https://doi.org/10.1038/s41591-024-02874-2

New AI regulation in the EU seeks to reduce risk without assessing public benefit

Barbara Prainsack & Nikolaus Forgó

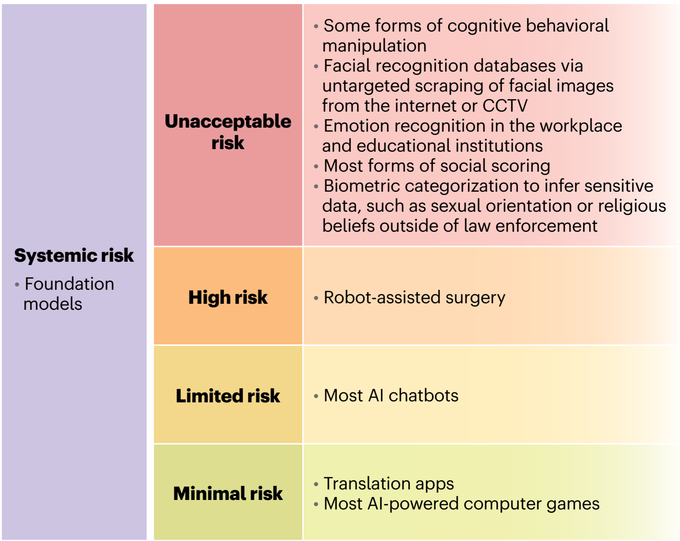
Check for updates

The European Union's new AI Act focuses on risk without considering benefits, which could hinder the development of new technology while failing to protect the public.

Not many regulatory initiatives have received as much public attention as the draft of the European Union (EU) Artificial Intelligence (AI) Act, a binding regulation that will be directly applicable in all EU member countries. The act will enter into force 20 days after its official publication, and its provisions will become applicable (in different stages) starting 6 months after that date[1].

In April 2021, the European Commission published their first proposal, and in December 2023, the European Commissioner for Internal Market announced that a deal had been reached. In the two and a half years in between, hundreds of pages of text were produced, open letters were written, and lobbying activities were carried out. Some questions remained controversial until the end, such as the permissibility of real-time biometric surveillance in the context of serious crimes, and the regulation of foundation models (general-purpose AI systems trained on broad datasets that can be used in a wide range of applications). The final version of the act represents a compromise on both points: real-time biometric identification, such as facial recognition for law enforcement purposes, will remain legal in specific cases such as the prevention of terrorism. And for foundation models, a special risk category was added for "high-impact general-purpose AI models that might pose systemic risk" such as GPT-4 (ref. 2). A model is presumed to have high impact when the cumulative amount of compute used for its training, measured in floating point operations, is greater than 10^{25}, a threshold that is reached by fewer than 20 companies in the world[3].

The key legislative approach taken in the AI Act, however, remained largely uncontested in the process. As with the EU's General Data Protection Regulation, the AI Act will apply across domains and will not be limited to specific fields, such as health. In addition, the AI Act tailors regulation to the presumed level of risk, ranging from unacceptable to minimal[4] (Fig. 1). AI applications that are seen to pose unacceptable risk (such as some forms of cognitive behavioral manipulation) may not be used at all. Applications in the minimal risk category (such as AI-enabled video games) require very little oversight. Most regulatory concern is focused on the categories in between and, in particular, on high-risk applications.



Systemic risk • Foundation models	**Unacceptable risk**	• Some forms of cognitive behavioral manipulation • Facial recognition databases via untargeted scraping of facial images from the internet or CCTV • Emotion recognition in the workplace and educational institutions • Most forms of social scoring • Biometric categorization to infer sensitive data, such as sexual orientation or religious beliefs outside of law enforcement
	High risk	• Robot-assisted surgery
	Limited risk	• Most AI chatbots
	Minimal risk	• Translation apps • Most AI-powered computer games

Fig. 1 | The AI Act tailors regulation to different levels of presumed risk. Examples of each risk category are shown, together with the special category of 'systemic risk', which is reserved for foundation models. CCTV, closed-circuit television.

