, this is not the case. According to Article 3 No. 1 AI Act, an AI system is "a machine-based system designed to operate with *varying levels of autonomy*, 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."[63] This definition – an adaptation of the OECD's updated definition of AI[64] – extends the meaning of AI systems to cover (almost) all software systems,[65] as (i) there is no threshold for the level of autonomy required for a system to be classified as such, and (ii) the use of the word "may" implies that systems do not always have to exhibit adaptiveness after deployment, to be considered AI. Hence, the AI Act applies not only to machine learning, but also to logic- and knowledge-based approaches (recital 12 AI Act).

As a result, even deterministic software systems used in high-risk sectors are subject to the highest requirements. Consequently, as *Schrepel* puts it: "By not discriminating between AI systems based on their functioning, the AI Act indirectly sanctions those that are safer and easier to control"[66] – contrary to the components of a true risk-based approach.

7. Double regulatory burdens due to the horizontal approach

The AI Act does not replace existing EU law, but applies concurrently to it. As a result, companies and individuals will have to observe not only the AI Act, but also other related legislations – such as EU data protection law (Article 2(7) AI Act), EU copyright law, EU consumer and product safety law (Article 2(9) AI Act). As many principles and provisions of the AI Act overlap with those of pre-existing legislations, such a *horizontal approach* inevitably leads – in many areas – to legal uncertainty, different interpretations, contradictions and, ultimately, to double regulatory burdens – contrary to the idea of risk-based regulation.

Consider the following three examples from data protection law, medical law and product safety for machinery:

[61] European Commission, *AI Act Proposal*, COM(2021) 206 final, explanatory memorandum, 2; European Commission, *Impact Assessment*, SWD(2021) 84 final, Part 1/2, 28 et seq.

[62] Thibault Schrepel, *Decoding the AI Act: A Critical Guide for Competition Experts* (ALTI Working Paper, Amsterdam Law & Technology Institute – Working Paper 3-2023, October 2023) 11 https://ssrn.com/abstract=4609947.

[63] Emphasis added.

[64] OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, amended on 03/05/2024 by the 2024 OECD Ministerial Council Meeting (MCM) https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.

[65] According to Recital (12) AI Act, the only software systems that should not be regarded as AI are "systems that are based on the rules defined solely by natural persons to automatically execute operations".

[66] Schrepel, supra n 62, 11.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

(i) As the AI Act and EU data protection law apply in parallel, both the EDPB and the EDPS have already pointed out during the negotiations, that it is important to clearly avoid any inconsistencies and possible conflicts between the AI Act and data protection law.[67] However, these concerns have largely not been taken into account. For example, both the GDPR and the AI Act impose transparency obligations, but the scope and the requirements are regulated differently in the two laws.[68] Another example is the right to explanation and human intervention/oversight. While the GDPR requires human intervention (Article 22(3) GDPR) and a right to meaningful information for decisions based solely on automated processing, including profiling (Article 15(1)(h) GDPR), the AI Act requires human oversight (Article 14 AI Act) and a right to explanation of individual decision-making (Article 86 AI Act) for high-risk AI systems – whereby both the content as well as the prerequisites and legal consequences are regulated completely differently.

(ii) The relationship between the AI Act and the Medical Devices Regulation (MDR) is also currently unclear.[69] Given that both legislations apply simultaneously, without a formal hierarchy clause in either the AI Act or in the MDR to decide which of the overlapping rules should apply, a number of inconsistencies and contradictions arise. For example, its definitions – such as "importer," "putting into service," "provider" and "deployer" – differ from those of the MDR.[70] These differences do not only complicate compliance with both regulations, but will also make it very difficult for providers to integrate the documentation required under the AI Act into the MDR documentation "to ensure consistency, avoid duplication and minimise additional burdens" (cf. Art. 8(2)(2) AI Act).

(iii) A third example is the new Machinery Regulation (MR),[71] which applies alongside the AI Act when an AI system is used in a machine. This also results in a duplication of requirements. For example, both the AI Act and the MR require human oversight, but the two sets of rules differ in detail.[72] Another overlap and contradiction concerns the recoding and retention of decision-making data. While the MR requires "enabled" recording of "data on the safety-related decision-making process" the AI Act requires that high-risk AI systems automatically record logs of "events" throughout the system's lifespan. This discrepancy is likely to create practical challenges for companies to ensure compliance with both regulatory frameworks, resulting in double regulatory burdens.

