12

The European Union's AI Act

Beyond Motherhood and Apple Pie?

Nathalie A. Smuha and Karen Yeung[*]

12.1 INTRODUCTION

In spring 2024, the European Union formally adopted the "AI Act,"[1] purporting to create a comprehensive EU legal regime to regulate AI systems across sectors. In so doing, it signaled its commitment to the protection of core EU values against AI's adverse effects, to maintain a harmonized single market for AI in Europe and to benefit from a first mover advantage (the so-called "Brussels effect")[2] to establish itself as a leading global standard-setter for AI regulation. The AI Act reflects the EU's recognition that, left to its own devices, the market alone cannot protect the fundamental values upon which the European project is founded from unregulated AI applications.[3] Will the AI Act's implementation succeed in translating its noble aspirations into meaningful and effective protection of people whose everyday lives are already directly affected by these increasingly powerful systems? In this chapter, we critically examine the proposed conceptual vehicles and regulatory architecture upon which the AI Act relies to argue that there are good reasons for skepticism.

[*] Smuha primarily contributed to Sections 12.2 and 12.3, (drawing on Nathalie A. Smuha, *Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law* (Cambridge University Press, 2025, Chapter 5.4), while Yeung contributed primarily to Section 12.4 (drawing extensively on a keynote speech delivered on September 12, 2022, ADM+S Centre Symposium, *Automated Societies*, RMIT, Melbourne, Australia. A recording is available at https://podcasters.spotify.com/pod/show/adms-centre/episodes/2022-ADMS-Symposium-Keynote-by-Professor-Karen-Yeung-e1nmp1r/a-a8gu1ph (accessed August 2, 2024)).

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689, July 12, 2024.

[2] Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (Oxford University Press, 2020). See in this regard also Nathalie A. Smuha, "From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence," (2021) *Law, Innovation and Technology*, 13(1): 57–84.

[3] See Karen Yeung, Andrew Howes, and Ganna Pogrebna, "AI governance by human rights–centered design, deliberation, and oversight: An end to ethics washing," in Markus D. Dubber, Frank Pasquale, and Sunit Das (eds), *The Oxford Handbook of Ethics of AI* (Oxford University Press, 2020), pp. 76–106.

Despite its laudable intentions, the Act may deliver far less than it promises in terms of safeguarding fundamental rights, democracy, and the rule of law. Although the Act appears to provide meaningful safeguards, many of its key operative provisions delegate critical regulatory tasks largely to AI providers themselves without adequate oversight or effective mechanisms for redress.

We begin in Section 12.2 with a brief history of the AI Act, including the influential documents that preceded and inspired it. Section 12.3 outlines the Act's core features, including its scope, its "risk-based" regulatory approach, and the corollary classification of AI systems into risk-categories. In Section 12.4, we critically assess the AI Act's enforcement architecture, including the role played by standardization organizations, before concluding in Section 12.5.

12.2 A BRIEF HISTORY OF THE AI ACT

Today, AI routinely attracts hyperbolic claims about its power and importance, with one EU institution even likening it to a "*fifth element after air, earth, water and fire*."[4] Although AI is not new,[5] its capabilities have radically improved in recent years, enhancing its potential to effect major societal transformation. For many years, regulators and policymakers largely regarded the technology as either wholly beneficial or at least benign. However, in 2015, the so-called "Tech Lash" marked a change in tone, as public anxiety about AI's potential adverse impacts grew.[6] The Cambridge Analytica scandal, involving the alleged manipulation of voters via political microtargeting, with troubling implications for democracy, was particularly important in galvanizing these concerns.[7] From then on, policy initiatives within the EU and elsewhere began to take a "harder" shape: eschewing reliance on industry self-regulation in the form of non-binding "ethics codes" and culminating in the EU's "legal turn," marked by the passage of the AI Act. To understand the Act, it is helpful to briefly trace its historical origins.

12.2.1 *The European AI Strategy*

The European Commission published a European strategy for AI in 2018, setting in train Europe's AI policy[8] to promote and increase AI investment and uptake across

[4] Statement by the European Parliament's Special Committee on Artificial Intelligence in a Digital Age (AIDA), "Draft report on artificial intelligence in a digital age" (European Parliament, 2021) (2020/2266(INI)) 9.

[5] See in this regard also Chapter 1 of this book by Wannes Meert, Tinne De Laet, and Luc De Raedt.

[6] The first use of this term is ascribed to Adrian Wooldridge in his *The Economist* article titled "The coming tech-lash," November 2013.

[7] See, for example, Jim Isaak and Mina J Hanna, "User data privacy: Facebook, Cambridge Analytica, and privacy protection" (2018) *Computer*, 51(8): 56-59.

[8] European Commission, Artificial Intelligence for Europe, COM (2018) 237 final, Brussels, April 25, 2018.

Europe in pursuit of its ambition to become a global AI powerhouse.[9] This strategy was formulated against a larger geopolitical backdrop in which the US and China were widely regarded as frontrunners, battling it out for first place in the "AI race" with Europe lagging significantly behind. Yet the growing Tech-Lash made it politically untenable for European policymakers to ignore public concerns. How, then, could they help European firms compete more effectively on the global stage while assuaging growing concerns that more needed to be done to protect democracy and the broader public interest? The response was to turn a perceived weakness into an opportunity by making a virtue of its political ideals and creating a unique "brand" of AI: infused with "European values" – charting a "third way," distinct from both the Chinese state-driven approach and the US' laissez-faire approach to AI governance.[10]

At that time, the Commission resisted calls for the introduction of new laws. In particular, in 2018 the long-awaited General Data Protection Regulation (GDPR) finally took effect,[11] introducing more stringent legal requirements for collecting and processing personal data. Not only did EU policymakers believe these would guard against AI-generated risks, but it was also politically unacceptable to position this new legal measure as outdated even as it was just starting to bite. By then, the digital tech industry was seizing the initiative, attempting to assuage rising anxieties about AI's adverse impacts by voluntarily promulgating a wide range of "Ethical Codes of Conduct" proudly proclaiming they would uphold. This coincided with, and concurrently nurtured, a burgeoning academic interest by humanities and social science scholars in the social implications of AI, often proceeding under the broad rubric of "AI Ethics." In keeping with industry's stern warning that legal regulation would stifle innovation and push Europe even further behind, the Commission decided to convene a High-Level Expert Group on AI (AI HLEG) to develop a set of *harmonized* Ethics Guidelines based on European values that would serve as "best practice" in Europe, for which compliance was entirely voluntary.

12.2.2 *The High-Level Expert Group on AI*

This 52 member group was duly convened, to much fanfare, selected through open competition and comprised of approximately 50% industry representatives, with the remaining 50% from academia and civil society organizations.[12] Following a public

[9] Nathalie A. Smuha, "The EU approach to ethics guidelines for trustworthy artificial intelligence" (2019) *Computer Law Review International*, 20(4): 98.

[10] See also Anu Bradford, *Digital Empires: The Global Battle to Regulate Technology* (Oxford University Press, 2023).

[11] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, OJ L 119, May 4, 2016, pp. 1–88.

[12] Both the composition and the mandate of the AI HLEG was criticized, mostly due to the larger representation of industry, and the fact that the Commission tasked the group with drafting voluntary

consultation, the group published its Ethics Guidelines for Trustworthy AI in April 2019,[13] coining "Trustworthy AI" as its overarching objective.[14] The Guidelines' core consists of seven requirements that AI practitioners should take into account throughout an AI system's lifecycle: (1) *human agency and oversight* (including the need for a fundamental rights impact assessment); (2) *technical robustness and safety* (including resilience to attack and security mechanisms, general safety, as well as accuracy, reliability and reproducibility requirements); (3) *privacy and data governance* (including not only respect for privacy, but also ensuring the quality and integrity of training and testing data); (4) *transparency* (including traceability, explainability, and clear communication); (5) *diversity, nondiscrimination and fairness* (including the avoidance of unfair bias, considerations of accessibility and universal design, and stakeholder participation); (6) *societal and environmental well-being* (including sustainability and fostering the "environmental friendliness" of AI systems, and considering their impact on society and democracy); and finally (7) *accountability* (including auditability, minimization, and reporting of negative impact, trade-offs, and redress mechanisms).[15]

The group was also mandated to deliver Policy Recommendations which were published in June 2019,[16] oriented toward Member States and EU Institutions.[17]

guidelines rather than asking its input on new binding rules. Yeung was one of these members. Smuha served as the group's coordinator from its initial formation until July 2019.

[13] High-Level Expert Group on AI, "Ethics Guidelines for Trustworthy AI," Brussels, April 8, 2019. The Guidelines were endorsed by the Commission in a Communication that was published the same day, encouraging AI developers and deployers to implement them in their organization. See European Commission, Building Trust in Human-Centric Artificial Intelligence, COM (2019) 168 final, Brussels, April 8, 2019.