How to assess risk

There is much to be liked about the risk-based approach that the AI Act is taking. It has some clear advantages over one-size-fits-all alternatives that over-regulate on the low-risk end of the spectrum while missing important problems on the other end. However, the risk-based approach also raises serious practical and political issues. It remains unclear what is considered a high-risk AI system. Two annexes to Article 6 of the European Commission's proposal provide only a partial answer[5]. Annex 2 considers all technologies that fall under the EU's product safety legislation as high risk, including medical devices, whereas Annex 3 identifies entire fields as high risk, including biometrics, education, employment and law enforcement (but not health). Some fields of application are explicitly exempt from the regulation, such as AI used solely for military purposes.

Risk assessment based on the intended field of use may fall short of its stated purpose. The exclusion of military AI is likely to raise complicated issues with technologies developed for military purposes that could be applied in other domains (dual-use technologies). For example, autonomous drones developed for military surveillance could be used for civil purposes, such as crop monitoring in agriculture. General-purpose AI tools, such as large language models, can be used for multiple fields; it will be very difficult to delineate the field of application for these upfront[6]. Moreover, given the rapid pace of technology development, risk assessments may be outdated by the time their legal consequences emerge. Finally, some continuously evolving AI applications may pose risks that no one was able to foresee.

There is a lack of precision in the AI Act, which could undermine its own intentions. For example, the provision that any AI system used as a medical device is considered high risk, regardless of how it will be used, means that any such device must comply with the act's high-risk provisions in addition to the requirements for medical device certification. In contrast, a device classified as a lifestyle gadget, such as a smartwatch, avoids these additional regulatory burdens, even if it poses the exact same level of risk. This situation creates competitive advantages for companies with sufficient economic power to legally challenge high-risk assessments. Similar to what has been observed with the EU's General Data Protection Regulation, the AI Act would result in smaller enterprises facing the full burdens of compliance, while large corporations can avoid effective oversight. Risk classification will be based on preliminary self-assessment, so the act is likely to increase the problem of developers deliberately misclassifying their innovations to avoid having to meet stringent requirements.

Public benefit

The legislation's strong focus on risk also neglects the potential benefits of a technology. Regulators should aim to increase the overall value that the use of a technology creates for the public, so both risks and benefits should matter. 'Data solidarity' is an approach to assessing both the risks and the benefits of a technology systematically[7]. This posits that the public value of data use is high when it can be assumed that the data use will result in clear benefits for the public while posing only minimal risks of substantial and undue harm to any person or group. Since October 2023, an online tool has been publicly available to support structured public value assessment of specific instances of data use[8].

Regulation that examines only risk could hinder the development of applications for which the expected public value outweighs the risks. However, it is important to note that regulation itself does not stifle innovation. Claims that it does are often linked to the concern that in a race to develop AI, Europe may fall behind other countries – in particular, China and the United States. This suggestion, however, is both sweeping and misleading. First, regulation that clearly delineates what technology developers and providers can and cannot do gives businesses legal certainty and predictability, supporting technological advancement rather than hindering it. Technology development is obstructed by ambiguous regulation and a lack of public investment, not by regulation as such.

Second, not all innovation is necessarily beneficial for societies[9]. In recent decades, some innovation has increased capital gains without helping to solve societal challenges[10]. Analysts are increasingly drawing attention to the 'dark side' of innovation[11], arguing that a substantial proportion of innovation has exacerbated societal issues, such as climate change or rising inequalities. There is no doubt that technological innovation is of key importance to the field of healthcare, but not all innovation is equally beneficial. A more systematic consideration of the benefits that will materialize for different groups, and at whose cost, is sorely needed.

Third, the narrative that Europe is lagging behind China and the United States because of the EU's stringent regulation of digital technologies is overly simplistic. There are much wider reasons that explain Europe's inability to lead on the AI front, including the lack of an integrated digital market and insufficient public investment into technological capabilities and human resources in this field.

Harm mitigation

Effective regulation protects people from harm and supports technology development that will yield public benefits. In the context of AI, this means that the EU needs to go beyond its well-worn frame of fair market competition. The European Group on Ethics in Science and New Technologies, a permanent advisory body to the EU's commission president, argued in an opinion last year that the EU's fixation on fair market competition distracts from some of the structural issues that form the root causes of harms in the digital era[12]. Disinformation and misinformation, for example, can cost lives, as was illustrated during the COVID-19 pandemic[13]. With the use of generative AI, which allows people to create fake images, videos or audio files that are increasingly difficult to distinguish from genuine material, the problem will accelerate.

