cerre

THE AI ACT AND TECHNOLOGICAL NEUTRALITY

ISSUE PAPER

November 2025

Zach Meyers
Daniel Schnurr
Pierre Larouche

As provided for in CERRE's bylaws and procedural rules from its "Transparency & Independence Policy", all CERRE research projects and reports are completed in accordance with the strictest academic independence.

The project, within the framework of which this report has been prepared, received the support and/or input of the following CERRE member organisations: Amazon, IBPT, Microsoft, and TCS. However, they bear no responsibility for the contents of this report. The views expressed in this CERRE report are attributable only to the authors in a personal capacity and not to any institution with which they are associated. In addition, they do not necessarily correspond either to those of CERRE, or of any sponsor or of members of CERRE.

info@cerre.eu – www.cerre.eu

Executive Summary

The AI Act entered into force in August 2024, but the EU is still deep in the implementation process of setting up the institutional frameworks, procedures, and the significant number of secondary instruments which the Act envisages. At the same time, the AI Act has become a focal point in the debate about Europe's approach to regulation, its economic competitiveness, and the bloc's poor performance in commercialising innovation. This has culminated in proposals to make modifications to the AI Act, through a "Digital Omnibus on AI" regulation.

This issue paper focuses on evaluating how well the AI Act reflects one core element of better regulation: the principle of technological neutrality, looking in particular at the how well the AI Act is designed to adapt to changes in the 'AI value chain'. It argues that certain provisions of the AI Act and aspects of its emerging implementation risk locking in specific technological and business configurations. AI value chains are fluid and dynamic, and already more complex than anticipated in the AI Act's regulatory value chain. Against this backdrop, the paper sets out recommendations on how the AI Act's framework could better accommodate this complexity and address the identified issues in a technologically neutral way, to promote both regulatory effectiveness and Europe's competitiveness.

The AI Act's conception of the regulatory AI value chain

The paper analyses the regulatory value chain as construed by the AI Act. The framework was originally built around a technologically neutral, context-specific definition of AI Systems, with AI System providers as the primary accountable actors. The emergence of large general-purpose AI (GPAI) Models such as ChatGPT during the legislative process led lawmakers to add a parallel regime for these models, using training-compute thresholds to identify GPAI Models and those with systemic risk. The use of these unidimensional quantitative thresholds represents a significant shift away from technological neutrality and raises concerns about whether training compute is a sufficient, or even necessary, criterion to estimate a model's level of risk. Ultimately, the architecture of the AI Act results in a three-layer value chain of GPAI Model providers, AI System providers, and deployers or other operators. This value chain is used to distribute ex ante obligations of individual actors under the Act, particularly in the context of High-risk AI Systems.

Cooperation across the value chain

The AI Act requires suppliers of inputs to High-risk AI systems to set out in written agreements how they will share information and provide assistance. Suppliers of open-source inputs are exempt, except if the input is a GPAI model. Additional provisions address the post-market phase, where cooperation between actors is crucial to quickly identify and mitigate risks. The EU market surveillance regime applies to all relevant AI Systems, which requires providers to report suspected risks to authorities and to fully cooperate with their investigations. In addition, providers of High-risk AI Systems must establish post-market monitoring systems, maintain logs, report serious incidents to authorities within specified deadlines, and integrate their findings into ongoing risk-management processes. Deployers of High-risk AI Systems must monitor system use, follow providers' instructions, report serious incidents to both providers and authorities, and share relevant performance data.

While providers of GPAI Models with systemic risks are likewise subject to mandatory post-market monitoring and serious-incident reporting obligations, providers of GPAI Models without systemic risk have no comparable post-market duties. Moreover, the AI Act does not require non-systemic-risk GPAI Model providers and downstream system providers or deployers to share serious-incident information between them. While these omissions may reflect an intention to reduce the overall regulatory burden and allocate responsibility to the parties with better access to information, this creates potential gaps where serious incidents occur in AI systems based on non-systemic-risk GPAI Models or in non-high-risk use cases, with limited obligations on model providers to support post-market risk mitigation.

How developments in the AI ecosystem challenge the AI Act framework

The paper discusses developments in the AI ecosystem that create challenges for, or tensions with, the regulatory value chain enacted in the AI Act. It argues that fast-moving developments in the AI ecosystem are already straining the AI Act's built-in assumptions about how the "AI value chain" works. Innovation is increasingly organised through complex, interdependent ecosystems: models build on each other's outputs, open-weight and open-source models proliferate, and many actors combine multiple models and services in a single product. This reality is more fluid and networked than the largely linear model-system-deployer chain the Act presumes.

The paper highlights three specific issues. First, the Act's use of training compute (FLOP) thresholds to classify GPAI Models and GPAI Models with systemic risk departs from technological neutrality and may quickly become a poor proxy for risk, as techniques such as distillation, fine-tuning and specialised smaller models advance. Second, new distribution channels, cloud and model platforms hosting large numbers of models, create influential intermediaries that the Act barely contemplates, yet they could be pivotal for risk monitoring and information sharing. Third, the rise of agentic AI, which operates autonomously across changing tasks and contexts, does not fit neatly into either the "High-risk AI System" or "GPAI Model" categories and shifts much of the relevant risk to deployment choices.

Towards value-chain neutrality

Together, these trends point to the need to re-orient the AI Act towards "value-chain neutrality" as a core dimension of technological neutrality, and to strengthen post-market cooperation and information-sharing across all actors in the AI ecosystem.

In general, AI regulation should not favour one technical or organisational value-chain design over another, unless clearly justified by higher-order goals like accountability. Pre-market, it can still be useful to assign primary responsibility to one actor (e.g. the system provider) to avoid duplicated compliance. But post-market, where unforeseen harms are likely, and responsibilities are more diffuse, rigid role definitions become counterproductive. Instead, the law should foster "accountability across the value chain": every actor should have duties to share the information needed for others to meet their obligations and to resolve incidents quickly, rather than engaging in blame-shifting.

The paper argues that the AI Act's relatively open-textured post-market provisions already point in this direction, and calls for guidance, templates and possibly dispute-resolution mechanisms to operationalise flexible information-sharing. It also proposes a broader, cross-sector incident-sharing

infrastructure, with standardised reporting schemas and safeguards for trade secrets, security and competition law, so that lessons from incidents can systematically improve AI safety and resilience.

Policy recommendations

Next to the broader recommendation to consider value-chain neutrality as a central criterion for the AI Act's regulatory framework as well as several suggestions for the specific issues analysed in the paper, the paper derives the following five main recommendations for further implementation and development of the AI Act:

Recommendation 1: Law-makers should establish general principles for cooperation across the AI value chain to support effective risk identification and mitigation (especially for the post-market phase) rather than fully prescribing value-chain structures and roles, which are prone to being overly rigid and quickly becoming outdated. To support effective implementation, the AI Office could then issue complementary guidance on the information expected to be shared across the value chain.

Recommendation 2: As their relevance increases, the AI Office should provide guidance to more fully integrate General-purpose AI systems into the AI Act framework. As a first step, regulators should monitor how effectively contractual arrangements, market incentives and co-regulatory approaches, alongside existing obligations, mitigate risks arising from such systems. Further regulatory guidance should then build on observed best practices and identified market failures.

Recommendation 3: Industry players should develop new institutions and mechanisms for broader information sharing on incidents and risks in the post-market phase.

Recommendation 4: The AI Office should clarify the responsibilities of suppliers of inputs to High-risk AI Systems, so as to avoid chilling effects on the provision of important inputs, while allowing context-specific agreements and solutions to develop.

Recommendation 5: The AI Office should reconsider the role of computing thresholds as a proxy for classifying GPAI Models and GPAI Models with systemic risks in the light of current technical developments.

Table of Contents

About CERRE

Providing high quality studies and dissemination activities, the Centre on Regulation in Europe (CERRE) is a not-for-profit think tank. It promotes robust and consistent regulation in Europe's network, digital industry, and service sectors. CERRE's members are regulatory authorities and companies operating in these sectors, as well as universities.

CERRE's added value is based on:

- its original, multidisciplinary and cross-sector approach covering a variety of markets, e.g., energy, mobility, sustainability, tech, media, telecom, etc.;
- the widely acknowledged academic credentials and policy experience of its research team and associated staff members;
- its scientific independence and impartiality; and
- the direct relevance and timeliness of its contributions to the policy and regulatory development process impacting network industry players and the markets for their goods and services.

CERRE's activities include contributions to the development of norms, standards, and policy recommendations related to the regulation of service providers, to the specification of market rules and to improvements in the management of infrastructure in a changing political, economic, technological, and social environment. CERRE's work also aims to clarify the respective roles of market operators, governments, and regulatory authorities, as well as contribute to the enhancement of those organisations' expertise in addressing regulatory issues of relevance to their activities.

About the Authors

As the CERRE Director of Research, Zach Meyers has a wide remit, including managing our cross-sectoral programmes and projects. Previously the assistant director of the Centre on European Reform, Zach Meyers has a recognised expertise in economic regulation and network industries such as telecoms, energy, payments, financial services and airports. In addition to advising in the private sector, with more than ten years' experience as a competition and regulatory lawyer, he has consulted to several governments, regulators and multilateral institutions on competition reforms in regulated sectors. He is also a regular contributor to media. Zach holds a BA, LLB and a Master of Public & International Law from the University of Melbourne.

Daniel Schnurr is a CERRE Research Fellow and a Professor of Information Systems at the University of Regensburg, where he holds the Chair of Machine Learning and Uncertainty Quantification. Previously, he led the Data Policies research group at the University of Passau. He received his Ph.D. in Information Systems from the Karlsruhe Institute of Technology, where he also completed his B.Sc. and M.Sc. in Information Engineering and Management. Daniel Schnurr has published in leading journals in Information Systems and Economics on competition and data sharing in digital markets, regulation of data-driven market power, and competition and cooperation in telecommunications markets. His current research focuses on the role of artificial intelligence in competition, privacy and data sharing in digital markets as well as regulation of AI, cloud computing and the data economy.

Pierre Larouche holds the chair of Law and Innovation at Université de Montréal, where he also directs the PhD programme on Innovation, Science, Technology and Law. A graduate of McGill University, Bonn University and Maastricht University and a law clerk at the Supreme Court of Canada, Pierre Larouche was Professor of Competition Law at Tilburg University (Netherlands) from 2002 to 2017. There he founded and directed the Tilburg Law and Economics Center (TILEC), one of the largest research centres on economic governance. He also conceived and launched the Bachelor Global Law, an innovative law degree inspired by his meta-comparative and inter-disciplinary method. In his capacity as Associate Dean (2019-2024), he led the LL.B. reform at Université de Montréal. Pierre Larouche's research centers around economic governance, and in particular how law and regulation struggle to deal with complex phenomena such as innovation. An expert in competition law and civil liability, his works have been cited by the European Court of Justice and the UK Supreme Court, and they have influenced EU policy on electronic communications, competition and standardisation.

1. Introduction

The EU's Artificial Intelligence Act (the "AI Act") – the world's first comprehensive legislation seeking to regulate the uses of artificial intelligence – entered into force in August 2024.[1] Yet the EU is still deep in the process of setting up the institutional frameworks, procedures, and the significant number of secondary instruments which the Act envisages. The EU's AI Office has finalised its high-profile General-Purpose AI Code of Practice, a voluntary tool to help providers of AI foundation models comply with their AI Act obligations, subject to its endorsement by member-states and the Commission.[2] But the Commission is still consulting on how to implement the AI Act's rules on high-risk AI systems as well on transparency requirements for certain AI systems.[3] Many Member States are yet to nominate their national authorities responsible for implementing the AI Act and to produce national instruments in support of the implementation of the AI Act.

Despite being a relative newcomer to the EU's ever-growing digital rulebook, and its implementation being a work-in-progress, the AI Act has already become a focal point in the debate about Europe's economic competitiveness, the bloc's poor performance in commercialising innovation, and its approach to regulation.[4] This has culminated in the Digital Omnibus on AI proposal, containing a series of modifications to the AI Act, to delay implementation of parts of the Act, to extend to small mid-cap (SMC) firms the provisions simplifying compliance for SMEs, to streamline post-market surveillance for AI Systems built on GPAI models and to solve conflicts with the GDPR, among others.[5]

Despite this controversy, and the growing perception that the law risks stifling innovation, the AI Act reflects many principles of better regulation. Rather than adopting a precautionary approach to innovation, for example, the law largely articulates principles that AI developers are expected to internalise and operationalise in dialogue with public authorities.[6] This method of co-regulation allows a diversity of approaches to compliance and thus more flexibility to accommodate specific circumstances, different business practices and technological advancements. In consequence, this can lower compliance costs and the risks of unintended consequences. At the same time, a principle-based

[1] Regulation 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), http://data.europa.eu/eli/reg/2024/1689/oj.

[2] European Commission. Press Release: General-Purpose AI Code of Practice now available. https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_25_1787/IP_25_1787_EN.pdf

[3] European Commission (2025). Commission launches public consultation on high-risk AI systems. https://digital-strategy.ec.europa.eu/en/news/commission-launches-public-consultation-high-risk-ai-systems; European Commission (2025). Commission launches consultation to develop guidelines and Code of Practice on transparent AI systems. https://digital-strategy.ec.europa.eu/en/news/commission-launches-consultation-develop-guidelines-and-code-practice-transparent-ai-systems.

[4] Mario Draghi, 'The future of European competitiveness', September 2024.

[5] Proposal for a Regulation amending Regulations 2024/1689 and 2018/1139 (Digital Omnibus on AI), COM(2025)836 (19 November 2025).

[6] Larouche, P. (2025). Legal Framework for an Effective Implementation of the AI Act. CERRE. https://cerre.eu/wp-content/uploads/2025/02/Legal-Framework-for-an-Effective-Implementation-of-the-AI-Act_FINAL.pdf.

approach to regulation calls for additional regulatory guidance and an adequate institutional framework to facilitate effective implementation and reduce legal uncertainty.[7]

This issue paper focuses on a specific aspect of better regulation: the principle of technological neutrality. It assesses the degree to which the AI Act complies with the principle of technological neutrality, looking specifically at the question of how well the AI Act is designed to adapt to changes in the 'AI value chain': both those changes already occurring in the market today and those that may occur in future. Section 2 provides an overview of the principle of technological neutrality, including in relation to the vertical AI value chain. Section 3 describes the value chain assumed in the AI Act. Section 4 then explains how current and potential developments affect and challenge these assumptions. Section 5 provides policy recommendations on how the AI Act can be implemented and enforced in a way which best promotes technological neutrality, and areas where the law may require updates or amendments. Without greater attention to technological neutrality, policy-makers risk creating an unpredictable regulatory environment which will need constant changes. In turn, this risks unnecessarily stifling innovation and making it more difficult for AI firms to adapt to their customers' needs.

[7] Larouche, 2025; Schnurr, D. (2025). Effective Implementation of Requirements for High-risk AI Systems Under the AI Act: Transparency and Appropriate Accuracy. CERRE. https://cerre.eu/wp-content/uploads/2025/02/Effective-Implementation-of-Requirements-for-High-Risk-AI-Systems-Under-the-AI-Act_FINAL-1.pdf.