[14] Trustworthy AI was defined as: (1) lawful, or complying with all applicable laws and regulations; (2) ethical, or ensuring adherence to ethical principles and values; and (3) robust since, even with good intentions, AI systems can still lead to unintentional harm. The AI HLEG was however careful in stating that the Guidelines only offered guidance on complying with the two latter components (*ethical* and *robust* AI), indicating the need for the EU to take additional steps to ensure that AI systems were also *lawful*. See in this regard also Nathalie A. Smuha, Emma Ahmed-Rengers, Adam Harkens, Wenlong Li, James MacLaren, Riccardo Piselli, and Karen Yeung, "*How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act*," Social Science Research Network, 2021, https://ssrn.com/abstract=3899991.

[15] The Guidelines also included an assessment list to operationalize these requirements in practice, and a list of critical concerns raised by AI systems that should be carefully considered (including, for example, the use of AI systems to identify and track individuals, covert AI systems, AI-enabled citizen scoring, lethal autonomous weapons, and longer-term concerns, covering what is today often referred to as "existential risks").

[16] High-Level Expert Group on AI, 'Policy and Investment Recommendations for Trustworthy AI' (European Commission, June 26, 2019), https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

[17] In addition, the group was also mandated to support the Commission with outreach through the European AI Alliance, a multi-stakeholder online platform seeking broader input on Europe's AI policy. See European Commission, Call for Applications for the Selection of Members of the High-Level Expert Group on Artificial Intelligence, March 9, 2018, https://digital-strategy.ec.europa.eu/en/news/call-high-level-expert-group-artificial-intelligence.

While attracting considerably less attention than the Ethics Guidelines, the Recommendations called for the adoption of new legal safeguards, recommending "*a risk-based approach to AI policy-making*," taking into account "*both individual and societal risks*,"[18] to be complemented by "*a precautionary principle-based approach*" for "*AI applications that generate 'unacceptable' risks or pose threats of harm that are substantial*."[19] For the use of AI in the public sector, the group stated that adherence to the Guidelines should be mandatory.[20] For the private sector, the group asked the Commission to consider introducing obligations to conduct a "trustworthy AI" assessment (including a fundamental rights impact assessment) and stakeholder consultations; to comply with traceability, auditability, and ex-ante oversight requirements; and to ensure effective redress.[21] These Recommendations reflected a belief that nonbinding "ethics" guidelines were insufficient to ensure respect for fundamental rights, democracy, and the rule of law, and that legal reform was needed. Whether a catalyst or not, we will never know, for a few weeks later, the then President-elect of the Commission, Ursula von der Leyen, announced that she would "*put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence*."[22]

12.2.3 *The White Paper on AI*

In February 2020, the Commission issued a White Paper on AI,[23] setting out a blueprint for new legislation to regulate AI "*based on European values*,"[24] identifying several legal gaps that needed to be addressed. Although it sought to adopt a risk-based approach to regulate AI, it identified only two categories of AI systems: high-risk and not-high-risk, with solely the former being subjected to new obligations inspired by the Guidelines' seven requirements for Trustworthy AI. The AI HLEG's recommendation to protect fundamental rights as well as democracy and the rule of law were largely overlooked, and its suggestion to adopt a precautionary approach in relation to "unacceptable harm" was ignored altogether.

On enforcement, the White Paper remained rather vague. It did, however, suggest that high-risk systems should be subjected to a prior conformity assessment by providers of AI systems, analogous to existing EU conformity assessment procedures for products governed by the New Legislative Framework (discussed later).[25]

[18] Policy and Investment Recommendations for Trustworthy AI (n 16), 26.
[19] Ibid., 38.
[20] Ibid., 20.
[21] Ibid., 40.
[22] Ibid., 13.
[23] European Commission, White Paper on Artificial Intelligence – A European approach to excellence and trust, Brussels, February 19, 2020, COM (2020) 65 final.
[24] See also the Explanatory Memorandum of the White Paper.
[25] The White Paper provides the examples of Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing

In this way, AI systems were to be regulated in a similar fashion to other stand-alone products including toys, measuring instruments, radio equipment, low-voltage electrical equipment, medical devices, and fertilizers rather than embedded within a complex and inherently socio-technical system that may be infrastructural in nature. Accordingly, the basic thrust of the proposal appeared animated primarily by a light-touch market-based orientation aimed at establishing a harmonized and competitive European AI market in which the protection of fundamental rights, democracy, and the rule of law were secondary concerns.

12.2.4 *The Proposal for an AI Act*

Despite extensive criticism, this approach formed the foundation of the Commission's subsequent proposal for an AI Act published in April 2021.[26] Building on the White Paper, it adopted a "horizontal" approach, regulating "AI systems" in general rather than pursuing a sector-specific approach. The risk-categorization of AI systems was more refined (unacceptable risk, high risk, medium risk, and low risk), although criticisms persisted given that various highly problematic applications were omitted from the list of "high-risk" and "unacceptable" systems, and with unwarranted exceptions.[27] The conformity (self)assessment scheme was retained, firmly entrenching a product-safety approach to AI regulation, yet failing to confer any rights whatsoever for those subjected to AI systems; it only included obligations imposed on AI providers and (to a lesser extent) deployers.[28]

In December 2022, the Council of the European Union adopted its "general approach" on the Commission's proposal.[29] It sought to limit the regulation's scope by narrowing the definition of AI and introducing more exceptions (for example for national security and research), sought stronger EU coordination for the Act's enforcement; and proposed that AI systems listed as "high-risk" systems would not be automatically subjected to the Act's requirements. Instead, providers could self-assess whether their system is *truly* high-risk based on a number of criteria – thereby further diluting the already limited protection the proposal afforded. Finally, the Council took into account the popularization of Large Language Models (LLMs) and generative AI applications such as ChatGPT, which at that time were drawing

Council Decision 93/465/EEC, and to Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA and on information and communications technology cybersecurity certification (the Cybersecurity Act).

[26] Proposal for a Regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), COM (2021) 206 final, Brussels, April 21, 2021.

[27] See also Smuha et al. (n 14), 28.

[28] See ibid, 50.

[29] Council of the European Union, General Approach, 2021/0106(COD) Brussels, 25 November 2022 (adopted December 6, 2022).

considerable public and political attention, and included modest provisions on General-Purpose AI models (GPAI).[30]

By the time the European Parliament formulated its own negotiating position in June 2023, generative AI was booming and called for more demanding restrictions. Additional requirements for the GPAI models that underpin generative AI were thus introduced, including risk-assessments and transparency obligations.[31] Contrary to the Council, the Parliament sought to widen some of the risk-categories; restore a broader definition of AI; strengthen transparency measures; introduce remedies for those subjected to AI systems; include stakeholder participation; and introduce mandatory fundamental rights impact assessments for high-risk systems. Yet it retained the Council's proposal to allow AI providers to self-assess whether their "high-risk" system could be excluded from that category, and hence from the legal duties that would otherwise apply.[32] It also sprinkled the Act with references to the "rule of law" and "democracy," yet these were little more than rhetorical flourishes given that it retained the underlying foundations of the original proposal's market-oriented product-safety approach.

12.3 SUBSTANTIVE FEATURES OF THE AI ACT

The adoption of the AI Act in spring 2024 marked the culmination of a series of initiatives that reflected significant policy choices which determined its form, content and contours. We now provide an overview of the Act's core features, which – for better or for worse – will shape the future of AI systems in Europe.

12.3.1 *Scope*

The AI Act aims to harmonize Member States' national legislation, to eliminate potential obstacles to trade on the internal AI market, and to protect citizens and society against AI's adverse effects, in that order of priority. Its main legal basis is Article 114 of the Treaty of the Functioning of the European Union (TFEU), which enables the establishment and functioning of the internal market. The inherent single-market orientation of this article limits the Act's scope and justification.[33] For this

[30] Essentially, it provided that GPAI systems used for high-risk purposes should be treated as such. However, instead of directly applying the high-risk requirements to such systems, the Council proposed that the Commission should adopt an implementing act to specify how they should be applied, based on a consultation and detailed impact assessment and taking into account their specific characteristics.

[31] European Parliament, Amendments adopted by the European Parliament on 14 June 2023 on the proposal for an Artificial Intelligence Act, COM (2021)0206 – C9-0146/2021 – 2021/0106(COD), Amendment 168.

[32] See also *infra* (n 61).

[33] See also Stephen Weatherill, "The limits of legislative harmonization ten years after tobacco advertising: How the court's case law has become a 'drafting guide'" (2011) *German Law Journal*, 12(3): 827–864.

reason, certain provisions on the use of AI-enabled biometric data processing by law enforcement are also based on Article 16 TFEU, which provides a legal basis to regulate matters related to the right to data protection.[34] Whether these legal bases are sufficient to regulate AI practices within the *public* sector or to achieve nonmarket-related aims remains uncertain, and could render the Act vulnerable to (partial) challenges for annulment on competence-related grounds.[35] In terms of scope, the regulation applies to providers who place on the market or put into service AI systems (or general purpose AI models) in the EU, regardless of where they are established; deployers of AI systems that have their place of establishment or location in the EU; and providers and deployers of AI systems that are established or located outside the EU, while the output produced by their AI system is used in the EU.[36]

The definition of AI for the purpose of the regulation has been a significant battleground,[37] with every EU institution proposing different definitions, each attracting criticism. Ultimately, the Commission's initial proposal to combine a broad AI definition in the regulation's main text with an amendable Annex that exhaustively enumerates the AI techniques covered by the Act was rejected. Instead, the legislators opted for a definition of AI which models that of the OECD, to promote international alignment: "a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."[38]

AI systems used exclusively for military or defense purposes are excluded from the Act, as are systems used for "nonprofessional" purposes. So too are AI systems "solely" used for research and innovation, which leaves open a substantive gap in protection given the many problematic research projects that can adversely affect individuals yet do not fall within the remit of university ethics committees. The AI Act also foresees that Member States' competences in national security remain untouched, thus risking very weak protection of individuals in one of the potentially

[34] See Recital 3 of the AI Act.