[67] EDPB-EDPS, Joint Opinion 5/2021, 18 June 2021, para 57.

[68] The GDPR establishes the principle of transparency to facilitate the exercise of data subjects' rights under Art 15–22, including the right to erasure, to rectification and to data portability. In contrast, the AI Act contains transparency obligations only for high-risk AI systems (Art 13 AI Act) and for other certain AI systems (Art 50 AI Act). Moreover, Art 13 AI Act focuses on the interests of the deployer of an AI system rather than on the final user and/or data subject.

[69] Cf. Ebers, supra n 57.

[70] For a detailed discussion cf Wimmy Choi, Marlies van Eck and Cécile van der Heijden, "Theo Hooghiemstra and Erik Vollebregt, Legal analysis: European legislative proposal draft AI act and MDR/IVDR" (January 2022) 16ff, <www.government.nl/binaries/government/documenten/publications/2022/05/25/legal-analysis-european-legislative-proposal-draft-ai-act-and-mdr-ivdr/Report+analysis+AI+act+-+MDR+and+IVDR.pdf>.

[71] Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC (OJ L 165, 2962023, 1–102.

[72] Tobias Mahler, "Smart Robotics in the EU Legal Framework: The Role of the Machinery Regulation" (2024) 11(1) Oslo Law Review 1–18.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

All three examples show that the AI Act will create inefficiencies, regulatory uncertainty and increased compliance costs, due to conflicting or duplicative requirements in both the AI Act and other EU laws.

8. Overlap of enforcement structures

Since the AI Act applies in addition to other existing EU laws, there is also a risk that the same use of an AI system may be subject to different regulatory authorities within a Member State. Article 70 AI Act leaves the designation of competent authorities to the Member States.[73] Member States may choose to entrust the enforcement of the AI Act either to existing bodies (such as national data protection authorities) or to entirely new administrative bodies (such as the Agency for the Supervision of AI in Spain). This will most likely lead to overlapping enforcement structures, duplication of procedures, inconsistencies between these procedures and, in the worst case, double fines for the same set of facts. As a result, coordination between the different regulatory bodies is necessary to avoid double and/or over-enforcement, which is contrary to the constitutional principles of *ne bis in idem* (Article 50 EU Charter of Fundamental Rights) and the principle of proportionality.[74]

In the absence of specific provisions on cooperation mechanisms in the AI Act and other EU legislation, the overlap of enforcement structures will be a significant challenge, increasing legal uncertainty and compliance costs – again, contrary to the risk-based approach.

IV. How to implement the AI act in accordance with a truly risk-based approach

The foregoing analysis shows that key provisions of the AI Act do not reflect the characteristics of a true risk-based regulation; leading to legal uncertainty and potential over-regulation, as well as unjustified increases in compliance costs. However, this is nothing that cannot be fixed. The AI Act includes sufficient tools to support future-proof legislation and to implement it in line with a genuine risk-based approach. Accordingly, the following analysis focuses on how to implement the AI Act in accordance with a truly risk-based approach.

1. The risk-based approach as a guiding principle for implementing the AI act

The implementation of the AI Act can take several forms. In this paper, the term "implementation" is used to describe the measures taken to ensure compliance with the obligations imposed by the AI Act. Implementation includes:

- the adoption of more specific legal provisions (normative implementation),
- the interpretation, application, and enforcement of the AI Act by public authorities (administrative implementation) and
- its interpretation by the courts (judicial implementation).

Implementation can take place both at the level of the EU (e.g. when the European Commission issues guidelines or adopts implementing/delegated acts) and at Member

[73] Cf. also recital 157 AI Act.
[74] Cf. *Case 14/68 Wilhelm [1969] ECR 1, para. 11.*

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

States' level (e.g. when Member States adopt more specific rules or when Member State authorities enforce the AI Act).

When implementing the AI Act, both the European Commission and the Member States are required to respect the choice of the European legislator in following a risk-based approach. This follows both from the preparatory work[75] and from the *ratio legis* as laid down in recital (26) AI Act as explained above in Section I.1.