2. The principle of technological neutrality

2.1 Background and history

The principle of technological neutrality is embedded in the European Commission's better regulation toolbox[8] and has been applied for many years in the EU's telecommunications, data protection, and cybersecurity laws.[9] Behind the crisp and evocative "technological neutrality" label, one finds at least three main constructions: [10]

- Technological neutrality can be seen as an application of the *non-discrimination* principle in matters relating to technology. The law should not discriminate as between technologies. In other words, functionally equivalent technologies should be treated in the same way. From the perspective of the law-maker, this implies that the same regulatory principles should apply to the same types of market actors regardless of the technology they use. From the perspective of the addressees, laws should describe the results to be achieved but should leave firms and users free to adopt the technology of their choice to achieve the required result. This "level-playing field" interpretation of technological neutrality is valuable; however, the same outcome could also be derived, albeit with a few more steps in the reasoning, by working from the general principle of non-discrimination.
- A second construction of technological neutrality assigns it a more distinctive meaning, centred on *legislative sustainability.* Higher law-making bodies, such as legislatures, lack the time and resources to revisit legislation frequently (say, every six months, which is an eternity in the current phase of AI development) in order to accommodate technological change. To the extent possible, law should be future proofed rather than becoming anachronistic.[11] Furthermore, it would not be conducive to stability if law was tinkered with so often. Here technological neutrality means that law is framed in such a way as to be able to withstand technological evolution over time.
- Technological neutrality has a more substantive dimension in its third construction, whereby the law should not determine the path of technological innovation or should even avoid curtailing potential innovation paths. This *non-interventionist* stance is summed up in the often-quoted slogan "the state should not be picking winners". Instead, under the umbrella of a technologically neutral law, firms should be able to bring a variety of inventive solutions to the market. The fate of these solutions should be determined by the decisions of customers as to whether to adopt a given solution or not. This approach can promote competition – allowing firms to experiment with different technologies and allowing users to select the most

[8] https://commission.europa.eu/document/download/9c8d2189-8abd-4f29-84e9-abc843cc68e0_en?filename=BR%20toolbox%20-%20Jul%202023%20-%20FINAL.pdf p 176.
[9] See for instance Directive 2018/1972 (European Electronic Communications Code) [2018] OJ L 321/36.
[10] See Ilse van der Haar, "Technological Neutrality: What Does It Entail?" TILEC Discussion Paper 2007-009 (2007) and Anna Butenko and Pierre Larouche, "Regulation for Innovativeness or Regulation of Innovation?" (2015) 7 Journal of Law, Innovation and Technology 52-82. See also Maxwell and Bourreau, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2529680.
[11] E Puhakainen and KE Väyrynen, "The Benefits and Challenges of Technology Neutral Regulation: A Scoping Review" (2021) PACIS 2021 Proceedings 48. Available at https://aisel.aisnet.org/pacis2021/48/.

attractive and best value solutions. Competition in turn provides incentives to invest in more efficient and/or effective approaches to compliance.

Achieving technological neutrality is particularly challenging in the context of fast-moving technologies like AI, since the risks of the technology and its economic and social impacts are far from being well-understood. Nevertheless, technological neutrality is important both to protect the effectiveness of regulation, but also – importantly given current concerns about the impact of regulation on EU competitiveness – to ensure regulation is as consistent as possible with promoting investment, innovation and competition.

Even if the three constructions of technological neutrality set out above may seem quite different, in practice they lead to converging recommendations to lawmakers. These include:

- *Relying on functional or economic rather than technological definitions:* For instance, the definition of "electronic communications networks", in Directive 2002/21,[12] was part of a first attempt at setting technologically neutral definitions in EU law. Stripped of the enumerations aimed to signal that no technology was excluded, the definition comes down to "transmission systems... and other resources which permit the conveyance of signals by... electromagnetic means... irrespective of the type of information conveyed". Leaving aside the limitation to electromagnetic signals,[13] this definition focuses on functions as opposed to technologies. Since its adoption in 2002, it has accommodated many different technologies, as they arose.
- *Focusing on outcomes:* Legislation specifies the results that it hopes to achieve, rather than how regulated firms should achieve those results. In the context of AI regulation, for example, that may mean requiring the risks of the technology to be addressed (such as risks that an AI System's outputs are discriminatory or dangerous) instead of specifying how an AI System must be designed or trained. As the EU's better regulation toolbox explains, '[e]xcessively prescriptive and detailed regulation can create barriers to entry for innovative solutions, even if the innovation could contribute to achieving the policy goal of regulation'.[14]
- *Taking a principles-based approach to regulation:* Building on the previous recommendation, laws tend to be more technologically neutral and future-proof when they focus on broad, overarching principles rather than being excessively prescriptive, or including implicit assumptions about technological models or solutions. This would imply that legislation would be more general and shorter, leaving details to be elaborated elsewhere.
- *Relying on independent regulatory authorities to bridge the gap between general legislation and evolving realities:* The principle of technological neutrality ties in with the now-mainstream institutional setup in EU regulation, whereby such general legislation, focusing on principles, is then further developed and implemented by independent regulatory authorities. These authorities have the resources and expertise to handle technological evolution, and

[12] Directive 2002/21 on a common regulatory framework for electronic communications networks and services (Framework Directive) [2002] OJ L 108/33, Art. 2(a), now Art. 2(1) of the EECC, supra, note 8.
[13] Understandable, given that the Directive is concerned with electronic communications, and not with printed communication, for instance.
[14] European Commission. (2023). Better Regulation TOOLBOX. https://commission.europa.eu/document/download/9c8d2189-8abd-4f29-84e9-abc843cc68e0_en?filename=BR%20toolbox%20-%20Jul%202023%20-%20FINAL.pdf, p. 176.

their implementing instruments can be more readily and quickly adapted to address technological specificities, as the need may arise.

- *Using adaptive approaches to regulation:* For example, the AI Act requires providers of High-risk AI Systems to comply with the law taking into account "the generally acknowledged state of the art" and what is "proportionate".[15] This allows the requirements of the law to be dynamically updated as technological capabilities and commercial realities evolve. Such flexibility and dynamism are often associated with the use of soft-law instruments (recommendations, guidelines, etc.) which are easier to change and are typically subject to consultation as opposed to lengthier lawmaking processes.[16] In addition, industry-driven norms – mostly standards – can also be used to allow bottom-up industry consensus to filter into implementation by giving a concrete translation to the more abstract principles contained in legislation.

These recommendations are not absolute. For example, technological neutrality does not mean that regulation can or should be developed without careful consideration of technological capabilities so that the outcomes that regulation imposes are practical and feasible. Specifically, technical realities may often introduce trade-offs that should be considered when drafting and implementing regulation.[17] Similarly, the principle of technological neutrality does not preclude that certain business models or technological approaches should be treated differently if there is an objective reason to do so – for example, if a particular business model or technology-specific criterion acts as reasonable proxy for the level of risk of a product.[18] Indeed, at the most general level, the mere existence of the AI Act implies that law-makers have decided to single out a broad technological realm (AI) for special supervision. Complex questions may arise, for example, about whether products with similar capabilities should always be regulated to the same extent. In some cases, the benefits of technological neutrality must be weighed against other requirements – such as a desire to ensure law is sufficiently specific to ensure firms cannot 'game' it; the need for efficiency and administrability, which can sometimes justify technology-specific measures; or industrial policy objectives which might aim to support particular technologies on the basis that they promote European 'digital sovereignty'. Furthermore, the more generally formulated regulation associated with technological neutrality may sometimes be more difficult for smaller firms to apply, by imposing greater burdens on them to engage with regulators and to assess the impacts of their technological choices, thus impacting competition.[19]

[15] AI Act rec 64.

[16] For instance, the Commission Guidelines on the definition of AI Systems are less technologically neutral than the AI Act itself. One can argue whether this is a good development, but in any event these Guidelines can and will be revised regularly to adapt them to technological developments.

[17] Fast, V., Schnurr, D., & Wohlfarth, M. (2023). Regulation of data-driven market power in the digital economy: Business value creation and competitive advantages from big data. Journal of Information Technology, 38(2), 202-229; Bourreau, M., Krämer, J., & Buiten, M. (2022). Interoperability in digital markets. CERRE Report. https://cerre.eu/publications/interoperability-in-digital-markets/; Krämer, J., Colangelo, G., Richter, H., Schnurr, D. (2023). Data Act: Towards a Balanced EU Data Regulation. CERRE Book. https://cerre.eu/publications/data-act-towards-a-balanced-eu-data-regulation/.

[18] https://resolve.cambridge.org/core/journals/european-journal-of-risk-regulation/article/technology-neutrality-as-a-way-to-futureproof-regulation-the-case-of-the-artificial-intelligence-act/B4B5FD9D31DEB2B7B31C5745C68032D1#fn7.

[19] Maxwell and Bourreau, *supra* note 9.

Consistent with principle of technological neutrality not being absolute, the AI Act does involve some technology-specific choices. For example, the AI Act treats a general-purpose AI model (GPAI Model) presenting systemic risk if it has 'high-impact capabilities', which the Act approximates by reference to a cumulative amount of computation used for the training of the model, measured in floating point operations.[20] The Commission is empowered to adjust this threshold over time. Reliance on technology-specific thresholds as a proxy might be understandable given the demand from industry for specific and measurable thresholds. However, as we discuss further in Section 4, this particular technology-specific characteristic (the computing power required to train a model) may not be particularly enduring as a proxy for a model's capabilities, given the proliferation of methods that also allow models with a smaller number of parameters, requiring less training compute, to achieve capabilities that were recently unattainable.[21]

For the principle of technological neutrality to deliver effective regulation requires a particular type of relationship between regulators and regulated firms – which will often involve delegating more responsibility to regulated firms.[22] Technologically neutral regulation requires that regulated firms avoid a 'checkbox' approach to compliance and engage in open dialogue with regulators about how to apply regulatory principles to new contexts rather than exploiting information asymmetry. In turn, regulators have a responsibility to provide ongoing, up-to-date guidance and to take an approach to enforcement which rewards firms for open and good faith engagement.

2.2 The vertical dimension of technological neutrality

To date, the principle of technological neutrality has been discussed and applied mostly in what can be described as "flat" or "horizontal" settings. From a static perspective, the main concern is that various technologies are presently available to fulfil a given function, and the law should remain neutral among them, unless there is an imperious reason to do otherwise. From a dynamic perspective, technologies evolve over time, and such evolution may follow any one of several potential innovation paths, the direction of which cannot be predicted in advance.

However, technological neutrality also has a vertical dimension. Assume, for the sake of argument, that two or more technological paths can fulfil the same function, yet involve different vertical relationships. For instance, as a matter of technology, office productivity software can be delivered physically on a workstation or hosted in the cloud on a Software-as-a-Service basis, with either option available in proprietary or open-source format. Presumably, the three constructions of technological neutrality (non-discrimination, legislative sustainability and non-intervention) apply with equal force in such a setting. Yet these technological choices, while left to firms and customers, also have business implications, since they are associated with different value chains.[23]

[20] AI Act art 51(2).
[21] See Section 4.1 and also Belcak, P., Heinrich, G., Diao, S., Fu, Y., Dong, X., Muralidharan, S., ... & Molchanov, P. (2025). Small Language Models are the Future of Agentic AI. https://arxiv.org/pdf/2506.02153.
[22] Rebecca Crootof and BJ Ard, 'Structuring Techlaw' (2021) 34 Harvard Journal of Law & Technology 347.
[23] These business implications are to some extent also present in the "flat" version of technological neutrality, but they are not so salient. In the vertical dimension, technological choices by firms and customers will typically lead to different value chains and therefore different business models, and business lines.

Accordingly, one could argue that technological neutrality should also extend to situations where technological choices and innovation paths have a "vertical" dimension. In such cases, proper care should be taken to avoid undermining technological neutrality through the back door by making assumptions about value chains in legislation or regulation. For example, that may mean avoiding making specific assumptions about how value chains in AI are arranged and the distribution of responsibilities between different players, particularly when the technology which may sometimes dictate the design of a value chain can itself be an important parameter of competition. In other words, technological neutrality might include a form of "value-chain neutrality", at least in circumstances where the value chain is influenced by the choice of technology deployed (and not just by choices of business model or commercial strategy). Neutrality in this case is all the more important in sectors such as the digital economy (broadly construed), where disruptive innovation is known to occur and has often led to significant welfare gains. Disruptive innovation properly understood[24] is by its very nature inimical to existing value chains, as they may reflect established value networks or dominant architectures. For instance, some of the most consequential disruptive innovations in this century involved the displacement of the Blackberry with the iPhone architecture for smartphones, or the replacement of physical content supports (CD, DVD/BluRay) with streaming. Fortunately, in both cases, neither the Blackberry architecture nor the physical support models were baked into the pre-existing regulatory framework, and accordingly disruptive innovation was not hindered by law.[25]

The history of electronic communications offers a counterexample in point. Directive 2002/21 defined "electronic communications networks" in a technologically neutral fashion, as seen above, but it also added a definition of "electronic communications service" as "service normally provided for remuneration which consists wholly or mainly in the conveyance of signals on electronic communications networks".[26] Here as well, the definition appears technologically neutral. Traditional voice telephony (over the PSTN) was seen as the quintessential electronic communications service, in 2002. Later, over-the-top (OTT) services, including VoIP services such as Skype, were introduced. These services were technically operated at a higher architecture layer (hence the name), and legally they were found to fall outside the definition of "electronic communications service".[27] This regulatory distinction between traditional voice telephony and VoIP might have offered certain advantages from the point of view of liberalisation policy,[28] but as certain regulatory obligations applied only to electronic communications services but not OTT services, even though in many cases the services performed the same functions, it also introduced competitive distortions that were a constant source of recrimination, until the Electronic Communications Code (EECC) removed the distinction.[29]

[24] C. Christensen, *The Innovator's Dilemma* (Boston: Harvard Business School Press, 1997) and more recent account in J. Gans, *The Disruption Dilemma* (Cambridge: MIT Press, 2016).

[25] Although one could argue that the implementation of copyright law was somehow over reliant on the presence of physical media (and the attendant value chain) and could not easily switch to streaming, leading to law-originating frictions in the disruption process.

[26] Directive 2002/21, *supra* note 11, Art. 2(c). This definition has been modified in the successor legislation, the EECC, *supra* note 8, Art. 2(4).

[27] See on this point CJEU, *Skype v. IBPT* ECLI:EU:C:2019:460, para. 42.

[28] It gave a freer rein to operators challenging the incumbents with VoIP services.

[29] The new definition of "electronic communications service" in the EECC *supra* note 11 is meant to encompass VoIP as well. In any event, by the time the EECC was adopted in 2019, that debate had lost much of its salience, given the progressive demise of traditional voice telephony.

3. The AI Act's conception of the regulatory AI value chain

The AI Act was originally conceived around the concept of "AI Systems", discussed under Heading 3.1. below. During the legislative process, a separate regime was added for "General-purpose AI models" (GPAI Models), addressed under Heading 3.2. Under Heading 3.3. we examine the resulting relationship between AI Systems and GPAI Models, while under Heading 3.4 we focus on the responsibilities of the various actors in the AI value chain under the AI Act.