[35] See in this regard also Nathalie A. Smuha, "The paramountcy of data protection law in the age of AI (Acts)," in Brendan Van Alsenoy, Julia Hodder, Fenneke Buskermolen, Miriam Čakurdová, Ilektra Makraki and Estelle Burgot (eds), *Twenty Years of Data Protection. What Next? – EDPS 20th Anniversary*, Luxembourg (2024), Publications Office of the European Union, 226–39.

[36] See in more details Article 2(1) of the AI Act.

[37] For a discussion of the importance of AI definitions, see also Bilel Benbouzid, Yannick Meneceur and Nathalie A. Smuha, "Four shades of AI regulation. A cartography of normative and definitional conflicts" (2022) *Réseaux*, 232–33(2–3), 29–64.

[38] Article 3(1) of the AI Act. The definition's emphasis on the system making inferences seems to exclude more traditional or rule-based AI systems from its scope, despite their significant potential for harm. Ultimately, it will be up to the courts to decide how this definition must be interpreted in case of a dispute.

most intrusive areas for which AI might be used.[39] Finally, the legislators also included certain exemptions for open-source AI models and systems,[40] and derogations for microenterprises.[41]

12.3.2 *A Risk-based Approach*

The AI Act adopts what the Commission describes as a "risk-based" approach: AI systems and/or practices are classified into a series of graded "tiers," with proportionately more demanding legal obligations that vary in accordance with the EU's perceptions of the severity of the risks they pose.[42] "Risks" are defined rather narrowly in terms of risks to "health, safety or fundamental rights." The Act's final risk categorization consists of five tiers: (1) systems that pose an "unacceptable" risk are prohibited; (2) systems deemed to pose a "high risk" are subjected to requirements akin to those listed in the Ethics Guidelines; (3) GPAI models are subjected to obligations that primarily focus on transparency, intellectual property protection, and the mitigation of "systemic risks"; (4) systems posing a limited risk must meet specified transparency requirements; and (5) systems that are not considered as posing significant risks do not attract new legal requirements.

12.3.2.1 Prohibited Practices

Article 5 of the AI Act prohibits several "AI practices," reflecting a view that they pose an unacceptable risk. These include the use of AI to manipulate human behavior in order to circumvent a person's free will[43] and to exploit the vulnerability of natural persons in light of their age, disability, or their social or economic situation.[44] It also includes the use of AI systems to make criminal risk assessments and predictions of natural

[39] More generally, yet less unusual, the legislator also carved out from the AI Act all areas that fall outside the scope of EU law.

[40] Article 2 of the AI Act provides that "this Regulation does not apply to AI systems released under free and open-source licences, unless they are placed on the market or put into service as high-risk AI systems or as an AI system that falls under Article 5 or 50" (covering respectively prohibited AI practices and systems requiring additional transparency measures). Moreover, Article 53 of the AI Act excludes providers of AI models that are released under a free and open-source licence from certain transparency requirements if the license "allows for the access, usage, modification, and distribution of the model" and if certain information (about the parameters including the weights, model architecture, and model usage) is made publicly available. The exclusion does not apply to general-purpose AI models with "systemic risks" though, which shall be discussed further below.

[41] For instance, Article 63 of the AI Act states that microenterprises can comply with certain elements of the quality management system required by Article 17 in "a simplified manner," for which "the Commission shall develop guidelines."

[42] See in this regard Karen Yeung and Sofia Ranchordas, *An Introduction to Law and Regulation*, 2nd ed. (Cambridge University Press, 2025), especially Chapter 9, Section 9.9.2.

[43] Article 5(1)(a) of the AI Act.

[44] Article 5(1)(b) of the AI Act.

persons without human involvement,[45] or to evaluate or classify people based on their social behavior or personal characteristics (social scoring), though only if it leads to detrimental or unfavorable treatment in social contexts that are either unrelated to the contexts in which the data was originally collected, or that is unjustified or disproportionate.[46] Also prohibited is the use of emotion recognition in the workplace and educational institutions,[47] thus permitting their use in other domains despite their deeply problematic nature.[48] The untargeted scraping of facial images from the internet or from CCTV footage to create facial recognition databases is likewise prohibited.[49] Furthermore, biometric categorization is not legally permissible to infer sensitive characteristics, such as political, religious, or philosophical beliefs, sexual orientation or race.[50]

Whether to prohibit the use of real-time remote biometric identification by law enforcement in public places was a lightning-rod for controversy. It was prohibited in the Commission's original proposal, but subject to three exceptions. The Parliament sought to make the prohibition unconditional, yet the exceptions were reinstated during the trilogue. The AI Act therefore allows law enforcement to use live facial recognition in public places, but only if a number of conditions are met: prior authorization must be obtained from a judicial authority or an independent administrative authority; and it is used either to conduct a targeted search of victims, to prevent a specific and imminent (terrorist) threat, or to localize or identify a person who is convicted or (even merely) suspected of having committed a specified serious crime.[51] These exceptions have been heavily criticized, despite the Act's safeguards. In particular, they pave the way for Member States to install and equip public places with facial recognition cameras which can then be configured for the purposes of remote biometric identification if the exceptional circumstances are met, thus expanding the possibility of function creep and the abuse of law enforcement authority.

12.3.2.2 High-Risk Systems

The Act identifies two categories of high-risk AI systems: (1) those that are (safety components of) products that are already subject to an existing ex ante conformity assessment (in light of exhaustively listed EU harmonizing legislation on health and safety in Annex I, for example, for toys, aviation, cars, medical devices or lifts) and (2) stand-alone

[45] Article 5(1)(d) of the AI Act.
[46] Article 5(1)(c) of the AI Act.
[47] Article 5(1)(f) of the AI Act.
[48] See also Smuha et al. (n 14) 27.
[49] Article 5(1)(e) of the AI Act.
[50] Article 5(1)(g) of the AI Act. The four latter practices were introduced by the European Parliament in its June 2023 negotiating mandate (along with other spurious practices that, unfortunately, did not survive the trilogue with the Commission and the Council).
[51] Article 5(1)(h) of the AI Act.

high-risk AI systems, which are mainly of concern due to their adverse fundamental rights implications and exhaustively listed in Annex III, referring to eight domains in which AI systems can be used. These stand-alone high-risk systems are arguably the most important category of systems regulated under the AI Act (since those in Annex I are already regulated by specific legislation), and will hence be our main focus.

Only the AI applications that are explicitly listed under one of those eight domains headings are deemed high-risk (see Table 12.1). While the list of applications under each domain can be updated over time by the European Commission, the domain headings themselves cannot.[52] The domains include biometrics; critical infrastructure; educational and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and essential public services and benefits; law enforcement; migration, asylum and border control management; and the administration of justice and democratic processes. Even if their system is listed in Annex III, AI providers can self-assess whether their system *truly* poses a significant risk to harm "*health, safety or fundamental rights*" and only then are they subjected to the high-risk requirements.[53]

High-risk systems must comply with "essential requirements" set out in Articles 8 to 15 of the AI Act (Chapter III, Section 2). These requirements pertain, inter alia, to:

- the establishment, implementation, documentation and maintenance of a risk-management system pursuant to Article 9;
- data quality and data governance measures regarding the datasets used for training, validation, and testing; ensuring the suitability, correctness and representativeness of data; and monitoring for bias pursuant to Article 10;
- technical documentation and (automated) logging capabilities for record-keeping, to help overcome the inherent opacity of software, pursuant to Articles 11 and 12;
- transparency provisions, focusing on information provided to enable deployers to interpret system output and use it appropriately as instructed through disclosure of, for example, the system's intended purpose, capabilities, and limitations, pursuant to Article 13;
- human oversight provisions requiring that the system can be effectively overseen by natural persons (e.g., through appropriate human–machine interface tools) so as to minimize risks, pursuant to Article 14;
- the need to ensure an appropriate level of accuracy, robustness, and cybersecurity and to ensure that the systems perform consistently in those respects throughout their lifecycle, pursuant to Article 15.

[52] Article 7 of the AI Act establishes a procedure for the Commission to amend Annex III through delegated acts. The domain headings can only be adapted by the EU legislator through a revision of the regulation itself.

[53] Article 6(3) of the AI Act. To avoid misuse of this provision, the AI Act states that such providers must justify why, despite being included in Annex III, their system does not pose a significant risk. Article 6 establishes a procedure for the European Commission to challenge their justification and to impose the high-risk requirements in case the justification is flawed.