Apart from the *ratio legis* of the AI Act itself, an interpretation of the AI Act in the light of EU primary law also supports the view that the implementation of the AI Act should follow a truly risk-based approach. As explained above, the principle of (legislative) proportionality is enshrined in Article 5 TEU. The CJEU has consistently held that:

> "the principle of proportionality is one of the general principles of Community law. By virtue of that principle, the lawfulness of the prohibition of an economic activity is subject to the condition that the prohibitory measures are appropriate and necessary in order to achieve the objectives legitimately pursued by the legislation in question; when there is a choice between several appropriate measures recourse must be had to the least onerous, and the disadvantages caused must not be disproportionate to the aims pursued."[76]

As a result, both secondary law (*ratio legis* of the AI Act) and primary law (principle of legislative proportionality) favour the implementation of the AI Act on the basis of a truly risk-based approach.

2. *Tools in the AI act to support future-proof legislation*

Although some key provisions of the AI Act are not consistent with a truly risk-based approach, the Regulation provides for sufficient tools to interpret, specify, and even amend it accordingly.

First, many provisions contain broad language that is subject to interpretation. To this end, the Regulation relies, not only on the courts (and ultimately on the CJEU), but also on the European Commission and its AI Office to develop guidelines (Article 96 AI Act).

Moreover, Article 97 AI Act gives the European Commission the power to adopt delegated acts, in particular to modify Annex III use-cases (Article 7(1) and 7(3) AI Act) and the conditions under which these systems shall not be considered to be high-risk (Article 6(6)–(7) AI Act), as well as to amend or supplement the thresholds, benchmarks and indicators for classifying GPAI models as "systemic risk" (Article 51(3) AI Act), including the criteria set out in Annex XIII for the designation of GPAI models with systemic risk (Article 52(4) AI Act).

Other tools provided by the AI Act are harmonised standards developed by CEN/CENELEC as well as codes of practice for GPAI. As soon as the European Commission approves harmonised standards and/or codes of practice, providers can rely on them to demonstrate compliance with their obligations (Article 40(1), 53(4)(1) and 55(2)(1) AI Act).

All of the instruments – guidelines, delegated and implementing acts of the European Commission, as well as harmonised standards and codes of practice – are powerful tools to clarify, concretise, amend, supplement, modify or even delete a large number of provisions of the AI Act to "take into account the rapid pace of technological development, as well as the potential changes in the use of AI systems" (recital 52 AI Act). Indeed, such measures are

[75] Cf. European Commission, supra n 61, 78.

[76] *Case C-331/48 The Queen v Minister of Agriculture, Fisheries and Food and Secretary of State for Health, ex parte Fedesa et al.*

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

necessary to implement the AI Act in accordance with a truly risk-based approach – in line with its *ratio legis* and the principle of legislative proportionality enshrined in Article 5 TEU.

For this, the following aspects should be taken into account.

3. Risk-benefit analysis and evidence-based high-risk categories

As explained above, the AI Act lacks an explicit, independent risk-*benefit* analysis and evidence-based (high-)risk categories. On the other hand, the European Commission possesses – not only the authority to issue guidelines for the classification of AI systems as high-risk under Annex III, but also – the power to amend, modify or remove use-cases for high-risk AI systems in Annex III (Article 7(1) and 7(3) AI Act) and to modify or add new conditions under which Annex III high-risk AI systems shall not be considered to be high-risk according to Article 6(3) AI Act (Article 6(6)–(7) AI Act).

Here, the Commission should take into account both the potential harm that AI systems may cause, and their economic and social benefits. To a certain extent, the Commission already has to assess the beneficial effects if it wants to amend Annex III by adding or modifying use-cases of high-risk systems, as Article 7(2)(j) AI Act requires the Commission in these cases to also take into account "the magnitude and likelihood of benefit of the deployment of the AI system for individuals, groups, or society at large."[77] However, the benefits should not be one of many factors to be considered under the umbrella of a risk analysis, but an *independent* criterion to be weighed against the risks.[78]

Moreover, the classification of AI systems as "high-risk" should be based on sufficient empirical evidence. This requirement is also present in the AI Act. Specifically, Article 6(5) AI Act states that the Commission, when adopting delegated acts to add new conditions under which an AI system referred to in Annex III is not to be considered high-risk, shall follow "concrete and reliable evidence." However, such evidence-based risk analysis should be carried out not only in this context, but generally when defining (or changing) the risk categories, together with a clear methodology, explanation and documentation.

4. Regulation of GPAI models

With respect to GPAI models, the European Commission also has flexible tools at its disposal to address at least some of the concerns raised above.