3.1 AI Systems

As originally proposed, the AI Act was built around the concept of "AI System". Its definition was the subject of considerable debate, as leading jurisdictions sought to coordinate their approaches in order to avoid regulatory fragmentation from the outset. Using the OECD as a discussion forum, a common definition was agreed to, which the AI Act closely tracked:

> *a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*.[30]

The definition therefore relies on the concepts of 'autonomy', 'adaptiveness' and a system's ability to infer how to generate outputs, rather than on the specific technologies used in the system's creation or operation. The AI System definition exemplifies the use of functional descriptions, which are characteristic of technological neutrality (see Section 2.1). In particular, the definition does not make assumptions about how these functions are realised in the system's implementation. For example, it remains agnostic as to whether the ability to infer arises from a logical framework (symbolic AI), from training on large labelled datasets (supervised learning), or from trial-and-error interactions with the system's environment (reinforcement learning).[31] It was designed to last, to avoid pre-determining or influencing technological evolution[32] and, last but not least, to pre-empt time- and resource-consuming discussions over whether a given technology falls under the AI Act or not. In explanatory documents, the OECD developed its definition in greater detail, with the help of the following illustration:

[30] AI Act art 3(1).
[31] See also European Commission (2025). Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act). https://ec.europa.eu/newsroom/dae/redirection/document/112455
[32] AI Act, Rec. 12.

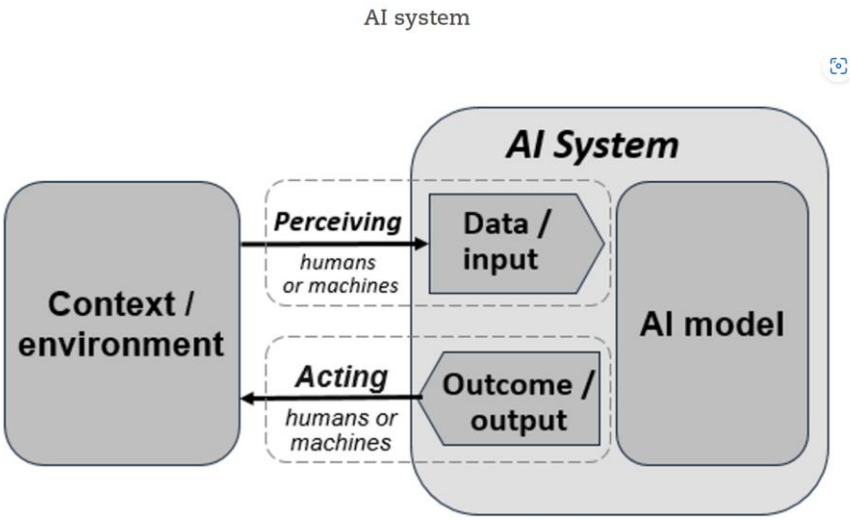

Figure 1: Stylised conceptualised view of an AI System (OECD)

As the illustration indicates, an AI System as defined by the OECD is built on an AI model, together with data input and outcomes output (see Figure 1). It draws its data from, and produces output for, a given context or environment.

Much like the OECD, the Commission in its original AI Act proposal chose to base the entire regulatory edifice on this concept of AI System.[33] An AI System is then the unit of development and marketing by 'providers', who are presumably overseeing its model, input and output components. Such a system may either be released as a standalone product or embedded in another product. Because that AI System relates to a given context or environment, the OECD and EU definitions imply that AI Systems are developed with a specific purpose or application in mind, ranging from, say, operating industrial machinery to assisting judicial decision-making.

Relying on that conception of an AI System relating to a specific context or environment as the basic unit of regulation, the AI Act proceeds to build an elaborate system of risk-based regulation around four tiers.[34] Stricter regulation applies where there is higher risk.

The strictest tier involves outright prohibition. In this case, risk is determined by reference to a system's capabilities and potential uses, in a (mostly) technologically neutral way. For example, AI Systems are prohibited in certain cases where they use 'purposefully manipulative or deceptive techniques', exploit certain vulnerabilities, or engage in 'social scoring'.[35]

The second tier subjects 'high-risk' AI Systems to a regime modelled on EU product regulation (the New Legislative Framework), with pre-market compliance and post-market surveillance.[36] Here, 'risk' is largely determined by using the functionality or the intended uses of the system as a proxy.[37] For example, AI Systems are high-risk in certain cases where they are 'intended to be used as a safety

[33] Which is introduced from the very beginning at Rec. 1 of the AI Act.
[34] Often illustrated in pyramidal form, to reflect the assumption that the strictest tiers also cover the fewest AI Systems.
[35] Art. 5 AIA.
[36] Larouche, 2025.
[37] See Annex III where all high-risk systems are described by reference to how they are 'intended to be used'.

component of a product'.[38] As stated above, the assumption is that an AI System will have a particular intended purpose, and hence that the AI Act can already make a determination of risk on that basis and assign certain use-cases to the 'high-risk' category. It is this second tier of High-risk AI Systems that has attracted much of the attention in the debates around the AI Act.[39]

The third tier involves transparency obligations for AI Systems interacting directly with humans or generating audio, image, video and text content.[40]

The fourth tier comprises all other AI Systems and does not provide for any additional pre-market regulation.

Following from the above, the AI Act assumes that AI Systems would be integrated into relatively predictable and static vertical value chains, involving:

- 'providers' that develop an AI System and place it on the market (along with distributors, authorised representatives and importers, all players involved in making AI Systems available in the EU);
- 'deployers' which use AI in their commercial or professional activities, but which (along with any other parties) can be treated as 'providers' of a High-risk AI System if they modify the intended purpose of an AI System so that it becomes a high-risk system, or make a 'substantial modification' to an existing High-risk AI System;[41] and
- 'product manufacturers' which incorporate AI Systems in certain types of regulated products, and can sometimes be treated as the 'provider' of a High-risk AI System in their own right.[42]

This brief overview shows how the concept of AI System relating to a given environment or context, viewed as a regulatory unit, percolates through the architecture of the AI Act, as it was originally proposed.

3.2 General-purpose AI Models

In November 2022, in the midst of the legislative process on what was then the proposed AI Act, OpenAI launched its generative AI product, ChatGPT. ChatGPT as such qualifies as an AI System and is based on GPT, an AI model developed by OpenAI.[43] ChatGPT marked a technological breakthrough, and it caught the attention and imagination of the general public. At the same time, it threw into doubt two basic assumptions of the proposed AI Act. Firstly, ChatGPT showed that AI Systems are not necessarily purpose-specific: a defining characteristic of the GPT model used by ChatGPT and similar products is its ability to have or develop unforeseen uses and capabilities.[44] Secondly, ChatGPT

[38] Art. 6(1)(a) AIA.
[39] Together with the regime for GPAI Models, discussed below.
[40] Art. 50 AIA.
[41] AI Act recital 84 and art 25.
[42] Art 25(3) AIA.
[43] GPT stands for Generative Pre-trained Transformer. It is a large language model (LLM). While the term and its abbreviation is the technical designation of a class of AI models, GPT has now become associated with OpenAI's family of models. ChatGPT grafts a chatbot interface on GPT. At the time ChatGPT was launched in November 2022, the underlying model was GPT-3.5. OpenAI has since developed new and more powerful LLMs. The most recent one is GPT-5.
[44] https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/.

highlighted the distinction between AI Systems and the models on which they run. Whereas the proposed AI Act was based on the assumption that firms would develop and market AI Systems, the ChatGPT model developed by OpenAI could conceivably be integrated into another set of input-output interfaces than a chatbot. These interfaces could also be part of an AI System made by a third-party producer.[45] This opened the door to a segmentation between AI models and the AI Systems built on and around these AI models, and therefore to a more complex value chain.

Conceivably, these developments could have been accommodated within the conceptual framework of the AI Act. Since any risk is made concrete when AI Systems are used for a certain purpose, and an AI model requires an input/output interface – in other words to be part of an AI System – to be useful,[46] responsibility could remain with AI System providers, irrespective of whether they source their AI models from a third-party or not. Multi-purpose AI Systems could be subject to different regulatory classifications depending on the purpose for which they are used (and the corresponding level of risk). At the same time, a tension was already emerging between prospective AI System providers and providers of AI models: AI System providers were arguing that they could not be responsible for risks arising from the AI model, whilst AI model providers were responding that they could not be responsible for risks that are only actualised once their AI model is integrated in a specific AI System. The prospect of reciprocal blame-shifting was not appealing to lawmakers. Furthermore, there was a perception that certain risks arose from the AI models as such, irrespective of any purpose for which they would be used. Finally, the prospect of multi-purpose AI Systems or AI models facing a set of fragmented and possibly conflicting regulations, depending on the purpose for which these systems or models are used, also clashed with the legislative intent to provide certainty through the AI Act.

As a result, the EU institutions decided to insert in the AI Act a new concept next to that of AI Systems, namely GPAI Models. A parallel regulatory regime was created for GPAI Models, based on whether the model carries 'systemic risk'. Unlike AI Systems which are supervised at Member State level, the European Commission (more specifically the AI Office) is in charge of policing GPAI Models. A GPAI Model is defined as:

> *an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market*.[47]

As for AI Systems, this definition seems largely technologically neutral, focusing mostly on the model's capabilities and generality, rather than the technological solutions used to produce it, even if the definition does suggest that (many such) GPAI Models will be trained using data and self-supervision.[48]

[45] For instance, OpenAI's GPT is used as a model to run Co-pilot, an AI System developed and marketed by Microsoft.
[46] AI Act recital 97.
[47] AI Act art 3(63).
[48] While self-supervised deep learning methods have been the primary drivers of the capabilities of today's state-of-the-art GPAI models, alternative AI approaches focus on deductive logic and trial-and-error learning to enhance the reasoning abilities of AI models.

However, Recital 98 of the AI Act does introduce some technological specificity: it states that models with (among other things) at least a billion parameters should be considered to fulfil the criteria of displaying '*significant generality*' and to be '*capable of competently performing a wide range of distinct tasks*'.

Nevertheless, the European Commission was under pressure to provide more specific criteria to enable firms developing AI models to ascertain whether their models qualify as GPAI Models or not.[49] In its recent Guidelines on the scope of the obligations for GPAI Models,[50] the Commission broke with technological neutrality when it determined that an 'indicative criterion' for whether a model would be considered a GPAI Model is that its training compute is greater than 10^{23} floating-point operations (FLOP) and it can generate language, text-to-image or text-to-video. This is typically met by models that are trained on large datasets and have 1 billion parameters or more.[51] In addition, the guidelines require that the model exhibits significant generality and is capable of competently performing a wide range of distinct tasks. The term 'indicative criterion' recognises that the number of FLOP is an 'imperfect proxy for generality and capabilities' and the guidelines provide examples of models which exceed the 10^{23} threshold but which nevertheless should not qualify as GPAI Models (for example because their range of output is limited).[52] The guidelines also acknowledge that the Commission's approach might change in future even if reliance on compute thresholds seems 'the most suitable approach at present'.[53]

As with AI Systems, there is a tiered approach to the regulation of GPAI Models, with different obligations applying depending on whether the model carries 'systemic risk' and whether it is open source. The AI Act treats a general-purpose AI model as if it presents systemic risk if it has 'high-impact capabilities'.[54] The AI Act itself provides a technological proxy for this criterion, here as well by reference to a cumulative amount of computation used for the training of the model. Specifically, a general-purpose AI model is presumed to have high-impact capabilities when the cumulative amount of computation used for its training is greater than 10^{25} FLOP.[55] The Commission is empowered to adjust this threshold over time.

The specification of these unidimensional quantitative thresholds, for both the definition of GPAI Models and for the sub-set of GPAI Models with systemic risk, represents a significant shift away from technological neutrality, and has given rise to significant debates about whether computational power used for training is a sufficient, or even a relevant, metric to estimate a model's level of risk. Reliance on technology-specific thresholds like computational power as a proxy might be understandable given the demand from industry for specific and measurable thresholds, and the lack of clarity today about which risks might arise from foundational AI models. However, as we note below, developments in

[49] This is a significant issue since AI models that would not qualify as GPAI Models are not covered by specific obligations under the AI Act.
[50] European Commission. (2025). Guidelines on the scope of the obligations for general-purpose AI models established by Regulation (EU) 2024/1689 (AI Act). C(2025) 5034 final.
[51] See Guidelines on the scope of the obligations for general-purpose AI models AI, No. 15 and AI Act, Rec. 98.
[52] Guidelines, text box after para 20.
[53] Guidelines paras 15-16.
[54] AI Act, Art. 51(1).
[55] AI Act, Art 51(2).

the AI sector raise questions about whether these thresholds are likely to remain useful proxies for risk, especially over the long run.

3.3 The relationship between GPAI Models and AI Systems

With the introduction of the concept of 'GPAI Models' in the AI Act, lawmakers also added an additional explicit link to the Act's conception of the value chain, namely the relationship between GPAI Models and AI Systems. More generally, as outlined above, the AI Act proposal already recognised that an AI provider typically depends on third-party suppliers of tools, services, components, and processes.[56]

However, more specific provisions became necessary to describe the relationship between GPAI Model providers and providers of AI Systems, given that both sets of providers are subject to different respective sets of specific obligations in the AI Act. Defining the relationship is problematic, though, since the technical and commercial relationship between GPAI Models and providers of AI Systems is complex and still changing. The AI Act treats GPAI Models primarily as components of AI Systems:

> *Although AI models are essential components of AI Systems, they do not constitute AI Systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI Systems.*[57]

Consequently, most GPAI Model providers must make available documentation to downstream companies that integrate the model into their AI Systems, so that those downstream providers can comply with their own obligations in the Act.[58]

It will be recalled that GPAI Models challenge the original architecture of the AI Act not only by introducing a potential new link in the value chain, but also by challenging the assumption that AI Systems (and their components) are always tied to a specific purpose. As a practical matter, it is possible that the general-purpose quality of a GPAI Model carries over into the AI System in which such model is embedded.[59] The AI Act acknowledges this with a new – and somewhat subdued – concept of general-purpose AI System (GPAI System), that was introduced at the same time as the GPAI Model regime was inserted. An AI System that integrates a GPAI Model becomes a GPAI System if "due to this integration, this system has the capability to serve a variety of purposes".[60] A GPAI System can be used directly, or it may be integrated into other AI Systems. GPAI Systems are subject to several specific obligations – if the GPAI System can be used for a high-risk purpose, market surveillance authorities must carry out evaluations of that system;[61] and providers of GPAI Systems must cooperate with providers of High-risk AI Systems to enable the latter to comply with the AI Act.[62]

[56] See Art 25(4) and also Section 3.4.
[57] AI Act, Recital 97.
[58] Art 53(1)(b). Exception for open source.
[59] See also Section 4.4 for a discussion of agentic AI Systems.
[60] AI Act, Recital 100.
[61] AI Act, Rec. 161 and Art 75.
[62] Recital 85.