TABLE 12.1 *High-risk AI systems listed in Annex III*

1. Biometric AI systems	• remote biometric identification systems (excluding biometric verification the sole purpose of which is to confirm that a specific natural person is the person he or she claims to be); • biometric categorisation according to sensitive or protected attributes or characteristics based on the inference of those attributes or characteristics; • emotion recognition systems.
2. Critical infrastructure	AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity.
3. Education and vocational training	AI systems intended to be used: • to determine access or admission or to assign natural persons to educational and vocational training institutions at all levels • to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels; • for the purpose of assessing the appropriate level of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions at all levels; • for monitoring and detecting prohibited behaviour of students during tests in the context of or within educational and vocational training institutions at all levels.
4. Employment, workers management and access to self-employment	AI systems intended to be used: • for the recruitment or selection of natural persons, in particular to place targeted job advertisements, to analyse and filter job applications, and to evaluate candidates; • to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behaviour or personal traits or characteristics or to monitor and evaluate the performance and behaviour of persons in such relationships.
5. Access to and enjoyment of essential private services and essential public services and benefits	AI systems intended to be used: • by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services; • to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud; • for risk assessment and pricing in relation to natural persons in the case of life and health insurance;

(continued)

TABLE 12.1 (*continued*)

	• to evaluate and classify emergency calls by natural persons or to be used to dispatch, or to establish priority in the dispatching of, emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems.
6. Law enforcement, in so far as their use is permitted under relevant Union or national law	AI systems intended to be used by or on behalf of law enforcement authorities, or by Union institutions, bodies, offices or agencies in support of law enforcement authorities or on their behalf: • to assess the risk of a natural person becoming the victim of criminal offences; • as polygraphs or similar tools; • to evaluate the reliability of evidence in the course of the investigation or prosecution of criminal offences; • for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups; • for the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of the detection, investigation or prosecution of criminal offences.
7. Migration, asylum and border control management, in so far as their use is permitted under relevant Union or national law	AI systems intended to be used by or on behalf of competent public authorities or by Union institutions, bodies, offices or agencies: • to assess a risk, including a security risk, a risk of irregular migration, or a health risk, posed by a natural person who intends to enter or who has entered into the territory of a Member State; • to assist competent public authorities for the examination of applications for asylum, visa or residence permits and for associated complaints with regard to the eligibility of the natural persons applying for a status, including related assessments of the reliability of evidence; • in the context of migration, asylum or border control management, for the purpose of detecting, recognising or identifying natural persons, with the exception of the verification of travel documents.
8. Administration of justice and democratic processes	AI systems intended to be used: • by a judicial authority or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution; • for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda. This does not include AI systems to the output of which natural persons are not directly exposed, such as tools used to organise, optimise or structure political campaigns from an administrative or logistical point of view.

Finally, Articles 16 and 17 require that high-risk AI providers[54] establish a "quality management system" that must include, among other things, the aforementioned risk management system imposed by Article 9 and a strategy for regulatory compliance, including compliance with conformity assessment procedures for the management of modifications for high-risk AI. These two systems – the risk management system and the quality management system – can be understood as the AI Act's pièce de resistance. While providers have the more general obligation to demonstrably ensure compliance with the "essential requirements," most of these requirements are concerned with technical functionality, and are expected to offer assurance that AI systems will function as stated and intended, that the software's functional performance will be reliable, consistent, "without bias," and in accordance with what providers claim about system design and performance metrics. To the extent that consistent software performance is a prerequisite for facilitating its "safe" and "rights-compliant" use, these are welcome requirements. They are, however, not primarily concerned, in a direct and unmediated manner, with guarding against the dangers ("risks") that the AI Act specifically states it is intended to protect against, notably potential dangers to health, safety and fundamental rights.

This is where the AI Act's characterization of the relevant "risks," which the Article 9 risk management system must identify, estimate and evaluate, is of importance. Article 9(2) refers to "*the known and reasonably foreseeable risks that the high-risk AI system can pose to health, safety or fundamental rights*" when used in accordance with its intended purpose and an estimate and evaluation of risks that may emerge under conditions of "*reasonably foreseeable misuse*."[55] Risk management measures must be implemented such that any "*residual risk associated with each hazard*" and the "*relevant residual risk of the high-risk AI system*" is judged "*acceptable*."[56] High-risk AI systems must be tested prior to being placed on the market to identify the "most appropriate" risk management measures and to ensure the systems "perform consistently for their intended purposes," in compliance with the requirements of Section 2 and in accordance with "appropriate" preliminarily defined metrics and probabilistic thresholds – all of which are to be further specified.

While, generally speaking, the imposition of new obligations is a positive development, their likely effectiveness is a matter of substantial concern. We wonder, for instance, whether it is at all *acceptable* to delegate the identification of risks and their evaluation as "acceptable" to AI providers, particularly given the fact that their assessment might differ very significantly from those who are the relevant risk-bearers

[54] Articles 23 to 27 also set out some obligations for importers, distributors and deployers of high-risk AI systems.

[55] Article 9(2)(a) and (b) of the AI Act.

[56] Article 9(5) of the AI Act.

and who are most likely to suffer adverse consequences if those risks ripen into harm or rights-violations. Furthermore, Article 9(3) is ambiguous: purporting to limit the risks that must be considered as part of the risk management system to "*those which may be reasonably mitigated or eliminated through the development or design of the high-risk AI system, or the provision of adequate technical information.*"[57] As observed elsewhere, this could be interpreted to mean that risks that *cannot* be mitigated through the high-risk system's development and design or by the provision of information can be ignored altogether,[58] although the underlying legislative intent, as stated in Article 2, suggests an alternative reading such that if those "unmitigatable risks" are unacceptable, the AI system cannot be lawfully placed on the market or put into service.[59]

Although the list-based approach to the classification of high-risk systems was intended to provide legal certainty, critics pointed out that it is inherently prone to problems of under and over-inclusiveness.[60] As a result, problematic AI systems that are not included in the list are bound to appear on the market, and might not be added to the Commission's future list-updates. In addition, allowing AI providers to self-assess whether their system actually poses a significant risk or not undermines the legal certainty allegedly offered by the Act's list-based approach.[61] Furthermore, under pressure from the European Parliament, high-risk AI *deployers* that are bodies governed by public law, or are private entities providing public services, must also carry out a "fundamental rights impact assessment" before the system is put into use.[62] However, the fact that an "automated tool" will be provided to facilitate compliance with this obligation "in a simplified manner" suggests that the regulation of these risks is likely to descend into a formalistic box-ticking exercise in which formal documentation takes precedence over its substantive content and real-world effects.[63] While some companies might adopt a more prudent approach, the effectiveness of the AI Act's protection mechanisms will ultimately depend on how its oversight and enforcement mechanisms will operate on-the-ground, which we believe, for reasons set out below, are unlikely to provide a muscular response.

[57] Article 9(3) of the AI Act.

[58] See Nathalie A. Smuha, *Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law* (Cambridge University Press, 2025), Chapter 5.4.

[59] Article 26(5) also states that: "*where deployers have reason to consider that the use of the high-risk AI system in accordance with the instructions may result in that AI system presenting a risk within the meaning of Article 79(1), they shall, without undue delay, inform the provider or distributor and the relevant market surveillance authority, and shall suspend the use of that system.*"

[60] See Karen Yeung, "Response to European Commission White Paper," Social Science Research Network, 2020, https://ssrn.com/abstract=3626915; Nathalie A. Smuha et al., n (14).

[61] That said, as noted in n (53), AI providers who self-assess their high-risk system as excluded from the Act's requirements will still need to justify their assessment and register their system in a newly established database, managed by the Commission. See Article 49(2) of the AI Act.

[62] Article 27 of the AI Act.

[63] Article 27(5) of the AI Act.

12.3.2.3 General-Purpose AI Models

The AI Act defines a general-purpose AI (GPAI) model as one 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 can be integrated into a variety of downstream systems or applications (GPAI systems).[64] The prime example of GPAI models are Large Language Models (LLMs) that converse in natural language and generate text (which, for instance, form the basis of Open AI's Chat-GPT or Google's Bard), yet there are also models that can generate images, videos, music or some combination thereof.

The primary obligations of GPAI model-providers are to draw up and maintain technical documentation, comply with EU copyright law and disseminate "sufficiently detailed" summaries about the content used for training models before they are placed on the market.[65] These minimum standards apply to all models, yet GPAI models that are classified as posing a "systemic risk" due to their "high impact capabilities" are subject to additional obligations. Those include duties to conduct model evaluations, adversarial testing, assess and mitigate systemic risks, report on serious incidents, and ensure an adequate level of cybersecurity.[66] Note, however, that providers of (systemic risk) GPAI models can conduct their own audits and evaluations, rather than rely on external independent third party audits. Nor is any public licensing scheme required.