This applies in particular to the thresholds for classifying GPAI models as "systemic risk," which can be amended by delegated acts (Article 51(3) and Article 52(4) AI Act). Thus, the Commission can revise the criteria, especially the FLOP threshold which is not based on empirical evidence but rather the result of a political compromise. Article 51(3) AIA even goes beyond the mere update of the FLOPs threshold, by empowering the Commission to also "supplement benchmarks and indicators in light of evolving technological developments, such as algorithmic improvements or increased hardware efficiency, when necessary, for these thresholds to reflect the state of the art." Accordingly, the Commission has the opportunity to use real-world evidence to set and define the systemic risk threshold by going beyond FLOPs and adding or replacing them with new benchmarks.

In addition, there is a need to develop a common methodology to define systemic risks pursuant to Article 55(1) AI Act and to establish measures as well as benchmarks for assessing and managing these risks pursuant to Article 55(1) AI Act. It is therefore essential

[77] This criterion applies also, as per Art. 7(3)(a) AI Act, when the Commission wants to remove high-risk AI systems from the list in Annex III.

[78] In line with Schrepel's claim to follow a "law + technology" approach; cf. Thibault Schrepel, 'Law + Technology' (Stanford CodeX Working Paper, 19 May 2022) https://ssrn.com/abstract=4115666.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

that GPAI model providers, standard setting organisations, national authorities, civil society organisations and other relevant stakeholders work together with the AI Office to develop codes of practice in this area, considering international approaches.[79]

This is likely to be a major challenge as the Codes of Practice shall be ready at the latest by 1 May 2025 (Art. 56(9)(1) AI Act) – a very short timeframe for drafting a Code based on very vague rules. Against this background, the focus should be primarily on the risk-based approach – addressing only relevant risks and sticking to the letters of the AI Act.

5. Overly broad AI definition and mandatory requirements for high-risk AI systems

Proper risk-based implementation can also mitigate the overly broad definition of AI and the problem that deterministic software systems used in high-risk sectors are subject to the same mandatory (strict) requirements as unpredictable AI systems based on machine learning. While the definition of AI systems in Article 3 AI Act cannot be changed by the Commission but only be interpreted, many of the requirements in Article 8–15 AI Act are worded broadly enough to be applied in a manner that takes into account the fact that AI systems pose different risks due to their different levels of autonomy and adaptability:

- For example, the requirements laid down in Article 10 AI Act on data and data governance appear to apply only to AI systems "which make use of techniques involving the training of AI models with data," thus excluding other AI systems that are not based on machine learning.
- Moreover, the wording of Article 13(1) AI Act that high-risk AI systems shall be "sufficiently" transparent to enable deployers to interpret the output of a system and use it "appropriately" is open enough to distinguish between different AI systems with different levels of transparency.
- With respect to human oversight, Article 14(3) AI Act explicitly emphasises that the necessary oversight measures must be "commensurate with the risks, level of autonomy and context of use of the high-risk AI system."
- Furthermore, Article 15(1) AI Act speaks only of an "appropriate" level of accuracy, robustness and cybersecurity, whereas Article 15(4) AI Act provides specific rules for high-risk AI systems "that continue to learn."

Against this background, it seems sufficient that the European Commission issues guidelines on the application of Articles 8 to 15 (Article 96(1)(a) AI Act) to clarify how these articles apply with regard to different AI technologies.

6. Double regulatory burdens and overlap of enforcement structures

Guidelines of the European Commission can also help companies to deal with overlaps, inconsistencies and (potential) contradictions between the AI Act and other EU legislation to avoid double regulatory burdens and over-enforcement. Article 96(1)(e) AI Act requires the Commission to develop guidelines with "detailed information" on the relationship of the Act with other relevant Union law, including with regard to "consistency in their enforcement."

To this end, the European Commission should conduct an in-depth analysis to identify overlaps and contradictions between the AI Act and other horizontal or sectoral legislation. Based on this research, guidelines could then be rolled out to

[79] For example, the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems https://www.mofa.go.jp/files/100573473.pdf.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

help clarify the relationship between these laws (i.e. lex specialis, lex generalis, and complementary laws).[80]

Moreover, the European Commission should carry out an assessment of the relationship between the AI Act governance bodies and other governance bodies on EU and Member State level. Guidelines could then clarify whether a particular governance body is the lead authority and how the different bodies should cooperate or relate with each other.[81]

Whereas guidelines could make an important contribution to clarifying the interplay between the AI Act and other EU laws (including the relationship between different enforcement bodies), they can only interpret existing laws but not change them. Yet, many of the discussed issues cannot be solved by interpretation alone, as they are rooted in legal inconsistencies, mostly at EU level. Therefore, only a revision of sector specific EU laws can deal with these issues.