The AI Act also acknowledges that GPAI Models may be "further modified or fine-tuned into new models"[63] but does not describe when or how a third party that alters an existing GPAI Model would become the 'provider' of such model. To this end, the Guidelines on the scope of obligations for GPAI Models specify that when the training compute used for the modification of the GPAI Model (e.g., through fine tuning) is greater than a third of the training compute of the original model, this constitutes an indicative criterion for significant changes of the model's generality, capabilities, or systemic risk. In such cases, the downstream modifier becomes the provider of the modified GPAI Model.[64] If the original training compute is unknown to the downstream provider, the thresholds should be replaced with a third of the 10^{25} threshold for GPAI Models with systemic risks or of the 10^{23} threshold for GPAI Models, respectively.

With respect to this threshold, the guidelines further state that "the criterion is […] primarily forward-looking, and in line with the risk-based approach of the AI Act. Therefore, the Commission's approach may change in the future as technology and the market evolve."[65]

3.4 Responsibilities across the AI Act value chain

The AI Act explicitly acknowledges that AI Systems are supplied by a value chain of multiple actors.[66] While most obligations for risk mitigation of High-risk AI Systems under the AI Act fall on the providers of such systems (see Art. 16), the AI Act lays out additional provisions on the cooperation between the actors involved. This complements individual obligations for other operators along the value chain (including product manufacturers, deployers, authorised representatives, importers and distributors) as well as for GPAI Model providers. Several of these obligations concern the post-market phase, i.e., after an AI System has been placed on the market or put into service, as cooperation across actors is particularly important to identify and mitigate risks at this stage.

Responsibilities between the providers of High-risk AI Systems and third-party input suppliers

Art. 25(4) AI Act specifies the duties between the provider of a High-risk AI System and any third-party supplying tools, services, components (including models), processes, or the underlying AI System that are used or integrated in the High-risk AI System. In particular, the parties must specify by written agreement the necessary information, capabilities, technical access and other assistance that the third party will provide to the provider of the High-risk AI System. This specification must be based on the generally acknowledged state of the art. To support implementation, Art. 25(4) further states that the AI Office may develop and recommend voluntary model terms for contracts that can serve as templates for these agreements. Third parties that provide their inputs under a free and open-source license are exempted from this duty unless the input they provide is a GPAI Model. It is further clarified that intellectual property rights, confidential business information and trade secrets must be observed with regard to the duties of the involved parties (see Art. 25(5) AI Act).

[63] Recital 97.
[64] European Commission. Guidelines on the scope of the obligations for general-purpose AI model.
[65] European Commission. Guidelines on the scope of the obligations for general-purpose AI model, No. 67.
[66] See Recital 83 and Larouche, 2025; Schnurr, 2025.

Distributors, importers, deployers or third parties may become providers of High-risk AI Systems themselves under conditions laid out in Art. 25(1) AI Act. In this case, the original provider of the AI System "shall closely cooperate with new providers" making available necessary information, expected technical access and other assistance required for fulfilment of the obligations set out by the AI Act, unless the provider has clearly specified that its AI System is not to be changed into a High-risk AI System (Art. 25(2) AI Act).

Provision of information and documentation by GPAI Model providers to downstream AI System providers

Providers of GPAI Models must supply up-to-date information and documentation to AI System providers integrating such models, enabling them to understand the models' capabilities and limitations and to meet their obligations under the AI Act, with the minimum scope defined in Annex XII.[67] Providers of open-source GPAI Models are exempt from this obligation where the models do not present systemic risks.[68] The Transparency Chapter of the Code of Practice for GPAI Models provides a model documentation form intended to serve as template for sharing information and document with downstream providers of AI Systems, among other information to be shared with the AI office or national competent authorities.[69]

Post-market monitoring and serious incident reporting across the value chain

With respect to risk management for High-risk AI Systems, the AI Act acknowledges the need to consider the entire lifecycle of AI Systems. Given the unique properties of AI, it is generally difficult to fully identify and mitigate risks ex-ante. Several commentators have thus pointed to the important role of effective post-market monitoring for risk mitigation.[70] Hence, swift identification and correction of risks and harm that materialise after deployment is critical.

[67] Art. 53(1)(b) AI Act.
[68] Art. 51(2) AI Act.
[69] Code of Practice for General-Purpose AI Models. Transparency Chapter. https://ec.europa.eu/newsroom/dae/redirection/document/118120; Model Documentation Form https://ec.europa.eu/newsroom/dae/redirection/document/118118.
[70] Schnurr, 2025; Weidinger, L., Rauh, M., Marchal, N., Manzini, A., Hendricks, L. A., Mateos-Garcia, J., ... & Isaac, W. (2023). Sociotechnical safety evaluation of generative AI Systems. Available at https://doi.org/10.48550/arXiv.2310.11986.

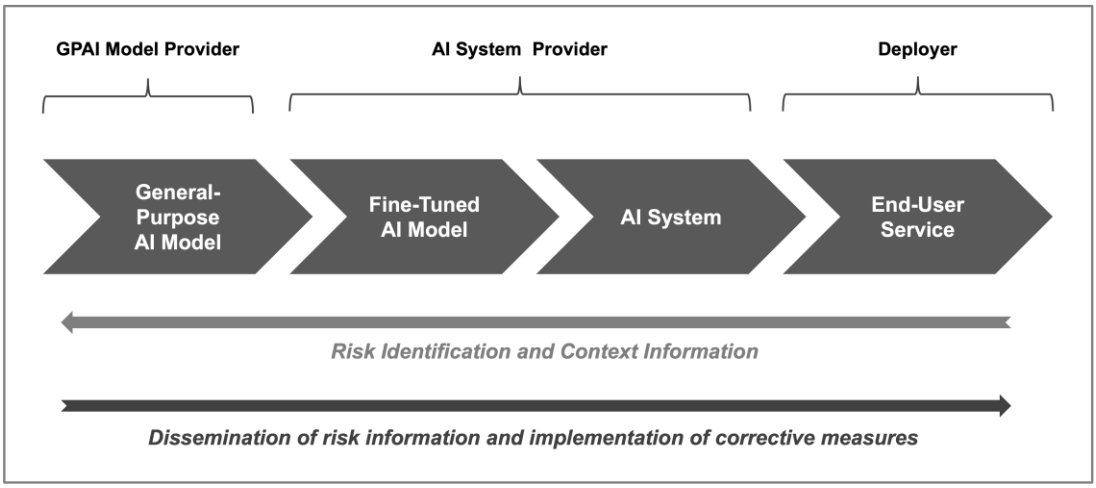

Figure 2: Stylised and simplified view of the AI value chain and shared responsibilities for post-market identification and mitigation of fisks and incidents

Post-market monitoring for providers of AI Systems in general

As a starting point, the market surveillance regime of Regulation 2019/1020[71] is made applicable to all AI Systems falling within the scope of the Regulation, irrespective of which tier of the risk-based pyramid they fall under, i.e. whether they are High-risk AI Systems or not. While that Regulation does not require the introduction of monitoring systems, it nevertheless puts providers under a duty to inform these authorities if they have reason to believe that their AI System presents a risk to health, safety or fundamental rights.[72] Providers must also fully cooperate with the market surveillance authorities both at the inquiry and at the remedial stage.[73]

Specific additional monitoring provisions for providers of High-risk AI Systems

In addition to the above, according to Art 72(1) AI Act, providers of High-risk AI Systems must establish a post-market monitoring system, to collect and review experience gained from the use of their High-risk AI System after it was placed on the market or put into service, in order to identify "any need to immediately apply any necessary corrective or preventive actions" (Art. 3(25) AI Act). In this vein, the post-market monitoring system is also envisioned as a key measure to efficiently and timely address the risks that may emerge from AI Systems which continue to 'learn' after deployment (this may, for example, apply to AI agents as discussed in Section 4.4).[74] Accordingly, the AI Act calls for an analysis of the interaction with other AI Systems as part of post-market monitoring.

The post-market monitoring system shall be based on a post-market monitoring plan (Art. 72(3)) and is part of the broader quality management system that providers of High-risk AI Systems must put in place under Art. 17(1)(h) AI Act. A description of the post-market system, together with the post-market monitoring plan, must be included in the technical documentation required for High-risk AI

[71] Regulation 2019/1020 on market surveillance and compliance of products [2019] OJ L 169/1, as subsequently amended.
[72] Regulation 2019/1020, Art. 4(3)(c), combined with AI Act, Art. 79 (1).
[73] Regulation 2019/1020, Art. 4(3)(b) and (d), combined with AI Act, Art. 79.
[74] Recital 155 AI Act.

Systems (Art. 11(1) and Annex IV(9) AI Act. Insights generated from the post-market monitoring system regarding possibly arising risks should feed into the provider's overall risk management system (Art. 9(2)(c) AI Act). Moreover, its operation is to be facilitated by record-keeping, i.e., the automatic recording of events (logs) over the lifetime of the system (Art. 12(1) and (2)(c) AI Act), although the AI Act does not explicitly state who can access these records and when they would need to be transferred to other actors in the value chain.

The creation and documentation of the post-market monitoring system should be "proportionate to the nature of the AI technologies and the risks of the High-risk AI System" (Art. 71(1) AI Act). Its main purpose is the active and systematic collection, documentation and analysis of data to allow the provider to evaluate the performance of High-risk AI Systems through their lifetime and the continuous compliance with the requirements for High-risk AI Systems (Recital 155). The AI Act recognises that such data may be only available from other actors in the value chain and specifically mentions that this data may be "provided by deployers" or "collected through other sources" (Recital 155).

To facilitate implementation, the Commission shall establish a template for the post-market monitoring plan and the list of elements included through an implementing act by 2 February 2026.

High-risk AI System providers' duties regarding serious incidents

Art. 73(1) AI Act requires providers of High-risk AI Systems to report any serious incident to the market surveillance authorities of the Member States where that incident occurred. A serious incident refers to an incident or malfunction leading to (i) death or serious damage to health, (ii) serious and irreversible disruption of the management and operation of critical infrastructure, (iii) infringements of obligations under Union law intended to protect fundamental rights or (iv) serious damage to property or the environment (see Art. 3(49)). Procedures for such incident reporting are required as part of the quality management system for High-risk AI Systems (Art. 17(1)(i)).

Serious incidents must be reported immediately after the provider has established a causal link between the AI System and the serious incident or the reasonable likelihood of such a link (Art. 73(2)). The period for the reporting shall consider the severity of the incident. In any event, the incident must be reported within 15 days after the provider (or where applicable, the deployer) becomes aware of the incident, although this minimum period may be further reduced in case of a widespread infringement or specific types of serious incidents (see Art. 73(3) and (4)).

Reporting of a serious incident is only the first step, as providers must subsequently perform "the necessary investigations in relation to the serious incident and the AI System concerned", including corrective action and a risk assessment of the incident (Art. 73(6)). Although it is further specified that the provider must cooperate with the competent authorities (and possibly the notified body concerned), the AI Act does not specify a duty to cooperate among actors in the value chain related to such investigations.

While Article 73(7) requires the Commission to provide dedicated guidance on providers' duty to report serious incidents by 2 August 2025, the Commission has only recently issued its draft guidance and conducted a consultation to collect stakeholders' feedback.[75]

Duties for deployers of High-risk AI Systems

The AI Act tasks deployers of High-risk AI Systems with monitoring their operation in accordance with the instructions for use that providers must supply under the transparency obligations of Article 13. In addition, Article 26(5) requires deployers to "where relevant, inform providers in accordance with Article 72," which sets out the framework for providers' post-market monitoring. However, the article does not clarify under which circumstances such relevance is presumed, nor does it specify the type or scope of information that must be shared with the provider beyond the reference to the instructions of use. Art. 72(2) notes that "relevant data […] on the performance of High-risk AI Systems throughout their lifetime" to be collected by providers for post-market monitoring "may be provided by deployers". Thus, the AI Act establishes a general duty for deployers to provide relevant performance data to providers but does not provide more specific guidance on the scope or limits of such data beyond the requirement that it will enable providers "to evaluate the continuous compliance of AI Systems with the requirements" for High-risk AI Systems. Art. 12(2)(b) identifies automated records of events (logs) as a tool to facilitate post-market monitoring. This suggests that deployers must make such records available to providers as part of the relevant performance data, even though the AI Act does not state this explicitly. As Article 26(5) designates the instructions for use as a reference point, providers of High-risk AI Systems may further specify the scope of relevant data to be shared in this document.

With respect to serious incidents, deployers of High-risk AI Systems are required to immediately inform the system provider upon detection, and thereafter notify the importer or distributor as well as the relevant market surveillance authorities.[76] If the deployer cannot reach the provider, the reporting obligations set out in Article 73 apply mutatis mutandis. In addition, deployers must inform the provider or distributor and the relevant market surveillance authority, in case they have reason to consider that the use of the High-risk AI System "may result in that AI System presenting a risk within the meaning of Article 79(1)".

Post-market obligations for providers of GPAI Models

Surprisingly, considering that Art. 74 AI Act imposes post-market obligations on providers of any AI System covered by the AI Act (whether or not high-risk), the AI Act provides no obligations for providers of GPAI Models without systemic risks with respect to post-market monitoring. Nor does it specify any duty to supply information to providers of High-risk AI Systems for the purposes of such monitoring. Thus, duties to provide information may arise only under the more general obligations which apply to all third-party input suppliers set out in Article 25(4). Furthermore, the AI Act contains no specific provisions regarding the reporting of serious incidents by providers of GPAI Models without systemic risks. This omission may reflect an intention to avoid imposing broad obligations to monitor

[75] European Commission (2025). AI Act: Commission issues draft guidance and reporting template on serious AI incidents and seeks stakeholders' feedback. https://digital-strategy.ec.europa.eu/en/consultations/ai-act-commission-issues-draft-guidance-and-reporting-template-serious-ai-incidents-and-seeks.
[76] Art. 26(5) AI Act.

their customers' AI systems on model providers below the systemic risk threshold, thereby reducing the overall regulatory burden and allocating responsibility to the parties with better access to information about whether an incident relates to the model or the system. The post-market obligations contained in the AI Act with respect to GPAI Models thus only concern the sub-set of GPAI Models with systemic risk. Providers of GPAI Models with systemic risks face mandatory provisions on post-market monitoring and serious incident reporting and correction. Art. 55(1)(b) AI Act requires these providers to assess and mitigate possible systemic risks at Union level associated with their models. Recital 114 states that continuous assessment and mitigation of systemic risks can be achieved for example by implementing post-market monitoring and cooperating with relevant actors along the AI value chain, among other measures. The Safety and Security Chapter of the Code of Practice for GPAI Models lays out examples for post-market monitoring methods (including the collection of end-user feedback and conducting frequent dialogues with affected stakeholders), while also establishing a mandatory access provision for an adequate number of independent external evaluators to facilitate post-market monitoring.[77]

In addition, the AI Act requires providers of GPAI Models with systemic risk to keep track of, document and report relevant information about serious incidents caused by the development or use of the model as well as possible corrective measures.[78] Information must be reported to the AI Office and national competent authorities without undue delay. While these provisions establish duties for serious incident direct reporting towards authorities, they do not impose specific obligations for model providers to inform downstream providers or other actors in the value chain about such incidents.[79] This can be particularly problematic if a GPAI model provider becomes aware of an incident (possibly through information reported by a deployer or system provider) that may pose broader risks for other systems built on the same model, while the respective system providers and deployers remain unaware of the risk or underlying vulnerability. Conversely, under the AIA, deployers and providers of AI systems are also not obliged to share information about serious incidents with the provider of the GPAI model that may serve as an input to their system. The wording of the AI Act on serious incidents has been criticised as possibly ambiguous, especially in the context of GPAI Models with systemic risks.[80] In this context, compliance could be facilitated by clarifying and aligning the AI Act terminology with recent proposals from the OECD that explicitly differentiate between actual harms as materialised outcomes (including AI incidents, serious AI incidents and AI disasters) and potential harms (including AI hazards and serious AI hazards).