More problematically, while the criteria to qualify GPAI models as posing a "systemic risk" are meant to capture their "*significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain*,"[67] the legislator opted to express these criteria in terms of a threshold pertaining to the size of the data on which the models are trained. Models trained with more than 10^{25} floating-point operations reach this threshold and are presumed to qualify as posing a systemic risk.[68] This threshold, though amendable, is rather arbitrary, as many existing models do not cross that threshold but are nevertheless capable of posing systemic risks. More generally, limiting "systemic risks" to those arising from GPAI models is difficult to justify, given that even traditional rule-based AI systems with far more limited capabilities can pose systemic risks.[69] Moreover, as Hacker has observed,[70] the industry is moving

[64] Article 3(63) of the AI Act. It does exclude AI models used for research, development or prototyping activities before their placement on the market.
[65] Article 53(1) of the AI Act.
[66] Article 55(1) of the AI Act.
[67] Article 3(65) of the AI Act.
[68] Article 51(2) of the AI Act.
[69] See in this regard also Smuha, n (58), Chapter 5.4.
[70] See Philipp Hacker, "What's missing from the EU AI Act: Addressing the four key challenges of large language models," *VerfassungsBlog*, December 13, 2023, https://verfassungsblog.de/whats-missing-from-the-eu-ai-act/.

toward smaller yet more potent models, which means many more influential GPAI models may fall outside the Act, shifting the regulatory burden "to the downstream deployers."[71] Although these provisions can, in theory, be updated over time, their effectiveness and durability are open to doubt.[72]

12.3.2.4 Systems Requiring Additional Transparency

For a subset of AI applications, the EU legislator acknowledged that specific risks can arise, such as impersonation or deception, which stand apart from high-risk systems. Pursuant to Article 50 of the AI Act, these applications are subjected to additional transparency obligations, yet they might also fall within the high-risk designation. Four types of AI systems fall into this category. The first are systems intended to interact with natural persons, such as chatbots. To avoid people mistakenly believing they are interacting with a fellow human being, these systems must be developed in such a way that the natural person who is exposed to the system is informed thereof, in a timely, clear and intelligible manner (unless this is obvious from the circumstances and context of the use). An exception is made for AI systems authorized by law to detect, prevent, investigate, and prosecute criminal offences.

A similar obligation to provide transparency exists when people are subjected either to an emotion recognition system or a biometric categorization system (to the extent it is not prohibited by Article 5 of the AI Act). Deployers must inform people subjected to those systems of the system's operation and must, pursuant to data protection law, obtain their consent prior to the processing of their biometric and other personal data. Again, an exception is made for emotion recognition systems and biometric categorization systems that are permitted by law to detect, prevent, and investigate criminal offences.

Finally, providers of AI systems that generate synthetic audio, image, video or text must ensure that the system's outputs are marked in a machine-readable format and are detectable as artificially generated or manipulated.[73] Deployers of such systems should disclose that the content has been artificially generated or manipulated.[74] This provision was already present in the Commission's initial AI Act proposal, but

[71] If a GPAI system is deployed for the purpose of one of the high-risk applications listed in Annex III – and if it is self-assessed as posing a significant risk – it will need to comply with the standard requirements for high-risk systems as listed in Chapter III, Section 2.

[72] It should however be noted that the European Commission can also designate certain GPAI models as posing a systemic risk through a decision, either ex officio or based on a qualified alert by a scientific panel that the AI Act will set up for this purpose. It is also able to amend the thresholds through delegated acts. Moreover, at least in theory, also systems that do not fall under the specified threshold can be considered as posing a systemic risk if they show high impact capabilities evaluated on the basis of "appropriate technical tools and methodologies, including indicators and benchmarks," which the Commission can supplement over time.

[73] Article 50(2) of the AI Act.

[74] Article 50(4) of the AI Act.

it became far more relevant with the boom of generative AI, which "democratized" the creation of deep fakes, enabling them to be easily created by those without specialist skills. As regards AI systems that generate or manipulate text, which is published with "*the purpose of informing the public on matters of public interest*," deployers must disclose that the text was artificially generated or manipulated, unless the AI-generated content underwent a process of human review or editorial control with editorial responsibility for its publication.[75] Here, too, exceptions exist. In each case, the disclosure measures must take into account the generally acknowledged state of the art, whereby the AI Act also refers to relevant harmonized standards,[76] to which we will return later.

12.3.2.5 Non-High-Risk Systems

All other AI systems that do not fall under one of the aforementioned risk-categories are effectively branded as "no risk" and do not attract new legal obligations. To the extent they fall under existing legal frameworks – for instance, when they process personal data – they must still comply with those frameworks. In addition, the AI Act provides that the European Commission, Member States and the AI Office (a supervisory entity that we discuss in the next section) should encourage and facilitate the drawing up of codes of conduct that are intended to foster the voluntary application of the high-risk requirements to those no-risk AI systems.[77]

12.3.3 *Supporting Innovation*

The White Paper on AI focused not only on the adoption of rules to *limit* AI-related risks, but also included a range of measures and policies to *boost* AI innovation in the EU. Clearly, the AI Act is a tool aimed primarily at achieving the former, but the EU still found it important to also emphasize its "pro-innovation" stance. Chapter VI of the AI Act therefore lists "measures in support of innovation," which fits into the EU's broader policy narrative which recognizes that regulation can facilitate innovation, and even provide a "competitive advantage" in the AI "race."[78] These measures mainly concern[79] the introduction of AI regulatory sandboxes, which are intended to offer a safe and controlled environment for AI providers to develop, test, and validate AI systems, including the facilitation of "real-world-testing." National authorities must oversee these sandboxes and help

[75] Ibid.
[76] Article 50(2) of the AI Act.
[77] Articles 95 and following of the AI Act.
[78] See European Commission, n (8), 2.
[79] One could argue that the abovementioned derogations for open-source AI systems can likewise be seen as an innovation-boosting measure. See *supra*, n (41).

ensure that appropriate safeguards are in place, and that their experimentation occurs in compliance with the law. The AI Act mandates each Member State to establish at least one regulatory sandbox, which can also be established jointly with other Member States.[80] To avoid fragmentation, the AI Act further provides for the development of common rules for the sandboxes' implementation and a framework for cooperation between the relevant authorities that supervise them, to ensure their uniform implementation across the EU.[81]

Sandboxes must be made accessible especially to Small and Medium Enterprises (SMEs), thereby ensuring that they receive additional support and guidance to achieve regulatory compliance while retaining the ability to innovate. In fact, the AI Act explicitly recognizes the need to take into account the interests of "small-scale providers" and deployers of AI systems, particularly costs.[82] National authorities that oversee sandboxes are hence given various tasks, including increasing awareness on the regulation, promoting AI literacy, offering information and communication services to SMEs, start-ups, and deployers, and helping them identify methods that lower their compliance costs. Collectively, these measures are aimed to offset the fact that smaller companies will likely face heavier compliance and implementation burdens, especially compared to large tech companies that can afford an army of lawyers and consultants to implement the AI Act. It is also hoped that the sandboxes will help national authorities to improve their supervisory methods, develop better guidance, and identify possible future improvements of the legal framework.

12.4 MONITORING AND ENFORCEMENT

Our discussion has hitherto focused on the substantive dimensions of the Act. However, whether these provide effective protection of health, safety and fundamental rights will depend critically on the strength and operation of its monitoring and enforcement architecture, to which we now turn. We have already noted that the proposed regulatory enforcement framework underpinning the Commission's April 2021 blueprint was significantly flawed, yet these flaws remain unaltered in the final Act. As we shall see, the AI Act allocates considerable interpretative discretion to the industry itself, through a model which has been described by regulatory theorists as "meta-regulation." We also discuss the Act's approach to technical standards and the institutional framework for evaluating whether high-risk AI systems are in compliance with the Act, to argue that the regime as a whole fails to offer adequate protection against the adverse effects that it purports to counter.

[80] Article 57(1) of the AI Act.
[81] Article 58 of the AI Act.
[82] See, for example, Article 34(2) of the AI Act.

12.4.1 *Legal Rules and Interpretative Discretion*

Many of the AI Act's core provisions are written in broad, open-ended language, leaving the meaning of key terms uncertain and unresolved. It will be here that the rubber will hit the road, for it is through the interpretation and application of the Act's operative provisions that it will be given meaning and be translated into on-the-ground practice.

For example, when seeking to apply the essential requirements applicable to high-risk systems, three terms used in Chapter III, Section 2 play a crucial role. First, the concept of "risk." Article 3 defines risk as "*the combination of the probability of an occurrence of harm and the severity of that harm*," reflecting conventional statistical risk assessment terminology. Although risks to health and safety is a relatively familiar and established concept in legal parlance and regulatory regimes, the Annex III high-risk systems are more likely to interfere with fundamental rights and may adversely affect democracy and the rule of law. But what, precisely, is meant by "risk to fundamental rights," and how should those risks be identified, evaluated and assessed? Secondly, even assuming that fundamental rights-related risks can be meaningfully assessed, how then is a software firm to adequately evaluate what constitutes a level of residual risk judged "acceptable"? And thirdly, what constitutes a "risk management system" that meets the requirements of Article 9?