These issues cannot be solved merely with better content management of digital platforms, or with mandatory fundamental rights impact assessments. These problems will persist as long as a large part of the digital public space is owned by quasi-monopolist technology giants who, legitimately, aim to maximize their profits. Algorithms will continue to prioritize materials that are shocking or controversial enough to create traffic. Foundation models will keep being created with little or no public oversight or accountability. What is needed instead is for the EU regulator to increase democratic control over digital technologies in an effective manner. AI has a major impact on people's lives, from its most mundane forms in supporting harmless entertainment to its use in clinical practice, so the creation and operation of these technologies need to be moved back into the realm of effective control by the people.

Harm-mitigation measures should be in place to ensure that people who experience harm from AI-based and other digital technologies receive support. Today, people experiencing such harm, such as being declined insurance, or having medical information passed on to third parties illegally, often lack access to legal remedies, especially when they cannot prove who caused the harm. In addition to the reduction of risk, the introduction of harm-mitigation mechanisms is overdue[14].

Publicly owned infrastructures and technologies, such as publicly owned foundation models, would increase democratic control over AI[15]. The EU should also invest in education, research and knowledge transfer to increase European technical competitiveness. Without such investments, the development and ownership of critical infrastructure will remain with the private sector. This will not only lock in the public sector's dependence on tech giants but also limit the possibility for effective regulation.

Barbara Prainsack [1,2,3] ✉ **& Nikolaus Forgó** [1,4]

[1]Research Platform Governance of Digital Practices, University of Vienna, Vienna, Austria. [2]Department of Political Science, University of Vienna, Vienna, Austria. [3]Institute of Advanced Study, Berlin, Germany. [4]Department of Innovation and Digitalisation in Law, University of Vienna, Vienna, Austria.

✉e-mail: barbara.prainsack@univie.ac.at

Published online: 18 March 2024

References

1. European Union. *EUR-Lex* https://go.nature.com/4361deo (accessed 3 March 2024).
2. European Parliament. https://go.nature.com/3RImQOd (accessed 3 March 2024).
3. Moës, N. & Ryan, F. *The Future Society* https://go.nature.com/3P1vBke (2023).
4. Laux, J., Wachter, S., & Mittelstadt, B. in *Regulation & Governance* Vol. 18 (eds. Levi-Faur, D. et al.) 3–32 (2024).
5. European Commission. *EUR-Lex* https://go.nature.com/3P1SHHj (2021).
6. Novelli, C., Casolari, F., Rotolo, A., Taddeo, M. & Floridi, L. *AI Soc.* https://doi.org/10.1007/s00146-023-01723-z (2023).
7. Prainsack, B., El-Sayed, S., Forgó, N., Szoszkiewicz, Ł. & Baumer, P. *Lancet Digit. Health* **4**, e773–e774 (2022).

8. El-Sayed, S. et al. *PLUTO - Public Value Assessment Tool* https://pluto.univie.ac.at (2023).
9. Pfotenhauer, S. M., Juhl, J. & Aarden, E. *Res. Policy* **48**, 895–904 (2019).
10. Birch, K., Chiappetta, M. & Artyushina, A. *Policy Stud.* **5**, 468–487 (2020).
11. Coad, A., Nightingale, P., Stilgoe, J. & Vezzani, A (eds.). *The Dark Side of Innovation* (Routledge, 2022).
12. European Commission, Directorate-General for Research and Innovation, European Group on Ethics in Science and New Technologies. *Publications Office of the European Union* https://doi.org/10.2777/773647 (2023).
13. Expert Panel on the Socioeconomic Impacts of Science and Health Misinformation. *CCA CIC* https://go.nature.com/3TfblhA (23 January 2023).
14. McMahon, A., Buyx, A. & Prainsack, B. *Med. Law Rev.* **28**, 155–182 (2019).
15. Jones, E. *Ada Lovelace Institute* https://go.nature.com/3Pq1Mdx (12 October 2023).

Competing interests

B.P. is a member of the Austrian National Bioethics Commission, and Chair of the European Group on Ethics in Science and New Technologies (EGE). N.F. is a member of the Austrian Data Protection Board and the AI Board, both independent consultancy bodies to the Austrian Government, and also serves as principal investigator in several European Union-funded research projects on AI (full list: https://id.univie.ac.at/en/third-party-funded-projects/). This Comment was written in a personal academic capacity by both authors.