While the AI Act identifies various explicit roles for actors along the AI value chain, these roles relate primarily to AI Systems as the focal point. By comparison, the role of GPAI Model providers in the value chain and their relationship with other actors is addressed in less detail. This is particularly evident for providers of GPAI Models without systemic risks. Unlike providers of models with systemic risks, they are not subject to specific obligations regarding the establishment or support of post-market monitoring, nor are they required to report serious incidents or corrective measures.

[77] 71 Measure 3.5 in the Code of Practice for General-Purpose AI Models. Safety and Security Chapter. https://ec.europa.eu/newsroom/dae/redirection/document/118119.
[78] Art. 51(1)(c) and Recital 115 AI Act.
[79] 73 See, in comparison, the duties of deployers to inform providers of High-risk AI Systems, among others.
[80] Karathanasis, T. (2024). AI incident notification in the EU AI Act: How does it work and is it effective? Available at https://hal.univ-grenoble-alpes.fr/hal-04844964/document.

In conjunction with the quantitative thresholds for GPAI Models and GPAI Models with systemic risks, this framework may give rise to problematic situations. GPAI Model providers that fall below the systemic-risk threshold are not subject to duties of collaboration or information reporting, even where the use of their models in AI Systems leads to very serious incidents such as the death of people. This may reflect a deliberate choice to limit regulatory burdens on these GPAI Model providers (in the absence of systemic risk). Yet in the end, where serious incidents occur in connection with AI Systems that are neither themselves high-risk, nor resting on GPAI Models with systemic risk, the only reporting obligation under the AI Act rests with the system provider, under the general requirements set out in Regulation (EU) 2019/1020, where the provider has reason to believe that its AI system presents a risk to health, safety, or fundamental rights.[81]

[81] Regulation 2019/1020, Art. 4(3)(c), combined with AI Act, Art. 79 (1).

4. Developments in the AI ecosystem and the regulatory value chain of the AI Act

After significant modifications in 2023 to reflect the breakthrough in LLM models with ChatGPT, the AI Act was adopted in 2024. This meant that the value chain as conceived in the AI Act – the regulatory value chain – was solidified once the AI Act was enacted. Yet there is no reason to presume that the rapid evolution in the AI ecosystem that forced lawmakers to play catchup during the legislative procedure would stop or even slow down once the AI Act was enacted. As outlined in Section 2, the risk of disconnect due to technological evolution is meant to be addressed via the principle of technological neutrality.

This section introduces current developments in the AI ecosystem that create challenges for, or tensions with, the regulatory value chain enacted in the AI Act a year ago. After a general section on interdependence and integration (Section 4.1.), we review more specific issues relating to the compute thresholds used with respect to GPAI Models (Section 4.2.) and the distribution channels for GPAI Models (Section 4.3.), and the emphasis now put on agentic AI (Section 4.4.). In the light thereof, the last section examines how a value-chain neutral approach could look like, using post-market surveillance as an example (Section 4.5).

4.1 Open innovation: Interdependence and integration in the AI ecosystem

Much of the AI sector is characterised by remarkably high levels of public sharing of resources such as know-how, data repositories, libraries, and code, many shared through open-source facilities such as those of HuggingFace. What is more, many AI models now rely even more heavily on outputs, training data, or techniques derived from other models, and many AI firms are integrating products from different developers into new products.[82] From a business perspective, this chimes with the strategic management literature on open innovation and ecosystems.[83] Open innovation emphasises the distributed nature of innovation and seeks to move away from purely vertical models (hierarchy vs. markets) commonly used in the 20th century. It includes organisational structures such as alliances, networks, communities and platforms. Here we will use the term “ecosystem“ to refer to the broader set of actors in AI and their relationships.

For example, large models may be used to produce high-quality training data, to “teach” a smaller AI model about chains of reasoning, or to help fine tune a smaller model to specialise in a particular task.[84] The ability for AI developers to piggy-back off components and outputs of other models appears to explain how some smaller models, including China’s DeepSeek V3 and R1, have been able to achieve very high-performance outcomes in a short time and at relatively low cost. In turn, other models have now adopted many of the techniques used by DeepSeek (such as Mixture-of-Experts architectures).

[82] See Z Meyers and M Bourreau, CERRE, 2025.

[83] See H. Chesbrough, *Open Innovation - The New Imperative for Creating and Profiting from Technology* (Boston: Harvard Business School Press, 2003); Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal, 39*(8), 2255-2276.

[84] Autorité de la concurrence, ‘Opinion 24-A-05 on the competitive functioning of the generative artificial intelligence sector’, 28 June 2024, p 43.

Similarly, an AI firm may offer a product that combines (parts of) different AI models and systems into one software package – for example, where a product with one user interface can draw from different models, depending on which will best answer the user's query.[85]

As the sector is evolving, it is too early to tell whether this will eventually result in:

- A more "conventional" oligopolistic structure with a small number of "master" foundation models controlled by a few firms. In this scenario, these models would serve as essential inputs for many smaller or fine-tuned models, as well as for value-added services provided by independent AI developers, thereby creating a relatively conventional vertical value-chain structure.
- A more "contemporary" organisation along the lines of the open innovation literature, with a significant level of interdependency and feedback between different models and the services provided by AI developers more broadly. This outcome would seem to pose questions about how well adapted the AI Act is to manage complex interdependencies between AI models (which may mean some GPAI Model providers must rely on assurances and information from other GPAI Model providers) rather than assuming a linear vertical value chain. In other words, the characteristics and risks of a GPAI Model will be reliant on various different providers, including in many cases open-source.
- A combination of the two outcomes above. For example, an oligopolistic (or even monopolistic) structure may emerge for the use of AI in particular use cases, for example where that use case relies on datasets which only one market player has access to. This may co-exist with other AI use cases where there may be a more complex set of interdependencies and players. Similarly, there may be a small number of very large models used across many use cases, with a larger number of smaller and more targeted independently developed models.

In the current state of flux, it is difficult at this point to determine which outcome is more likely, and therefore how suitable and effective the AI Act's approach will remain in future. In part, the current situation appears to have arisen because the largest AI developers do not seem able to prevent their models being used to develop other, competing models even when they derive no commercial benefit from such downstream uses (and/or do not currently have incentives to do so, perhaps in order to maximise their influence and importance in the AI supply chain).[86] It is yet to be seen whether the best performing AI models will eventually be able to prevent smaller models from independently adopting the innovations of the best performing AI models at relatively low costs. Such restrictions could plausibly be developed either through technical means, enforceable contractual limitations, or new pricing models (for example higher prices for more extensive queries and outputs).

At the same time and as common economic sense would predict, any attempt to consolidate the AI ecosystem around the more conventional oligopolistic structure (first option above) is bound to be

[85] https://www.ceps.eu/wp-content/uploads/2022/09/CEPS-In-depth-analysis-2022-03_Reconciling-the-AI-Value-Chain-with-the-EU-Artificial-Intelligence-Act.pdf p 13.

[86] In general, model providers may have an economic incentive to allow downstream developers to modify or repurpose their models when the commercial benefits of such use outweigh the potential opportunity costs arising from competition. However, in cases such as DeepSeek-R1, some models appear to have been developed using competitors' models without any explicit agreement or compensation for that development.

met with countervailing moves by actors that seek to escape that fate.[87] The risk that large models become less open and more closed over time appears to be encouraging some AI System providers to rely more on genuinely open-source AI models, for example. Open-weight and open-source AI models have also fundamentally influenced the competitive environment. Today, several high-performing open-source models are available on the market, which developers can freely modify and fine-tune, typically requiring far less accelerated compute than would be necessary if they had to train such models from scratch.[88]

Notably, the Act provides exemptions in relation to AI systems[89] (other than those which are high-risk or prohibited) and GPAI models (other than those with systemic risks)[90] which are "free and open-source", that is "released under a free and open-source licence that allows them to be openly shared and where users can freely access, use, modify and redistribute them or modified versions thereof".[91] However, in practice, many AI models are open – but not to an extent that would necessarily qualify as "free and open-source". For example, model providers may release information about the training datasets but not the full source code for the model or might allow full access to a model but not allow its modification. Given this, the AI Act may create disincentives for models to become (or remain) open. Providers may face significant liability or obligations related to downstream uses of their models, yet adopting open policies may make these uses difficult to monitor and, consequently, make compliance with the AI Act's obligations more difficult.

4.2 Adequacy of Training Compute Thresholds

As pointed out in Section 3.2, the use of quantitative compute thresholds in terms of FLOP for delineating the responsibilities of AI providers breaks with the principle of technological neutrality. Rather than evaluating the qualitative characteristics that may qualify a model as general-purpose or as presenting systemic risks, the classification is reduced to a one-dimensional technical metric.

As noted above, in its Guidelines on the scope of obligations for general-purpose AI models, the Commission uses compute thresholds as an 'indicative criterion' for whether a model is a GPAI model or not (while making it clear that a model is not necessarily a GPAI model merely because it reaches this threshold). The Commission argues that "given the wide variety of capabilities and use cases for general-purpose AI models, it is not feasible to provide a precise list of capabilities that a model must display and tasks that it must be able to perform in order to determine whether it is a general-purpose AI model" (para. 14). While the Commission acknowledges that training compute is an imperfect proxy for generality and capabilities, it nevertheless considers this approach the most appropriate at present (para. 15). FLOP is selected as the indicative measure because this measure is proportional to both the number of parameters in an AI model (which is typically very large for large language and other generative models) and the size of the training data, thereby combining these dimensions into a single

[87] And that irrespective of any legal or regulatory action that would seek to avert an oligopolistic outcome, hence the need to consider existing market forces carefully when designing AI governance.
[88] Meta says that "Tens of thousands of startups are using or evaluating Llama 2 including Anyscale, Replicate, Snowflake, LangSmith, Scale AI, and so many others": see https://ai.meta.com/blog/llama-2-updates-connect-2023/.
[89] AI Act art 2(12).
[90] AI Act art 53(2).
[91] AI Act recital 102.

quantitative indicator. Accordingly, the specific threshold of 10^{23} FLOP for GPAI Models is designed to capture models with at least one billion parameters, as suggested in Recital 98 AI Act, that can generate language, text-to-image, or text-to-video outputs.

As a proxy for generality and capabilities, training compute is further employed in the AI Act to classify GPAI Models that pose systemic risks.[92] Article 51(2) provides that a GPAI Model is presumed to have high-impact capabilities, as defined in Article 51(1)(a), when its training compute exceeds the threshold of 10^{25} FLOP. Here, rather than being an 'indicative criterion', the threshold serves as a legal presumption. Such presumption implies that developers whose GPAI Model exceed the threshold must in any event notify the Commission pursuant to Article 52 AI Act and are immediately put in the position of having to defeat the presumption by demonstrating, pursuant to Article 52(2), that the specific characteristics of the model are such that the model does not possess the high-impact capabilities associated with systemic risk. Article 52(2) itself states that that such a demonstration is expected to succeed only "exceptionally". Recital 111 clarifies that high-impact capabilities refer to "capabilities that match or exceed the capabilities recorded in the most advanced general-purpose AI models" and identifies cumulative training compute, measured in FLOP, as one of the relevant proxies for model capabilities. It is further noted that this threshold "should be adjusted over time to reflect technological and industrial changes". The Commission is entitled to make this adjustment through a delegated act.

Proponents of training compute thresholds for the regulation of GPAI Models have highlighted the correlation between training compute and a model's performance based on the scaling laws identified primarily for large-language models.[93] They further assume a positive relationship between the capability of GPAI models and the risks they pose if misused or if they pursue misaligned objectives. In addition, greater capability is expected to increase both how widely a model will be used and how heavily it will be relied upon. Moreover, it has been emphasised that quantitative compute thresholds can be measured objectively, early in the lifecycle and can be verified by external actors.[94] As such, such thresholds have been advocated as a pragmatic and easily implementable means to detect potentially risky GPAI Models that may then be further scrutinised to determine appropriate risk mitigation measures.

On the other hand, the use of seemingly simple proxies in cases such as this leads to well-known difficulties. First of all, resources are diverted towards the application of the proxy, as opposed to the underlying policy issues. Instead of trying to ascertain whether a given AI model possesses the characteristics that justify including it in the GPAI Model or GPAI Model with systemic risk category, both firms and public authorities focus their efforts on trying to figure out whether the compute thresholds are met. Since these thresholds are not necessarily as clear-cut as one would imagine,[95] resources might also go towards gaming them.

[92] Article 51 and 52 AI Act, discussed in this paragraph, are also discussed in the Commission Guidelines on the scope of the obligations for general-purpose AI models {ref.}.

[93] Heim, L., & Koessler, L. (2024). Training compute thresholds: Features and functions in AI regulation. https://arxiv.org/pdf/2405.10799; Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., ... & Amodei, D. (2020). Scaling laws for neural language models. https://arxiv.org/pdf/2001.08361/1000.

[94] Heim and Koessler, 2024.

[95] See the Commission Guidelines on the scope of the obligations for GPAI Models, *supra*, {...}, Annex at para. 117 and ff.

In addition, and perhaps more fundamentally, proxies are only as good as the approximation they deliver. They almost always introduce Type I (false positive) and Type II (false negative) errors, in cases where the proxy would not be aligned with the underlying concern. Given the policy concerns underpinning the AI Act, it may be presumed that the compute threshold proxies were designed to err on the side of over-inclusiveness and thus to avoid Type II errors (where proxies would apply to leave models out of categories to which they should belong, on a more complete examination).[96] However, in light of more recent concerns about the impact of the AI Act on innovation and the bloc's competitiveness, policy-makers may also be increasingly focused on avoiding Type I errors.

Already today, doubts arise as to the adequacy of training computes as a proxy. The use of training compute assumes that scale is an adequate indication of impact and that, as increasing scale continues to dictate improvement, a training compute threshold will continue to single out the most impactful models. Yet the extent to which scaling laws will persist in the future and even whether they hold today has been controversial.[97] In this context, even proponents of compute thresholds have acknowledged that they could become less useful if scaling laws cease to hold.[98]

While it can be expected that AI models will further increase in scale, and as such their performance will further increase, recent advancements have been primarily achieved by complementary methodological approaches. AI providers are continuously leapfrogging each other in innovation and performance. However, very few AI businesses are profitable today, in part due to the vast cost of training large-scale models. Given the growing cost of developing the most powerful models (and the unclear willingness of customers to pay) many AI providers are exploring further means of differentiation, for example by improving the performance of general-purpose models at certain specific tasks. Furthermore, as mentioned above, this constitutes a predictable reaction by market actors to counter the risk of market power becoming cemented in an oligopoly.