The problem of interpretative discretion is not unique to the AI Act. All rules which take linguistic form, whether legally mandated or otherwise, must be interpreted before they can be applied to specific real-world circumstances. Yet how this discretion is exercised, and by whom, will be a product of the larger regulatory architecture in which those rules are embedded. The GDPR, for instance, contains a number of broadly defined "principles" which those who collect and process personal data must comply with. Both the European Data Protection Board (EDPB) and national level data protection authorities – as public regulators – issue "guidance" documents offering interpretative guidance about what the law requires. Compliance with this guidance (often called "soft law") does not guarantee compliance – for it does not bind courts when interpreting the law – but it nevertheless offers a valuable, and reasonably authoritative assistance to those seeking to comply with their legal obligations. This kind of guidance is open, published, transparent, and conventionally issued in draft form before-hand so that stakeholders and the public can provide feedback before it is issued in final form.[83]

In the AI Act, similar interpretative decisions will need to be made and, in theory, the Commission has a mandate to issue guidelines on the AI Act's practical implementation.[84] However, in contrast with the GDPR, the Act's adoption of the "New

[83] See Yeung and Ranchordas, n (42), Chapter 8.

[84] Article 96 of the AI Act. When issuing such guidelines, the Commission "*shall take due account of the generally acknowledged state of the art on AI, as well as of relevant harmonised standards and common*

Approach" to product-safety means that, in practice, providers of high-risk AI systems will likely adhere to technical standards produced by European Standardization Organizations on request from the Commission and which are expected to acquire the status of "harmonized standards" by publication of their titles in the EU's Official Journal.[85] As we explain below, the processes through which these standards are developed are difficult to characterize as democratic, transparent or based on open public participation.

12.4.2 *The AI Act as a Form of "Meta-Regulation"*

At first glance, the AI Act appears to adopt a public enforcement framework with both national and European public authorities playing a significant role. Each EU Member State must designate a national supervisory authority[86] to act as "market surveillance authority."[87] These authorities can investigate suspected incidents and infringements of the AI Act's requirements, and initiate recalls or withdrawals of AI systems from the market for non-compliance.[88] National authorities exchange best practices through a *European AI Board* comprised of Member States' representatives. The European Commission has also set up an AI Office to coordinate

specifications that are referred to in Articles 40 and 41, or of those harmonised standards or technical specifications that are set out pursuant to Union harmonisation law."

[85] See Articles 40 and 41 of the AI Act. A harmonized standard is a European standard developed by a recognized European Standardization Organization and its creation is requested by the European Commission. The references of harmonized standards must be published in the Official Journal of the EU. See https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards_en, accessed June 20, 2024.

[86] Member States are free to establish a new entity for this purpose, or they can designate an existing authority. They can also assign this task to several existing authorities, as long as they designate one of those authorities as the main authority and contact point for practical purposes. See Article 70 of the AI Act.

[87] Under the New Legislative Framework for product safety legislation, (national) market surveillance authorities have the task to monitor the market and, in case of doubt, to verify ex post whether the conformity assessment has correctly been carried out, and the CE mark duly affixed. This market surveillance authority can be a separate entity, or it can be the same authority that is also responsible for the supervision of the implementation of a regulation. As regards the regime of the AI Act, for all stand-alone high-risk systems, it provides that the national supervisory authority is also the market surveillance authority. For high-risk systems that are already covered by legal acts listed in Annex I (and that are hence already subject to a monitoring system, such as toys or medical devices), the competent authorities under those legal acts will remain the lead market surveillance authority, though cooperation is encouraged.

[88] The supervisory authorities should act independently and impartially in performing their tasks and exercising their powers. These powers consist of e.g. requesting the technical documentation and records that providers of high-risk systems must create and – if they exhausted all other reasonable ways to verify the system's conformity, they can also request access to the system's training, validation and testing datasets, the trained and training model of the high-risk AI system, including its relevant model parameters. Pursuant to Article 74(13) of the AI Act, national supervisory authorities can exceptionally also obtain access to the source code of a high-risk AI system, upon a reasoned request. Any information must be treated as confidential, and with respect to intellectual property rights and trade secrets.

enforcement at the EU level.[89] Its main task is to monitor and enforce the requirements relating to GPAI models,[90] yet it also undertakes several other roles, including (a) guiding the evaluation and review of the AI Act over time,[91] (b) offering coordination support for joint investigations between the Commission and Member States when a high-risk system presents a serious risk across multiple Member States,[92] and (c) facilitating the drawing up of voluntary codes of conduct for systems that are not classified as high-risk.[93]

The AI Office will be advised by a *scientific panel of independent experts* to help it develop methodologies to evaluate the capabilities of GPAI models, to designate GPAI models as posing a systemic risk, and to monitor material safety risks that such models pose. An *advisory forum of stakeholders* (to counter earlier criticism that stakeholders were allocated no role whatsoever in the regulation) is also established under the Act, to provide both the Board and the Commission with technical expertise and advice. Finally, the Commission is tasked with establishing a public EU-wide database where providers (and a limited set of deployers) of stand-alone high-risk AI systems must register their systems to enhance transparency.[94]

In practice, however, these public authorities are twice-removed from where much of the *real-world* compliance activity and evaluation takes place. The AI Act's regulatory enforcement framework delegates many crucial functions (and thus considerable discretionary power) to the very actors whom the regime purports to regulate, and to other tech industry experts. The entire architecture of the AI Act is based on what regulatory governance scholars sometimes refer to as "meta-regulation" or "enforced self-regulation."[95] This is a regulatory technique in which legally binding obligations are imposed on regulated organizations, requiring them to establish and maintain internal control systems that meet broadly specified, outcome-based, binding legal objectives.

Meta-regulatory strategies rest on the basic idea that one size does not fit all, and that firms themselves are best placed to understand their own operations and systems and take the necessary action to avoid risks and dangers. The primary safeguards through which the AI Act is intended to work rely on the quality and risk management systems within the regulated organizations, in which these organizations

[89] The establishment of the AI Office reflects the desire of both the European Parliament and the Council to have a stronger involvement at the EU level when it comes to implementing and enforcing the AI Act. Over time, the AI office could become a full-fledged European AI Agency.

[90] Articles 53 and following of the AI Act. For those models, the AI Office will also contribute to fostering standards and testing practices and enforcing common rules in all member states.

[91] Especially for those provisions that the Commission cannot adapt through a delegated act, but that can only be amended by the legislators (such as the domain headings under Annex III or the prohibited AI practices). See Article 112(11) of the AI Act.

[92] Article 74(11) of the AI Act.

[93] Article 95 of the AI Act.

[94] Article 71 of the AI Act.

[95] See Yeung and Ranchordas, n (42), Chapter 7 and literature cited therein.

retain considerable discretion to establish and maintain their own internal standards of control, provided that the Act's legally mandated objectives are met. The supervisory authorities oversee adherence to those internal standards, but they only play a secondary and reactionary role, which is triggered if there are grounds to suspect that regulated organizations are failing to discharge their legal obligations. While natural and legal persons have the right to lodge a complaint when they have grounds to consider that the AI Act was infringed,[96] supervisory authorities do not have any proactive role to ensure the requirements are met before high-risk AI systems are placed on the market or deployed.

This compliance architecture flows from the underlying foundations of the Act, which are rooted in the EU's "New Legislative Framework," adopted in 2008. Its aim was to improve the internal market for goods and strengthen the conditions for placing a wide range of products on the EU market.[97]

The AI Act largely leaves it to Annex III high-risk AI providers and deployers to self-assess their conformity with the AI Act's requirements (including, as discussed earlier, the judgment of what is deemed an "acceptable" residual risk). There is no routine or regular inspection and approval or licensing by a public authority. Instead, if they declare that they have self-assessed their AI system as compliant and duly lodge a declaration of conformity, providers can put their AI systems into service without any independent party verifying whether their assessment is indeed adequate (except for certain biometric systems).[98] Providers are, however, required to put in place a post-market monitoring system, which is intended to ensure that the possible risks emerging from AI systems that continue to "learn" or evolve once placed on the market or put into service can be better identified and addressed.[99] The role of

[96] Article 85 of the AI Act. Article 86 also grants affected persons who are subjected to (most) high-risk AI systems listed in Annex III the 'right to an explanation', covering the "*right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken*." This right however only applies if the decision "*produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights*," and national or Union law can provide exceptions to this right.

[97] It refers to a package of measures intended to: improve market surveillance; establish a framework of rules for product safety; enhance the quality of and confidence in the conformity assessment of products through stronger and clearer rules on notification requirements of conformity assessment bodies; and clarify the meaning of CE markings to enhance their credibility. This package of measures consists of Regulation (EC) 765/2008, which sets out the requirements for accreditation and the market surveillance of products, Commission Decision 768/2008 on a common framework for the marketing of products, which is effectively a template for future product harmonisation legislation and Regulation (EU) 2019/1020 on market surveillance and compliance of products, which aims to govern the role of various economic operators (manufacturers, authorised representatives, importers) and standardizing their tasks with regard to the placing of products on the market.

[98] See Article 43 of the AI Act.