V. Regulating AI outside the EU: lessons from the AI act

Lawmakers around the world are studying the AI Act to determine whether they should follow the European Union's lead and adopt similar laws to regulate AI systems. This is the so-called "Brussels effect" – a term coined by *Anu Bradford*, to describe Europe's global footprint in terms of triggering emulation in other legal systems. In its original formulation, the Brussels effect was seen mainly as a *de facto* phenomenon where companies voluntarily follow EU rules in standardising a product or service, making their business processes simpler.[82] However, it can also take a *de iure* effect where countries outside the EU adopt EU-like regulations.[83]

Over the past years, scholars have wondered whether the AI Act will unleash a new Brussels effect. While some claim this could be the case,[84] others disagree.[85] EU policymakers strongly believe in the Brussels effect. From the onset, the AI Act was designed with its extraterritorial effects in mind.[86]

In the following sections, it is argued that the AI Act is unlikely to unfold a *de iure* Brussels effect (1.). Moreover, considering the lessons the AI Act can teach lawmakers around the world, such an effect would also be undesirable (2.).

1. Why the AI act won't trigger a Brussels effect

One major obstacle for foreign legislators against simply copying and pasting the AI Act is the complexity of AI as a policy area. Unlike the GDPR – which has indeed served as a model for data protection regulation around the world – Artificial Intelligence does not present a single policy problem (e.g. how to protect the fundamental right to privacy), but rather a set

[80] Similarly, Axel Voss, "Ten steps to make the AI Act an EU success story" (06 March 2024) <https://www.axel-voss-europa.de/2024/03/06/ten-steps-to-make-the-ai-act-an-eu-success-story/>.

[81] See again Voss, ibid.

[82] Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press 2020) xiv.

[83] Ibid, 85.

[84] Fabian Lütz, "How the 'Brussels effect' could shape the future regulation of algorithmic discrimination" (2021) 1 Duodecim *Astra* 142–63; Charlotte Siegmann and Markus Anderljung, "The Brussels effect and artificial intelligence: How EU regulation will impact the global AI market" (2022) arXiv preprint arXiv:2208.12645 https://arxiv.org/abs/2208.12645.

[85] Alex Engler, "The EU AI Act Will Have Global Impact, but a Limited Brussels Effect" (8 June 2022) https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/; Almada and Radu, supra n 29; Ugo Pagallo, "Why the AI Act Won't Trigger a Brussels Effect" (16 December 2023) in *AI Approaches to the Complexity of Legal Systems* (Springer 2024, forthcoming) https://ssrn.com/abstract=4696148. 1

[86] Gabriele Mazzini and Salvatore Scalzo, "The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts" in Carmelita Camardi (ed), *La Vie Europea per l'intelligenza artificiale* (Cedam 2022) 1.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

of loosely connected problems – ranging from the protection of health and safety to a variety of fundamental rights. Even when legislators follow the definition of AI – as set out in the AI Act, which is based on the recently updated OECD principles – there is little agreement worldwide on *who* and *what* should be regulated: public administration, law enforcement bodies, the judiciary, alternative dispute resolution mechanisms, and/or the (entire) private sector? Autonomous weapons systems? Self-driving vehicles and other cyber-physical machines? Medical devices and expert systems? Social scoring? Employment? Social welfare? Biometric identification/categorisation systems? Credit scoring? Life and health insurance? Algorithmic recommender systems used by platforms? AI-based contracts?