As a result, many AI firms are shifting away from building more models on ever more data (which in turn requires expensive computing power) as the singular paradigm,[99] and have started to explore and rely more on alternative and complementary ways to improve their own or others' models.[100]

These include:

- Distilling smaller models from the large AI models, by having a large model (acting as "teacher") train the smaller model (the "student") to achieve comparable results and performance using a fraction of the resources of the large model.[101] Such knowledge

[96] The discrepancy between the ease of adding GPAI Models to the systemic risk category (Art. 51(1)(b)) and the procedure to take GPAI Models out of that category (Art. 52(2)) suggests that, at least for GPAI Models with systemic risk, the main worry is Type II errors.
[97] Hooker, S. (2024). On the Limitations of Compute Thresholds as a Governance Strategy. https://arxiv.org/pdf/2407.05694.
[98] Heim and Koessler, 2024.
[99] Bertin Martens, 'How DeepSeek has changed artificial intelligence and what it means for Europe', Bruegel, Policy Brief 12/25, March 2025.
[100] The use of synthetic data produced by AI has also been mooted as a way to avoid data bottlenecks, but in practice this approach has not proved as promising as hoped, due to concerns about the quality of synthetic data and its close association with the initial data on which it was based.
[101] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. In: NIPS Deep

distillation is considered to lie at the heart of DeepSeek's success[102] and has become a common practice in the release of most popular GPAI models, which are typically made available in different model sizes.[103]

- Relying more heavily on specific or highly curated datasets necessary to "fine-tune" AI models to work in particular use cases. In these cases, the data essential to fine-tune the model will depend on the intended use case – for example, a business customer wanting to use an AI model to optimise its business practices may want to fine-tune the model on the business's own data. Integrating alternative machine learning methods and approaches, especially reinforcement learning to empower LLMs with better reasoning capabilities.
- Integrating alternative machine learning methods and approaches, especially reinforcement learning to empower LLMs with better reasoning capabilities.[104]
- Relying more heavily on alternatives to more data, for example by instead improving the quality and structure of that data, so that AI models can identify the chain-of-thought that links a particular request or question to an answer, and can replicate that chain-of-thought to produce its own answers.[105] Such data can often be produced through manual categorisation of data or from other AI models (with the effect of increasing AI developers' reliance on other AI models).
- Applying sets of rules (such as not mentioning a competitor's products), which can then make an AI model better adapted to deliver results for a particular deployer or category of deployer.
- Focusing on more specialised tasks and the using a combination of smaller heterogeneous models/systems to establish larger systems achieve general-purpose capabilities (e.g., in the context of multi-agent systems).[106] These smaller models are also more easily deployable on resource-constrained hardware and devices.

These techniques illustrate that smaller models can often achieve performance levels comparable to, or approaching, those of their larger counterparts, particularly when applied to more narrowly defined tasks. However, it is not evident that, where a smaller or lighter model is derived from a larger one and where the original model exceeds the computational threshold for presumed systemic risk but the smaller model falls below it, the smaller or light model necessarily ceases to pose (all of) the systemic risks associated with the original model. The risks attached to the original model are likely to be inherited by the smaller version, even if its size or computational footprint is reduced. In such a situation, the compute threshold proxies could lead to Type II errors (under-inclusiveness).

Learning and Representation Learning Workshop. https://arxiv.org/abs/1503.02531; Gu, Y., Dong, L., Wei, F., & Huang, M. (2023). MiniLLM: Knowledge distillation of large language models. https://arxiv.org/pdf/2306.08543; Gemma Team: Riviere, M., Pathak, S., Sessa, P. G., Hardin, C., Bhupatiraju, S., ... & Garg, S. (2024). Gemma 2: Improving open language models at a practical size. https://arxiv.org/pdf/2408.00118.

[102] Criddle, C. & Olcott, E. (2025). OpenAI says it has evidence China's DeepSeek used its model to train competitor. https://www.ft.com/content/a0dfedd1-5255-4fa9-8ccc-1fe01de87ea6.

[103] See, e.g., Meta (2025). The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation. https://ai.meta.com/blog/llama-4-multimodal-intelligence/.

[104] OpenAI (2024). Learning to reason with LLMs. https://openai.com/index/learning-to-reason-with-llms/

[105] Maarten Grootendorst, 'A Visual Guide to Reasoning LLMs', available at https://newsletter.maartengrootendorst.com/p/a-visual-guide-to-reasoning-llms.

[106] Belcak et al. (2025). Small Language Models are the Future of Agentic AI.

In fact, with respect to techniques such as distillation, quantisation, or merging of model weights, the Commission's guidelines on the scope of obligations for general-purpose AI models state that these methods do not create a new model.[107] Instead, the resulting model is considered part of the original model's lifecycle if the technique is applied by the original provider or on its behalf. However, the same principle does not apply when such techniques are carried out by other providers.[108]

Other "simple thresholds" exist, such as the number of users. To be sure, this threshold is also imperfect, but at least it is technologically neutral and thus more likely to be future-proof, while also having other advantages such as being more straightforward to measure. It reflects the basic notion that risk is proportionate to the number of people exposed. It is also much harder to game.

In the end, the significance of the FLOP threshold proxy contained at Article 51(2) AI Act could be downplayed over time. For sure, that proxy played a useful role in bringing the implementation of the AI Act in motion, by avoiding lengthy discussions around the more elaborate criteria found at Article 51(1) (which still need further specification) and Annex XIII AI Act. The list of GPAI Models with systemic risk could thus be initially populated with a series of GPAI models that seem to fit at least the spirit, for lack of specification about the details of the letter, of Article 51. As the AI Office gains more experience with the application of Article 51, and as it specifies further the content of Article 51(1), one would expect that the categorisation as GPAI Model with systemic risk would be done on a more sophisticated basis. Both the GPAI Model providers and the AI Office should be in a position to gather and analyse the relevant data. Furthermore, the number of GPAI Model providers whose models potentially create systemic risk will probably remain manageable for an agency such as the AI Office, so that the efficiency value of a simple threshold proxy should wane with time once knowledge and expertise accumulate.[109]

Beyond the issues with the compute threshold proxy, the trend away from scaling creates other challenges under the AI Act, including:

- Where models begin as general-purpose but are 'fine-tuned', there may be an increasing 'grey area' which makes it hard to determine if a particular model is still general-purpose or it became purpose-specific (while remaining a model as opposed to an AI System). Smaller models bring us back to something close to the situation before November 2022, where AI System developers were at the centre of the action; and
- Putting more pressure on the need for predictability and clarity about when a GPAI Model is modified to such an extent that the modifier should be treated as the provider.

[107] European Commission. Guidelines on the scope of the obligations for general-purpose AI models. No. 23.
[108] Such modifications may result in the modifier being classified as a downstream provider, as discussed in Section 3.3. However, the quantitative threshold of one-third of the training compute, specified in the Commission's guidelines as an indicative criterion, requires that the modification be identifiable and verifiable in its measurement.
[109] In comparison, the proxy thresholds of the Merger Control Regulation (Regulation 139/2004 [2004] OJ L 24/1) Art. 1 (the notion of" Community dimension") retain their significance given the large number of merger cases involving firms from across the entire economy. Even then, they have been refined through subsequent Commission guidelines.

4.3 Channels to market for AI models and systems

As acknowledged in the AI Act, many providers of AI Systems remain reliant on GPAI Models. GPAI Model developers have a variety of ways of distributing models on the market depending on their business strategies, including (i) making the whole model available for download, for example through a platform or repository; or (ii) only allowing access via application programming interfaces (where the model and source code remain with the provider). API access allows GPAI Model providers, should they wish to do so, to impose contractual or technical terms and conditions on access, relating for instance to monitoring or potentially disallowing certain uses of the model. This implies a much greater scope for supervision and management of risk than full release. However, it also implies that some players in the AI ecosystem could exert considerable influence if these few players are capable of producing the largest and most capable models, especially if other (smaller or more specialised) model providers largely use the methods and outputs of the largest models.[110] This is one of the possible scenarios outlined above at the start of this heading. In such a scenario, competition and innovation could be constrained in the long run.

Many of the large platform operators today also operate 'platforms' for hosting GPAI Models and making them available to AI developers. For model providers, this provides a number of different channels to market, especially since most platforms do not currently appear to demand exclusivity. For downstream users of AI models, this can serve to ensure simpler access to both the models and access to the high-end computing power (often provided by large platform operators themselves, for their AI-hosting services) which is necessary for fine-tuning or performing inference using the models. For example:

- Google's cloud computing platform offers Model Garden, which hosts over 130 AI foundation models;
- Amazon Bedrock allows developers to access numerous AI models from providers such as Meta, Anthropic, AI21, Cohere, and Stability AI; and
- Microsoft Azure AI Model Catalogue hosts over 1,700 AI models for business customers.

In comparison, a platform such as HuggingFace offers access to an even larger set of models, but without the computing infrastructure associated with the platforms listed above.

In keeping with its conventional vertical value chain approach, the AI Act appears to assume that distribution would occur within a one-on-one contractual relationship in the value chain. It does not envisage more sophisticated channels to market such as those listed above, and in particular does not envisage that some players – like the providers of AI model platforms – may in future play a significant role in the AI ecosystem and influence the development of AI models. This is surprising given the significance that providers of platforms play in other areas of EU digital regulation (like the Digital Markets Act) and the role these platforms could potentially play in facilitating the post-market sharing of information along the AI value chain, particularly for disseminating information about serious incidents, critical vulnerabilities, or available model updates. Article 25 of the AI Act, which deals with the allocation of responsibilities across the AI value chain, imposes obligations on firms which supply 'an AI system, tools, services, components, or processes that are used or integrated in a high-risk AI

[110] https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/.

system', but not necessarily on the operators of platforms through which those components (such as AI models) are supplied.

Do these platforms raise concerns with respect to the public policy priorities underpinning the AI Act? How are these platforms to be qualified under the AI Act, if at all – for example would platforms qualify as a 'distributor' under the AI Act, which is defined as 'a natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the Union market'? The difficulty with this categorisation is that the role of distributors is primarily limited to AI systems, whereas platforms generally provide access to AI models. Alternatively, is the operator of a platform hosting a number of third-party models deemed under the Act to be a 'provider' of a GPAI model, on the basis that the platform operator is placing the model on the market? This would not seem correct if the model is available in the market through other means of distribution. Furthermore, many of the obligations applicable to GPAI model providers could likely not be fulfilled by platforms whose primary function is the distribution, rather than the development, of models. Finally, is the operator of a platform deemed to be a 'supplier' (under Art 25) of a tool used in any high-risk AI system? In any event, the appropriate role of a platform operator would seem to depend, at least in part, on the role they play in supervising and/or curating the models available on the platform. In other words, it would seem strange for a pure repository platform like HuggingFace to be regulated in the same way as a commercial platform.

The position of model platform operators similarly raises a number of competition policy questions – such as around self-preferencing and bundling – which are also worthy of further analysis.[111]

4.4 Agentic AI

Agentic AI[112] is widely regarded as a qualitative leap in the evolution of AI.[113] Most major AI providers have recently announced or launched early agentic AI services. However, the market remains at an early stage, as many initiatives are still limited to pilots or proof-of-concept projects rather than large-scale production deployments. Nonetheless, organisations across sectors are actively experimenting with agentic systems, particularly in customer service, process automation, technical support, and IT operations. While rapid growth is widely anticipated, significant uncertainty remains regarding how the technology and its practical applications will develop over time.

Compared to earlier AI Systems, AI agents embody a degree of rationalism that enables them to reason by exploring multiple options (planning and search[114]) and adapt effectively to diverse scenarios and circumstances in order to achieve their objectives. Furthermore, AI agents can directly interact with an environment, often through the use of tools such as software, APIs, and external systems.[115] Consequently, these agents can set and break down complex goals in dynamic

[111] Kramer and Boeston, CERRE, 2025.
[112] Agentic AI is concerned with how state-of-the-art AI can operate as an autonomous agent. It is different from the earlier discussion of "AI agents", which were then conceived as relatively narrow AI Systems designed to automate simple tasks such as organizing an agenda or managing e-mail inboxes.
[113] Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey. *IEEE Access*, *13*, 18912-18936.
[114] Schneider, J. (2025). Generative to agentic AI: Survey, conceptualization, and challenges. Working paper. https://arxiv.org/pdf/2504.18875.
[115] Ibid.

environments, pursuing them through autonomous adaptation and management of their resources. Thus, AI agents can assume a more proactive role instead of simply reacting to human requests relying on strategic planning, information processing and problem-solving. Based on these capabilities, AI agents are envisioned to operate autonomously in real-world settings, addressing unpredictable situations and open-ended problems for which there is no a priori specification of how the task should be accomplished. Specifically, agentic AI can excel in scenarios that demand rapid decision-making, the management of objectives over time, and continuous learning.[116]

From a technical perspective, the integration of reinforcement learning that enables agents to learn through trial and error with the goal to maximise cumulative rewards by interacting with an environment. This allows agents to continuously refine their strategies based on feedback. Goal-oriented architectures enable agents to concurrently pursue multiple objectives and manage priorities and trade-offs between those objectives. Typically, these architectures provide for a modular structure where larger goals are broken down into a hierarchy of sub-goals. In this vein, autonomy does not only apply to the completion of a single goal, but also means that agents can substitute lesser goals and individual strategies to meet larger, more long-term goals. Adaptive control mechanisms enable agents to recalibrate their internal parameters to external changes, making it possible to adjust to data shifts or unexpected disruptions.

This empowers AI Systems to become social actors as well as economic agents with considerable autonomy. While the AI Act highlights autonomy as a distinct characteristic of AI Systems, the high-risk provisions are mostly targeted to use cases, where AI Systems are leveraged to solve some specific and clearly defined task. For instance, the requirement of appropriate accuracy (see Art. 15 AI Act) suggests that there is clear ground truth regarding the task that the AI System is performing and that risks can be identified through measuring whether the systems' outputs diverge from this truth.[117] However, as agentic AI is increasingly starting to operate according to a dynamic open world paradigm (for example, by negotiating on behalf of a user, which may create an ethical conflict where the users' interests are furthered by lying or withholding information from a negotiating party), the applicability and suitability of such provisions are likely to be challenged. Moreover, with increasing autonomy of AI agents, it will become increasingly more difficult to evaluate and mitigate risks ex-ante, especially if agents build on general-purpose AI models to operate across domains.[118]

Agentic AI does not easily fit within the regulatory value chain of the AI Act:

- Firstly, agentic AI is a complete AI System, comprising a model (or even a set of models) combined with data input and outcomes output, as described in Section 2. In that sense, like other AI Systems, it operates at a more applied level than GPAI Models;

[116] Archaya et al. *supra* {...}.
[117] Schnurr, 2025.
[118] Beyond AI agents, other AI systems—particularly those built on GPAI models—can also operate across multiple domains and may therefore fall under various provisions of the AI Act. Moreover, a single product may incorporate several distinct AI systems or models, each subject to different regulatory requirements under the Act. As a result, similar compliance challenges to those identified for AI agents may already arise for such systems and products. A specific challenge in the case of AI agents is their autonomy, which can make it harder to predict which domains they will engage with after deployment.