[99] High-risk AI providers and deployers must also have a system in place to report to the relevant authorities any serious incidents or breaches of national and Union law, and take appropriate corrective actions.

public regulators is therefore largely that of *ex post* oversight, unlike the European regulation of pharmaceuticals, reflecting the regulatory regime as permissive rather than precautionary. This embodies the basic regulatory philosophy underpinning the New Legislative Framework, which builds on the "New Approach" to technical standardization. Together, these are concerned first and foremost with strengthening single market integration, and hence with ensuring a single EU market for AI.

12.4.3 *The New Approach to Technical Standardization*

Under the EU's "Old Approach" to product safety standards, national authorities drew up detailed technical legislation, which was often unwieldy and usually motivated by a lack of confidence in the rigour of economic operators on issues of public health and safety. However, the "New Approach" framework introduced in 1985 sought instead to restrict the content of legislation to "essential requirements," leaving technical details to European Harmonized Standards[100] thereby laying the foundation for technical standards produced by European Standardization Organizations (ESOs) in support of Union harmonization legislation.[101]

The animating purpose of the "New Approach" to standardization was to open up European markets in industrial products without threatening the safety of European consumers, by allowing the entry of those products across European markets if and only if they meet the "essential [safety] requirements" set out in sector-specific European rules developed by one of the three ESOs: the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC) and the European Telecommunications Standards Institute (ETSI).[102]

[100] The decision of the Court of Justice of the EU (CJEU) in *Cassis de Dijon* in 1979 was highly significant. The Court ruled that products lawfully manufactured or marketed in one Member State should in principle move freely throughout the Union where such products meet equivalent levels of protection to those imposed by the Member State of destination, and that barriers to free movement which result from differences in national legislation may only be accepted under specific circumstances, namely (1) the national measures are necessary to satisfy mandatory requirements (such as health, safety, consumer protection and environmental protection), (2) they serve a legitimate purpose which justifies overriding the principle of free movement of goods, and (3) they can be justified with regard to the legitimate purpose and are proportionate with the aims. See *Case 120/78 Cassis de Dijon* [1979] ECR 649 (*Rewe-Zentral v Bundesmonopolverwaltung für Branntwein*).

[101] Yet in practice, the framework did not create the necessary level of trust between Member States. Therefore, in 1989 and 1990, the "Global Approach" was adopted, which established general guidelines and detailed procedures for conformity assessment to cover a wide range of industrial and commercial products.

[102] See in this regard Jean-Pierre Galland, "Big Third-Party Certifiers and the Construction of Transnational Regulation" (2017) *The ANNALS of the American Academy of Political and Social Science*, 670(1), 263–279. This New Legislative Framework consists of a tripartite package of EU measures (1) EC Regulation No 765/2008 on accreditation and marketing surveillance (2) Decision No 768/2008/EC on establishing a common framework for the marketing of products (3) EC Regulation

Under this approach, producers can choose to *either* interpret the relevant EU Directive themselves *or* to rely on "harmonized (European) standards" drawn up by one of the ESOs. This meta-regulatory approach combines compulsory regulation (under EU secondary legislation) and "voluntary" standards, made by ESOs. Central to this approach is that conformity of products with "essential safety requirements" is checked and certified by *producers themselves* who make a declaration of conformity and affix the CE mark to their products to indicate this, thereby allowing the product to be marketed and sold across the whole of the EU. However, for some "sensitive products," conformity assessments must be carried out by an independent third-party "notified body" to certify conformity and issue a declaration of conformity. This approach was taken by the Commission in its initial AI Act proposal, and neither the Parliament nor the Council has sought to depart from it. By virtue of its reliance on the "New Approach," the AI Act lays tremendous power in the hands of private, technical bodies who are entrusted with the task of setting technical standards intended to operationalize the "essential requirements" stipulated in the AI Act.[103]

In particular, providers of Annex III high-risk AI systems that fall under the AI Act's requirements have three options. First, they can self-assess the compliance of their AI systems with the essential requirements (which the AI Act refers to as the conformity assessment procedure based on internal control, set out in Annex VI). Under this option, whenever the requirements are vague, organizations need to use their own judgment and discretion to interpret and apply them, which – given considerable uncertainty about what they require in practice – exposes them to potential legal risks (including substantial penalties) if they fail to meet the requirements.

Second, organizations can rely on a conformity assessment by a "notified body,"[104] which they can commission to undertake the conformity assessment. These bodies are independent yet nevertheless "private" organizations that verify the conformity of AI systems based on an assessment of the quality management system and the technical documentation (a procedure set out in Annex VII). AI providers pay for these certification services, with a flourishing "market for certification" emerging in response. To carry out the tasks of a notified body, it must meet the requirements of Article 31 of the AI Act, which are mainly concerned with ensuring that they possess the necessary competences, a high degree of professional integrity, and that they are independent from and impartial to the organizations they assess to avoid conflicts of interest. Pursuant to the AI Act, only providers of biometric identification systems

No 764/2008 to strengthen the internal market for a wide range of other products not subject to EU harmonisation.

[103] See Commission Implementing Decision of 22 May 2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence, Brussels, 22 May 2023, C(2023) 3215 final.

[104] This is because an organization that seeks to act as an independent third-party certifier first needs to receive accreditation from a national notifying authority which evaluates and monitors that these third-party certifiers meet certain quality and independence standards.

must currently undergo an assessment by a notification body. All others can opt for the first option (though in the future, other sensitive systems may also be obliged to obtain approval via third-party conformity assessment).

Third, AI providers can choose to follow voluntary standards currently under development by CEN/CENELEC following acceptance of the Commission's standardization request which are intended, once drafted, to become "harmonized standards" following citation in the Official Journal of the European Commission. This would mean that AI providers and deployers could choose to follow these harmonized standards and thereby benefit from a legal *presumption of conformity* with the AI Act's requirements. Although the presumption of compliance is rebuttable, it places the burden of proving non-compliance on those claiming that the AI Act's requirements were not met, thus considerably reducing the risk that the AI provider will be found to be in breach of the Act's essential requirements. If no harmonized standards are forthcoming, the Commission can adopt "common specifications" in respect of the requirements for high-risk systems and GPAI models, which likewise, will confer a presumption of conformity.[105]

Thus, although harmonized standards produced by ESOs are formally voluntary, providers are strongly incentivized to follow them (or, in their absence, to follow the common specifications) rather than carrying the burden of demonstrating that their own specifications meet the law's essential requirements. This means that harmonized standards are likely to become binding *de facto*, and will therefore in practice determine the nature and level of protection provided under the AI Act. The overwhelming majority of providers of Annex III high-risk systems can self-assess their own internal controls, sign and lodge a conformity assessment declaration, affix a CE mark to their software, and then notify the Commission's public register.

12.4.4 *Why Technical Standardization Falls Short in the AI Act's Context*

Importantly, however, several studies have found that products that have been self-certified by producers are considerably more likely to fail to meet the certified standard. For example, Larson and Jordan[106] compared toy safety recalls in the US, within a toy safety regime requiring independent third-party verification, and the EU's toy self-certification regime which relies on self-assessment and found stark differences. Over a two-year period, toy safety recalls in the EU were 9 to 20 times more frequent than those in the US. Their findings align with earlier policy studies finding that self-assessment models consistently produce substantially higher rates of worker injury compared with those involving independent third-party evaluation. Based on these studies, Larson and Jordon conclude that transnational product

[105] Article 41 of the AI Act.

[106] Derek B. Larson and Sara R. Jordan, "Playing it safe: toy safety and conformity and assessment in Europe and the US" (2018) *International Review of Administrative Sciences*, 85(4), 763–79.

safety regulatory systems that rely on the self-assessment of conformity with safety standards fail to keep products off the market, which do not comply with those standards.

What is more, even third-party certification under the EU's New Approach has shown itself to be weak and ineffective, as evidenced by the failure of the EU's Medical Device regime which prevailed before its more recent reform. This was vividly illustrated by the PIP breast implants scandal in which approximately 40,000 women in France, and possibly 10 times more in Europe and worldwide, were implanted with breast implants that were filled with industrial grade silicon, rather than the compulsory medical grade standard required under EU law.[107] This occurred despite the fact that the implants had been certified as "CE compliant" by a reputable German notified body, which was possible because, under the relevant directive,[108] breast implant producers could choose between different methods of inspection. PIP had chosen the "full quality assurance system," whereby the certifiers' job was to audit PIP's quality management system without having to inspect the breast implants themselves. In short, the New Approach has succeeded in fostering flourishing markets for certification services – but evidence suggests that it cannot be relied on systematically to deliver trustworthy products and services that protect individuals from harm to their health and safety.

Particularly troubling is the New Approach's reliance on testing the *quality of internal document keeping and management systems*, rather than an inspection and evaluation of the service or product itself.[109] As critical accounting scholar Mike Power has observed, the process of "rendering auditable" through measurable procedures and performance – is a test of "the quality of internal systems rather than the quality of the product or service itself specified in standards."[110] As Hopkins emphasizes in his analysis of the core features that a robust "safety case" approach must meet, "*without scrutiny by an independent regulator, a safety case may not be worth the paper it is written on*."[111] The AI Act, however, does not impose any external

[107] See in this regard also Victoria Martindale and Andre Menache, "The PIP scandal: an analysis of the process of quality control that failed to safeguard women from the health risks" (2013) *Journal of the Royal Society of Medicine*, 106(5), 173–77.