Moreover, there is little international consensus on the *how* of such a regulation, i.e. how to apply fundamental values such as human dignity and autonomy, fairness, transparency etc., in a given context. While global agreements in recent years – such as UNESCO's AI Recommendation and the OECD's AI Principles – recognise such fundamental principles, the ambiguity of these high-level agreements accommodates different political and ethical positions, allowing states to interpret them differently.[87] As *Roberts* and others point out: "Take AI fairness, a principle supported by all G20 member states, as applied to facial recognition technology. The implementation of this principle in the EU context involves the proposed banning of these technologies, while in China, ethnic-recognition technologies are permissible in the name of order and social stability."[88]

Even the adoption of a regulatory technique such as the risk-based approach, which appears to be politically neutral, involves fundamental policy choices: Does a country want to eliminate/ban (certain) risks, reduce them to an acceptable level, or strike a balance between risk reduction and the costs of regulation? If it is the latter: How does it balance risks and benefits in a specific sector and/or use-case?

Another major obstacle for foreign legislators to simply follow the EU's approach, is that the AI Act does not establish a comprehensive legal framework that can be adopted *tel quel*. Instead, it interacts in a very complex way with a rather sophisticated system of existing EU laws. In particular, the AI Act both complements existing product safety legislation[89] and builds, at the same time, on this legislation for the purpose of risk classification.[90] Moreover, the AI Act complements existing EU non-discrimination law with specific requirements aimed at minimising the risk of algorithmic discrimination. The AI Act further complements EU data protection law. This means that any processing of personal data by an AI system must comply with EU data protection law (e.g. the GDPR) and the AI Act. Last but not least, the Act provides for a number of future-proof instruments that will complement the Regulation, such as delegated and implemented acts, codes of practice and harmonised standards. As a result, it would not make sense to adopt the AI Act in isolation, as such a piece of legislation would be neither comprehensible nor meaningful without the rich body of existing EU law.

For all these reasons, it is unlikely that the AI Act will become the new global standard.

2. *Why the AI act shouldn't trigger a Brussels effect*

There are also important reasons as to why a *de iure* Brussels effect of the AI Act is not desirable.

[87] Brent Mittelstadt, 'Principles alone cannot guarantee ethical AI' (2019) 1 Nature Machine Intelligence 501-507 https://doi.org/10.1038/s42256-019-0114-4503, 503.

[88] Huw Roberts, Emmie Hine, Mariarosaria Taddeo and Luciano Floridi, "Global AI governance: barriers and pathways forward" (2024) 100(3) International Affairs 1275–1286 https://doi.org/10.1093/ia/iiae073, 8.

[89] Insofar as products using AI as a safety component must additionally comply with the specific requirements set out in the AI Act; see Art 2(9) and Art 6(1) AI Act.

[90] See Art 6(1)(b) AI Act.

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press

First, it is noteworthy that the economic, social, legal and political situation of countries are very different and, as a consequence, how countries and citizens are affected by AI.[91] Second, AI-specific regulation is still in its early stages, and it is unclear what the social and economic consequences of the AI Act will be. If it turns out that the AI Act has unforeseen significant negative effects, these will be duplicated around the world, with a Brussels effect.[92] Third, the existence of different AI regulations in different countries can – under the right conditions – stimulate experimentation and innovation in regulation through trial and error.[93]

Fourth, from a fundamental rights perspective, scholars rightly point out that the AI Act is the product of constitutional constraints.[94] Since the EU has no general competence to harmonise fundamental rights in the Member States, it relies instead on the competence to promote the internal market (Article 114 TFEU). As a result, the AI Act does not follow a fundamental rights approach, but instead uses product safety law and risk regulation. However, such an approach is ill-suited to protect human rights. Given that most countries (or even regions) are not subject to the same competence constraints as the EU, they can (and should) use other approaches to address fundamental rights issues in relation to AI.

Finally, legislators around the world should take into account that key provisions of the AI Act do not follow a truly risk-based approach, particularly with respect to proper risk-*benefit* analysis, limited reliance on empirical evidence, pre-defined, closed risk categories, systemic risks of GPAI models, the overly broad definition of AI, double regulatory burdens, and overlapping enforcement structures. It is also *and above all* for these reasons that regulators outside the EU should not blindly follow the EU approach.

[91] Nathalie A Smuha, "From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence" (2021) 13(1) Law, Innovation and Technology 57–84, at p 81 https://doi.org/10.1080/17579961.2021.1898300.

[92] Smuha, supra n 91, 80.

[93] Smuha, supra n 91, 69.

[94] Almada and Radu, supra n 29.

Cite this article: M Ebers (2025). "Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act". *European Journal of Risk Regulation* **16**, 684–703. https://doi.org/10.1017/err.2024.78

https://doi.org/10.1017/err.2024.78 Published online by Cambridge University Press