- At the same time, it is likely that certain types of AI agents will exhibit a degree of purpose-generality that brings them closer to GPAI Models than to AI Systems, as far as the structure of the AI Act is considered. While it is conceivable that agentic AI would be limited to a bounded set of purposes (e.g. health care, finances, etc.), these purposes are likely to be defined more broadly than the use-cases around which the High-risk AI Systems categories are articulated. In any event, given the theory behind agentic AI and predictable market trends,[119] chances are that agentic AI will evolve towards purpose-generality. Agentic AI could perhaps represent a "General-Purpose AI System" within the meaning of the AI Act, but that regulatory category is underdeveloped.[120]
- Finally, as far as risk assessment is concerned, the risks associated with agentic AI could emanate just as much from the deployer (the agent's principal) as from the provider, whereas the AI Act generally places the regulatory burden on the shoulders of the provider rather than the deployer. Compared to other AI systems, the deployer of an AI agent typically assumes a more significant role and exercises greater operational control in defining the environment, objectives, and constraints within which the agent operates. Since many risks stem from the interaction between an agent's autonomy and its deployment context, the deployer is particularly well-positioned to tailor safeguards, permissions, and oversight to that specific context.

Such uneasy fit is more than just a matter of finding the proper definition in which to pigeonhole agentic AI, it is also reflected in the substantive implications of agentic AI for the regulatory scheme of the AI Act.

First of all, the autonomy and general-purpose nature of agentic AI does not sit well with the rules surrounding AI Systems. Agentic AI could perhaps occasionally come close to the line delineating prohibited conduct. Depending on what it is doing, it could fall within one or the other High-risk AI System use-cases (especially those listed at Annex III of the AI Act) and jump between these use-cases from time to time. It will also be interacting with natural persons and generating content, within the meaning of Art. 50 AI Act, from time to time. Conceivably, then, the regulatory framework applicable to agentic AI would shift depending on the task (in a dynamic fashion) or on the customer and the counterparts (here as well in a dynamic fashion). While a single AI System can also fall under several categories under the AI Act depending on the use to which it is put,[121] such use is presumably more stable than in the case of general-purpose agentic AI. It can be argued that agentic AI will be designed to remain compliant throughout the various uses/tasks that it undertakes; in this case, the resulting effect could be that the regime of High-risk AI Systems would effectively spill over to uses that are not high-risk within the meaning of the AI Act. In response, one could argue that the regulatory regime

[119] Principals, be they individuals, firms or other organizations, are unlikely to prefer working with multiple agentic AI products if they are offered the option of having a single agentic AI to support them (provided that the performance is comparable).

[120] 108 Recital 85 contains the most detailed description of "GPAI System" in the AI Act, but few substantive provisions relate to that concept, namely Art. 25 (on the duty to cooperate with providers of High-risk AI Systems) and Art. 75 (on AI Office jurisdiction when a GPAI Model provider uses that model to offer a GPAI System as well).

[121] For instance, a facial recognition AI system can be put to different uses by different customers. Some of these uses are prohibited altogether under Art. 5 AI Act, others would make the AI system fall under the high-risk category, while others still will not be covered by the AI Act.

applicable to agentic AI should come closer to that of GPAI Models. Leaving aside that agentic AI is not a model and therefore that this regime would need to be added to the AI Act, the structure of GPAI Model regulation is not adequate for agentic AI. At the pre-market stage, there is no notion equivalent to "systemic risk" that could be used to single out those agentic AI products that are of greatest concern. In particular, applying the GPAI systemic risk delineation based on the computational proxy for training compute (see Section 4.2) would not meaningfully capture whether an agent entails systemic risks. At the post-market stage, there is no room for making any distinction amongst agentic AI products.

> More fundamentally, the inherent goal complexity and adaptability of agentic AI make it difficult to anticipate the strategies that will be implemented by agentic AI and the risks that will ensue. Any regulatory approach would have to abandon ex ante risk assessment and go more in the direction of seeking an "alignment" between agentic AI and human values. Here as well, the general-purpose nature of agentic AI, combined with its capability to autonomously interact with its environment, makes such a regulatory endeavour difficult, since it involves venturing into basic human ethics and morality (e.g. when is it acceptable to lie?), as opposed to the ethics of certain use-cases. In such a context, robustness will become an even bigger issue as agentic AI may move into unknown environments. As the relevant risks of agentic AI may only be identified once these systems are deployed in socio-economic contexts, regulatory sandboxes, post-market monitoring, and incident reporting are likely to play a particularly important role as instruments of risk mitigation. From a longer-term perspective, newly emerging, use-case-specific risks could be addressed by expanding the list of high-risk areas in the Annex to the AI Act. Nevertheless, given the autonomous interactions of agents with their environment, it will generally remain difficult to fully anticipate and identify all relevant risks, even within a single domain.

4.5 Towards value-chain neutrality

The previous heading show that the regulatory value chain embedded in the AI Act might be outdated already and is in any event unlikely to be sustainable.

It is now common for AI System providers to use GPAI Models as inputs in the development of user-facing applications. While such models may be developed or fine-tuned internally by firms, often drawing on open-source models, they are frequently procured as services from third-party providers and accessed via APIs. This practice gives rise to data-driven algorithmic supply chains, in which multiple interconnected actors contribute towards the production, deployment, use and functionality of AI services (see Figure 2 for a stylized and simplified view).[122] These supply chain interdependencies imply that risks originating at one layer can propagate across other actors' systems, yet they simultaneously allow for intervention and mitigation strategies to be implemented at various stages of the supply chain. As illustrated by the arrows in Figure 2, this requires that information is shared among actors of the supply chain, both in the upstream direction (primarily for data collection and the identification of risks and incidents) and downstream direction (primarily for the dissemination of risk information and the implementation of corrective measures). However, the sharing of information in this way must also be protective of trade secrets, since many of these players are vertically integrated,

[122] Cobbe, J., Veale, M., & Singh, J. (2023). Understanding accountability in algorithmic supply chains. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (pp. 1186-1197).

and may have a supplier/customer relationship at one point in the value chain while competing at another point in the value chain.

At the same time, supply chains are often transient and dynamic, rendering the independencies unstable, even though the bilateral relations between particular actors may remain relatively constant.[123] In the extreme case, supply chains may only materialise for a specific request to the AI service, as calls to specific upstream functionalities may only be instantiated after a specific user request. Even for more stable data flows and relationships between actors, agile development processes may lead to frequent changes of the involved interdependencies. Nonetheless, AI supply chains may also exhibit significant degrees of integration (especially around core inputs, such as GPAI Models), meaning that major providers are active at various upstream and downstream levels.

The AI Act was not designed to handle such diversity, volatility and fluidity of value chains in practice.

For instance, the requirement to conclude written contracts with all third-party input suppliers of High-risk AI Systems (unless the inputs are freely available as open source) could prove burdensome in practice, given the complexities of algorithmic supply chains outlined above. AI Systems (as most other current software and digital systems) typically rely on numerous dependencies and third-party components, making it difficult for providers of High-risk AI Systems to establish explicit agreements with every input supplier. Although the AI Office's expected provision of model terms for voluntary contracts may reduce the transaction costs associated with such agreements, smaller providers may still face difficulties in securing the consent of all third parties. Consequently, this requirement risks raising costs and creating entry barriers for providers operating in high-risk AI domains. This is particularly problematic in relation to inputs that fall outside the scope of the open-source exemption yet contribute little to facilitating compliance with the AI Act's requirements.

Beyond that, it is interesting to explore what technological neutrality would imply in the vertical dimension, where we derived the idea of "value-chain neutrality", since value chains are conditioned by technology (next to other considerations like commercial factors and business strategies). A technologically neutral regulation would then seek not to discriminate between various technological-driven value chain models, to be sustainable in the face of changes in value chains and not to pre-empt the choice of value chain model. That is, unless there was a higher-order justification to break from technological neutrality. What could that justification be? In the AI Act, what goal is served by embedding a value chain model in the AI Act?

Leaving aside the natural propensity of lawmakers to create definitional architectures, the most likely answer is that roles in the value chain must be specified to ensure clarity as regards accountability. To each role (GPAI Model provider, AI System provider, deployer, etc.) corresponds a specific set of regulatory obligations. The thrust of the AI Act is to vest primary accountability in one role and impose ancillary requirements (collaboration, information-sharing, reporting, etc.) on the others. The entity occupying that primary role presumably has strong incentives to bear on other links in the value chain to collaborate in complying with the regulatory regime.

Even then, the advantages of specifying roles and assigning accountability in a regulatory value chain must be balanced with the distortions this could impose in a context of diversity and fluidity of value

[123] Cobbe et al., 2023.

chain models. In particular, linking accountability to one role in the value chain can also be counterproductive in terms of clarity and predictability if parties in the value chain are uncertain as to which 'role' (in the regulatory scheme) they fall into because the actual value chain is not operating along the same lines as the regulatory value chain defined in legislation.

Furthermore, while for pre-market compliance it might be advantageous to specify regulatory roles clearly, the same does not go for post-market surveillance or for liability. At the pre-market compliance stage, clarity as to accountability helps market parties proceed to the market efficiently: one party bears responsibility for ensuring compliance, the others serve ancillary roles. This avoids a situation where many different parties carry out separate and parallel compliance exercises.

By definition, compliance is carried out against the backdrop of a set of expectations as to risk, which are largely fixed *ex ante*, as the AI Act itself exemplifies. Once a product is put in circulation, however, there is no guarantee that reality will unfold precisely as anticipated. Harm may occur in ways that were unforeseen or even unforeseeable. As mentioned above, this will quite likely be the case with AI Systems and GPAI Models: despite all the efforts going into compliance, post-market surveillance is likely to play a large role, as unexpected scenarios result in serious harm. In such cases, cooperation in dealing with actual and present harm might be more important than assigning clear regulatory roles to parties along the value chain.

Going one step further, if and when liability issues would arise, it would be unfortunate for the victims, and inefficient for the compensation system, if the roles specified in the regulatory value chain exacerbated defensiveness and blame-shifting attitudes amongst market parties. In the end, the costs and benefits of enshrining a value chain in regulation must be seen across the entire timeline of regulation: pre-market compliance, post-market surveillance and even liability. Depending on where the balance falls, there is also a case against embedding a regulatory value chain in a legislation such as the AI Act at all, aiming for more flexibility and vertical technological neutrality in such a fast-moving sector, where risks are often unpredictable.[124]

Ideally, a concept of "accountability across the value chain" could be developed along the following lines. The aim would be to avoid counterproductive finger-pointing and blame-shifting amongst parties in the value chain and foster a culture of cooperation and information sharing across the value chain, especially in the post-market phase. A guiding principle could be that every actor in the value chain would be responsible to ensure that itself and every other actor has the requisite information in hand to be held accountable, e.g. to comply with regulatory requirements if requested to, especially to contribute to solving post-market incidents quickly. In such an environment, public authorities and end-users, who might not always be cognizant of the various value chain models present in the ecosystem, could expect accountability from any link in the value chain.

To some extent, the post-market monitoring provisions of the AI Act already reflect the above. They are less clear-cut than the pre-market compliance sections. At least as far as AI Systems are concerned, the AI Act does not further list the specific elements of information sharing beyond general principles. This allows for more flexibility and adaptability to individual circumstances and should thus be seen as a positive element. Leaving value chain roles underspecified, as regards post-market surveillance, may likewise be beneficial in allowing the rules to adjust to the needs of specific contexts and application domains. To support effective implementation, the AI Office and the Commission may

[124] In the Knightian sense, they are really uncertainties rather than risks.

provide complementary guidance on the information expected to be shared across the value chain, potentially offering greater detail for individual risk domains. This could be further supported through the development of reference processes for information sharing. Drawing on the Code of Practice for GPAI Models, example methods for post-market monitoring could also be suggested for providers of High-risk AI Systems. At present, the provisions do not specify the consequences if actors in the value chain fail to provide the required information to others. Greater clarity regarding the enforcement framework, together with the creation of mechanisms and institutions for dispute resolution in such cases, will therefore be important for effective implementation.

As mentioned above, as AI Systems increasingly permeate individuals' daily lives and become prevalent across economic sectors, risks and incidents are likely to arise also outside the domains classified as high-risk under the AI Act. Recent examples include reports where the use of AI Systems has been associated with severe incidents including suicides, misinformation, and cyberattacks.[125] Moreover, given the universal applicability of GPAI Models in combination with agentic AI Systems, it will become even more difficult to assess risks ex ante and to anticipate all potential use cases in which these systems may be employed.

While swift identification and mitigation of post-market risks and serious incidents require vertical information flows and collaboration along individual AI value chains as well as the reporting to authorities (see Section 3.4), broader information sharing between providers of AI Systems and GPAI Models could further improve risk mitigation. In particular, because AI Systems and GPAI Models often rely on similar techniques and technical architectures, developers may benefit from learning about incidents that have occurred in other systems and models, thereby anticipating potential risks in the development of their own systems and models. In addition, the risks related to AI, specifically GPAI, are often cross-sectoral, so that the analysis of incidents in different sectors can be beneficial.[126] Broader availability of incident information could also help deployers remain informed about vulnerabilities and risks associated with their AI Systems and highlight the need for necessary updates. At the same time, it is important to ensure that such transparency does not inadvertently enable the exploitation of these issues by malicious actors. To this end, disclosure processes could follow established cybersecurity practices, such as coordinated or responsible vulnerability disclosure, which allow the responsible parties sufficient time to address the issue.[127]Databases or repositories for collecting and sharing information about incidents and vulnerabilities are common in other sectors, both digital and physical.[128] In cybersecurity, the Common Vulnerabilities and Exposures (CVE) system

[125] 111 The Washington Post. (2025). Instagram's chatbot helped teen accounts plan suicide — and parents can't disable it. https://www.washingtonpost.com/technology/2025/08/28/meta-ai-chatbot-safety-teens/; Anthropic. (2025). Threat Intelligence Report: August 2025. https://www-cdn.anthropic.com/b2a76c6f6992465c09a6f2fce282f6c0cea8c200.pdf; The Guardian. (2025). Elon Musk's AI firm apologizes after chatbot Grok praises Hitler. https://www.theguardian.com/us-news/2025/jul/12/elon-musk-grok-antisemitic.

[126] Lupo, G. (2023). Risky artificial intelligence: The role of incidents in the path to AI regulation. Law, Technology and Humans, 5(1), 133-152.

[127] Weulen Kranenbarg, M., Holt, T. J., & van der Ham, J. (2018). Don't shoot the messenger! A criminological and computer science perspective on coordinated vulnerability disclosure. *Crime Science, 7*(1), 1-9; Walshe, T., & Simpson, A. C. (2022). Coordinated vulnerability disclosure programme effectiveness: Issues and recommendations. Computers & Security, 123, 102936; ENISA. Economics of Vulnerability Disclosure. https://www.enisa.europa.eu/publications/economics-of-vulnerability-disclosure.

[128] Agarwal, A., & Nene, M. J. (2024, July). Addressing AI risks in critical infrastructure: formalising the AI incident reporting process. In 2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (pp. 1-6). IEEE.

has been critical for the unique identification of security vulnerabilities and for dissemination of related information across sectors and supply chains.[129] In aviation, the Aviation Safety Information Analysis and Sharing (ASIAS) system, operated by the US Federal Aviation Administration, serves as a central hub for the exchange of safety information and supports databases containing reported accidents, important findings, and safety recommendations. Similarly, AI incident information sharing could be facilitated by establishing a common database or information hub. Such an institution could build on existing databases that currently rely largely on user-contributed reports.[130] By establishing standardised schemas for describing AI incidents, ideally aligned with the templates to be developed by the European Commission, this database could further facilitate analysis and responses.[131] Standardizing the structure of incident reporting based on commonly accepted schema and taxonomy could be particularly valuable, as existing databases have been found to lack such standardisation as well as granularity required for consistent data collection and analysis, impeding effective incident management.[132] Additional measures may be necessary to mitigate structural ambiguities inherent to incident reporting obligations (such as handling multiplicity of incidents).[133] While its primary purpose would be to support the information exchange with authorities and the collaboration between AI providers, the database could also provide the public with some access, albeit at a less granular level.[134] In addition, the collected data could be made accessible for research to further promote AI safety and the development of risk mitigation measures.

As information sharing among potential or actual competitors may raise concerns about anti-competitive practices, clear guidelines are needed to delineate what information can be shared and possibly further discussed in the context of risk mitigation. In return, AI operators that comply with these guidelines should be provided with legal certainty that such collaboration and information exchange will not be deemed to constitute anti-competitive conduct.

[129] https://www.cve.org.

[130] Responsible AI Collaborative. (n.d.). AI Incident Database. https://incidentdatabase.ai/about/; AI, Algorithmic and Automation Incidents and Controversies (AIAAIC) Repository. (n.d.). About AIAAIC. https://www.aiaaic.org/about-aiaaic.

[131] See also Croxton, J., Robusto, D., Thallam, S. & Calidas, D. (20249: Message Incoming. Establish an AI Incident Reporting System. https://fas.org/publication/establishing-an-ai-incident-reporting-system/;Uba, C. (2025). Towards an AI Incident & Response Framework: Conceptualizing Cause, Locus, and Impact of AI Incidents. *AMCIS 2025 Proceedings.*

[132] Agarwal, A., & Nene, M. J. (2024, December). Standardised schema and taxonomy for AI incident databases in critical digital infrastructure. In 2024 IEEE Pune Section International Conference (PuneCon) (pp. 1-6). IEEE.

[133] Paeth, K., Atherton, D., Pittaras, N., Frase, H., & McGregor, S. (2025, April). Lessons for editors of AI incidents from the AI incident database. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 28, pp. 28946-28953).

[134] Croxton et al., 2024.

5. Conclusions and recommendations

In this issue paper, we have assessed the extent to which the EU AI Act aligns with the principle of technological neutrality, a core element of better regulation. In particular, our analysis has focused on how well the Act can adapt to changes in the 'AI value chain', both those already visible in the market today and those likely to emerge in the future. While the AI Act incorporates many elements of technologically neutral, principles-based regulation, our discussion also points to a number of concerns about its technological neutrality, including:

- The allocation of responsibilities across the AI value chain that is neutral towards the diversity and fluidity of value chain models, especially considering the emergence of new actors along the value chain (such as model platform operators and other intermediaries), the growing interdependence between different AI models and services, and the integration of different AI models and services into single packages and products;
- Significantly varying degrees in the specification of value chain roles between (GPAI) model providers and AI system providers, especially in the context of post-market monitoring and the sharing of information about serious incidents;
- A tension between the ex-ante classification and tiering of (systemic) risks and the ex-post materialisation of harm, as in the case of serious incidents;
- The rapid development and adoption of new machine learning methods and approaches that enable smaller models (in terms of the number of model parameters or training compute) to attain high levels of generality and capability, thereby challenging the adequacy of unidimensional quantitative proxies and thresholds for delineating systemic risks and responsibilities for risk mitigation.
- The potentially significant and ongoing role of open-source AI models in the value chain, particularly in the context of the aforementioned technical developments;
- The growing role of agentic AI systems (which operate autonomously across changing tasks and contexts) and new players in the AI value chain (such as those that specialise in fine-tuning general-purpose models), which raises questions around the lack of a clear distinction between general-purpose and purpose-specific AI services, and about whether every actor in the AI value chain has a clear place in the regulatory framework.

To address these concerns and to accommodate both the fluidity and increasing complexity of emerging AI value chains, the paper proposes the concept of "value-chain neutrality", derived as a vertical dimension of technological neutrality. In general, AI regulation should not favour one technical or organisational value-chain design over another, unless clearly justified by higher-order goals like accountability. Pre-market, it can still be useful to assign primary responsibility to one actor (e.g. the AI System provider) to avoid duplicated compliance. But post-market, where unforeseen harms are likely and responsibilities are more diffuse, rigid role definitions become counterproductive. Instead, the law should foster "accountability across the value chain": every actor should have duties to share the information needed for others to meet their obligations and to resolve incidents quickly.

On this basis, we derive the following five main recommendations:

Recommendation 1: Law-makers should establish general principles for cooperation across the AI value chain to support effective risk identification and mitigation (especially for the post-market phase) rather than fully prescribing value-chain structures and roles, which are prone to being overly rigid and quickly becoming outdated. To support effective implementation, the AI Office could then issue complementary guidance on the information expected to be shared across the value chain.

The analysis above shows that the AI value chain has already evolved significantly beyond the linear structure assumed in the AI Act's regulatory framework. New distribution channels, cloud platforms, and model platforms have introduced intermediaries that the Act barely contemplates. As a result, these new actors do not fit the regulatory value chain, and their duties regarding cooperation within the value chain, such as in the context of serious-incident reporting and information-sharing, are not explicitly specified. Even for actors explicitly covered by the Act, the current approach of enumerating obligations for each type of value-chain participant can lead to gaps (see Section 3.4). For instance, the AI Act does not specify any obligation for GPAI model providers that fall below the systemic-risk threshold and downstream system providers or deployers to share information on serious incidents, in either direction. While such omissions may reflect an intention to reduce the regulatory burden and to allocate responsibilities to the parties best positioned to access relevant information, they nonetheless create potential gaps when serious incidents arise in systems based on non-systemic GPAI models or in non-high-risk use cases.

Rather than trying to keep pace with ongoing developments in the AI value chain by exhaustively specifying its structure and prescribing detailed obligations, a principle-based approach is better suited to fostering effective cooperation for risk identification and mitigation, particularly in the post-market phase. Such an approach allows for flexibility to accommodate context-specific circumstances and adaptability to evolving technologies and value-creation processes. It is thus also likely to lower transaction costs for operators compared to a highly prescriptive regulatory framework governing cooperation across the AI value chain.

In this context, a guiding principle for the proposed concept of "accountability across the value chain" (see Section 4.5) could be that every actor in the value chain would be responsible to ensure that itself and every other actor has the requisite information in hand to be held accountable, e.g. to comply with regulatory requirements if requested to, especially to contribute to solving post-market incidents quickly. In such an environment, public authorities and end-users, who might not always be cognizant of the various value chain models present in the ecosystem, could expect accountability from any link in the value chain. To support effective implementation, the AI Office and the Commission could issue complementary guidance on the information expected to be shared across the value chain, potentially offering greater detail for individual risk domains. This could be further supported through the development of reference processes for information sharing.

Nonetheless, the proposed concept is not intended to be absolute. The AI Act's regulatory framework may still prescribe specific obligations for certain value-chain roles in selected cases, such as where actors exhibit unique characteristics, should bear specific responsibilities, or are intended to serve as central anchors for certain cooperation duties.

Recommendation 2: As their relevance increases, the AI Office should provide guidance to more fully integrate General-purpose AI Systems into the AI Act framework. As a first step, regulators should monitor how effectively contractual arrangements, market incentives and co-regulatory

approaches, alongside existing obligations, mitigate risks arising from such systems. Further regulatory guidance should then build on observed best practices and identified market failures.

As illustrated by the discussion of agentic AI, future AI Systems will increasingly perform general-purpose tasks rather than narrowly defined functions (see Section 4.4). This trend is already visible in current AI Systems, particularly those that perform functions (such as image recognition or text processing) that can be used across diverse application contexts, and which are therefore often embedded as inputs in downstream AI Systems. With the quickly increasing adoption of GPAI Models and General-purpose AI Systems, risks can also be expected to materialise more frequently outside the high-risk domains defined in the AI Act.

One potential response to this development would be to designate additional high-risk domains, an option already foreseen under the AI Act. However, this would extend the most stringent obligation, originally intended for a narrow subset of AI Systems, to a much broader range of applications. In the extreme, this could result in most AI Systems being classified as high-risk, thereby undermining the tiered risk framework designed to ensure proportionality and avoid stifling innovation. The designation of new high-risk domains should therefore remain a selective measure.

A more appropriate approach is to develop the existing but currently under-specified concept of General-purpose AI Systems within the AI Act. Given that both the underlying technologies and associated value-creation models are still in an early or quickly evolving phase, the principle of technological neutrality cautions against imposing overly prescriptive ex-ante obligations at this stage. Rather, as a first step, the AI Office and regulatory authorities should monitor how effectively contractual agreements and market incentives, in addition to existing obligations for GPAI Model providers, can foster cooperation across the AI value chain to prevent and mitigate risks of General-purpose AI Systems. This approach would preserve flexibility for experimentation, allow adaptation to sector-specific circumstances, and help avoid prematurely locking industries into rigid regulatory models.

Co-regulatory instruments, such as regulatory sandboxes, could further support this adaptive approach and are particularly suited to addressing the high uncertainty surrounding General-purpose AI Systems and GPAI Models. As best practices emerge or market failures become apparent, regulators can then intervene where necessary or facilitate compliance by issuing more detailed and prescriptive guidance.

Recommendation 3: Industry players should develop new institutions and mechanisms for broader information sharing on incidents and risks in the post-market phase.

As the roles of actors along the AI value chain become more fluid and the categorisation of operators into discrete roles (such as in the context of General-purpose AI Systems) becomes increasingly challenging, broader information-sharing on incidents and risks in the post-market phase could play a major role in facilitating effective risk identification and timely mitigation. This need is further reinforced by the expectation that risks and incidents will also arise outside the domains the AI Act classifies as high-risk, as AI Systems become increasingly pervasive in individuals' daily lives and across economic sectors (see Section 4.4).

Because AI Systems and GPAI Models often rely on similar techniques and technical architectures, developers can benefit from learning about incidents that have occurred in other systems and models, thereby anticipating potential risks in the development of their own systems and models. In addition, the risks related to AI, specifically general-purpose AI, are often cross-sectoral, so that the analysis of incidents in one sector can be beneficial to avoid incidents in other sectors. Broader availability of incident information can also help deployers remain informed about vulnerabilities and risks associated with their AI Systems and highlight the need for necessary updates. At the same time, it is important to ensure that such transparency does not inadvertently enable the exploitation of these issues by malicious actors.

Industry-driven self-regulatory initiatives could promote such information-sharing by establishing a common database or information hub for AI-related incidents, following models that exist in other sectors such as aviation or cybersecurity. Such an institution could build on existing databases, which currently rely largely on user-contributed reports. Promoting standardised reporting schemas, while implementing safeguards for trade secrets, security, and competition law (see the discussion in Section 4.5), could further promote the effectiveness of such an institution and its associated sharing mechanisms.

Beyond such a centralised database, intermediaries and platforms along AI value chains (see Section 4.3) could play an important role as potential distributors of information about serious incidents, risks, or system and model updates, building on their existing relationships with relevant developers, customers and end-users. In the absence of a self-regulatory initiative emerging, the Commission could consider interventions to deliver such a solution.

Recommendation 4: The AI Office should clarify the responsibilities of suppliers of inputs to High-risk AI Systems so as to avoid chilling effects on the provision of important inputs, while allowing context-specific agreements and solutions to develop.

Article 26 of the AI Act establishes general information-sharing and assistance obligations for a broad set of input suppliers to High-risk AI systems, exempting only open-source inputs other than GPAI Models. Although the Act appears to leave substantial flexibility for input suppliers and system providers to determine the precise terms of these obligations, implementing them may prove challenging in practice. Providers of High-risk AI Systems often rely on numerous inputs that were not originally developed for high-risk contexts. In such cases, even the requirement to conclude written agreements with all relevant input suppliers may be difficult to fulfil in today's fluid, dynamic, and sometimes algorithmically configured supply chains. This difficulty arises even when the inputs in question are not critical components for the high-risk task performed by the system.

Even where written agreements are technically feasible, input suppliers may have limited incentives to enter into them if they fear that doing so could expose them to additional liability or compliance burdens. Conversely, it remains unclear whether input suppliers can effectively and legally exclude the use of their products in High-risk AI Systems through licensing terms. As a result, the current provisions, and the uncertainty surrounding them, may create barriers to the development and diffusion of AI Systems in high-risk domains, even when many input components do not materially affect system safety. This is problematic, as high-risk domains are also areas where AI Systems have the potential to generate significant economic and societal benefits.

Recommendation 5: The AI Office should reconsider the role of computing thresholds as a proxy for classifying GPAI Models and GPAI Models with systemic risk in the light of current technical developments.

The use of quantitative training compute (FLOP) thresholds in the AI Act to classify GPAI Models and GPAI Models with systemic risk reflects an understandable desire to provide legal certainty and consistency in delineating critical roles in the AI value chain. However, as outlined in the discussion above, these thresholds represent a significant departure from technological neutrality and risk becoming rapidly outdated. Emerging techniques such as model distillation, fine-tuning and the deployment of specialised smaller models can produce systems with capabilities and risk profiles comparable to, or derived from, much larger models, while falling below the relevant compute thresholds. In these cases, compute-based proxies may systematically under- or over-estimate risk and possibly distort incentives for model development.

While no single alternative metric is without shortcomings, other indicators, such as the number of users, may provide a more robust and technologically neutral proxy, particularly for systemic risks where the breadth of exposure is itself a key concern. Over time, as the AI Office and GPAI Model providers accumulate data and experience and given that the number of models potentially creating systemic risk is likely to remain manageable, the efficiency value of simple proxies can be expected to decline, and more sophisticated classification approaches, especially for the designation of systemic risks, should become feasible. Such approaches should then rely on more direct assessments and evaluations of the systemic risks associated with GPAI Models.

cerre

Avenue Louise 475 (box 10)
1050 Brussels, Belgium
+32 2 230 83 60
info@cerre.eu
www.cerre.eu

Centre on Regulation in Europe (CERRE)
CERRE Think Tank
CERRE Think Tank