[108] Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, OJ L 169, July 12, 1993, 1–43.

[109] This is borne out in Laura Silva-Cataneda, "A forest of evidence: Third-party certification and multiple forms of proof – a case study on oil palm plantations in Indonesia" (2012) *Agriculture and Human Values*, 29(3): 361–70. In her study, she found that in practice, auditors regard the company's documents as the ultimate form of evidence. Villagers who disagree with the company may point to localized and personalized markers but not to documents, and this is regarded by the auditors as a "lack of evidence." Hence, in contrast to the company's documentary arsenal, auditors' unwillingness to recognize the validity of evidence other than in documentary while disregarding the local knowledge of local communities exacerbated the power imbalance between them.

[110] See Michael Power, *The Audit Society: Rituals of Verification* (Oxford University Press, 1997), p. 84.

[111] As Hopkins clarifies, under a safety case regime, when regulators make site visits, "rather than inspecting to ensure that hardware is working, or that documents are up to date, they must audit against the

auditing requirements. For Annex III high-risk AI systems, the compliance evaluation remains primarily limited to verification that there is requisite documentation in place. Accordingly, we are skeptical of the effectiveness of the CE marking regime for delivering meaningful and effective protections for those affected by rights-critical products and services regulated under the Act.[112]

What, then, are the prospects that the technical standards which the Commission has tasked CEN/CENELEC to produce will translate into practice the Act's noble aspirations to protect fundamental rights, health, safety and uphold the rule of law? We believe there are several reasons to worry. Technical standardization processes may appear "neutral" as they focus on mundane technical tasks, conducted in a highly specialized vernacular, yet these activities are in fact highly political. As Lawrence Busch puts it: "Standards are intimately associated with power."[113] Moreover, these standards will not be publicly available. Rather, they are protected by copyright and thus only available on payment.[114] If an AI provider self-certifies its compliance with an ESO-produced harmonized standard, that will constitute "deemed compliance" with the Act. But, if, in fact, that provider has made no attempt to comply with the standard, no-one will be any the wiser unless and until action is taken by a market surveillance authority to evaluate that AI system for compliance, which it cannot do unless it has "sufficient reasons to consider an AI system to present a risk."[115]

In addition, technical standardization bodies have conventionally been dominated by private sector actors who have had both the capacity to develop particular technologies and can leverage their market share to advocate for the standardization of the technology in line with their own products and organizational processes.

safety case, to ensure that the *specified controls are functioning* as intended." See Andrew Hopkins, "Explaining the 'safety case,'" Working Paper 87, Australian National University, April 2012, p. 6.

[112] The EU is currently struggling to implement a wide-ranging change in how medical devices are regulated – from the 1993 Medical Device Directive (MDD) to the 2017 Medical Device Regulation (MDR). Phased introduction of the MDR was due to be completed by May 2020, but was extended until this year due to COVID-19 pressures. This new regulatory framework is designed to ensure more thorough testing of devices before they can be used on patients, requiring clinical investigation and more rigorous monitoring of performance of devices once on the market. The MDR's implementation, however, has not gone smoothly.

[113] Lawrence Bush, *Standards: Recipes for Realities* (The MIT Press, 2011), p. 13.

[114] However, in *Public.Resource.Org, Inc., Right to Know CLG vs. European Commission* (C-588/21 P) the CJEU ruled that the Commission must indeed grant access to the four requested harmonized standards on the basis that harmonized standards form part of EU law and that the rule of law requires that access to harmonized standards must be freely available without charge. There is thus an overriding public interest in free access to the harmonized standards.

[115] See Article 79(2) of the AI Act. Supervisory authorities (in their capacity of market surveillance authorities) are empowered to have access to documentation, datasets and code upon reasoned request, together with other "appropriate technical means and tools enabling remote access" and datasets. However, only if the documentation is "insufficient to ascertain whether a breach of obligations under EU law intended to protect fundamental rights has occurred" can the MSA organize the testing of the high-risk system through technical means (see Article 77(3) of the AI Act).

Standards committees tend to be stacked with people from large corporations with vested interests and extensive resources. As Joanna Bryson has pithily put it, "even when technical standards for software are useful they are ripe for regulatory capture."[116] Nor are they subject to democratic mechanisms of public oversight and accountability that apply to conventional law-making bodies. Neither the Parliament nor the Member States have a binding veto over harmonized standards, and even the Commission has only limited powers to influence their content, at the point of determining whether the standard produced in response to its request meets the essential requirements set out in the Act, but otherwise the standard is essentially immune from judicial review.[117]

Criticisms of the lack of the democratic legitimacy of these organizations has led to moves to open up their standard-setting process to "multi-stakeholder" dialogue, with civil society organizations seeking to get more involved.[118] In practice, however, these moves are deeply inadequate, as civil society struggles to obtain technical parity with their better-resourced counterparts from the business and technology communities. Stakeholder organizations also face various de facto obstacles to use the CEN/CENELEC participatory mechanisms effectively. Most NGOs have no experience in standardization and many lack EU level representation. Moreover, active participation is costly and highly time-consuming.[119]

Equally if not more worrying is the fact that these "technical" standard-setting bodies are populated by experts primarily from engineering and computer science, who typically have little knowledge or expertise in matters related to fundamental rights, democracy, and the rule of law. Nor are they likely to be familiar with the analytical reasoning that is well established in human rights jurisprudence to determine what constitutes an interference with a fundamental right and whether it may be justified as necessary in a democratic society.[120] Without a significant cadre of human rights lawyers to assist them, we are deeply skeptical of the competence

[116] Joanna J. Bryson, "Belgian and Flemish policy makers' guide to AI regulation," KCDS-CiTiP Fellow Lectures Series: Towards an AI Regulator?, Leuven, October 11, 2022.

[117] Although the CJEU decided in the *James Elliot* case that it has jurisdiction to interpret harmonized standards in preliminary ruling procedures, according to Ebers (2022), it is unlikely that the Court would be willing to rule on the validity of a harmonized standard, either in an annulment action (per Article 264 TFEU) or a preliminary ruling procedure (per Article 267 TFEU). And even if it were, the CJEU is unlikely to review and invalidate its substantive content – its jurisdiction would be limited to reviewing whether the Commission made an error in making the decision to publish a harmonized standard in the official journal. See Martin Ebers, "Standardizing AI: The case of the European Commission's proposal for an 'Artificial Intelligence Act,'" in L. A. DiMatteo, C. Poncibò, and M. Cannarsa (eds.), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (Cambridge University Press, 2022), pp. 321–344.

[118] See for example the ANEC and BEUC standardization project: https://anec.eu/projects/ai-standards, accessed June 20, 2024.

[119] CENELEC/CEN standardization committees are dispersed across all corners of Europe, yet most of the meetings now tend to take place online.

[120] Our experiences when piloting the AI HLEG's Trustworthy AI Assessment List showed an across-the-board lack of understanding of what a fundamental rights impact assessment entails, with the majority

and ability of ESOs to translate the notion of "risks to fundamental rights" into tractable technical standards that can be relied upon to facilitate the protection of fundamental rights.[121]

Furthermore, unlike risks to safety generated by chemicals, machinery, or industrial waste, all of which can be materially observed and measured, fundamental rights are, in effect, political constructs. These rights are accorded special legal protection so that an evaluation of alleged interference requires close attention to the nature and scope of the relevant right and the specific, localized context in which a particular right is allegedly infringed. We therefore seriously doubt whether fundamental rights can *ever* be translated into generalized technical standards that can be precisely measured in quantitative terms, and in a manner that faithfully reflects what they are and how they have been interpreted under the European Charter on Fundamental Rights and the European Convention on Human Rights.

Moreover, the CENELEC rules nevertheless state that any harmonized standard must contain "objectively verifiable requirements and test methods,"[122] which does not alleviate our difficulties in trying to conceive of how "risks to fundamental rights" can be subject to quantitative "metrics" and translated into technical standards such that the "residual risk" can be assessed as "acceptable." Taken together, this leaves us rather pessimistic about the capacity and prospects for ESOs (even assuming a well-intentioned technical committee) to produce technical standards that will, if duly followed, provide the high level of protection to European values that the Act claims to aspire to, and which will constitute "deemed compliance" with the regulation. And if, as expected, providers of high-risk AI systems will choose to be guided by the technical standards produced by ESOs, this means that the "real" standard-setting for high-risk systems will take place within those organizations, with little public scrutiny or independent evaluation.

12.5 CONCLUSION

In this chapter, we have recounted the European Union's path toward a new legal framework to regulate AI systems, beginning in 2018 with the European AI strategy and the establishment of a High-Level Expert Group on AI, culminating in the AI Act of 2024. Since most of the AI Act's provisions will only apply two years after its entry into force,[123] we will not be in a position to acquire evidence of its effectiveness until the end of 2026. By then, both those regulated by the Act, and the supervisory

of respondents mystified by the requirement to consider the impact of their AI system on fundamental rights in the first place.

[121] But see recent efforts by Equinet, "Equality-compliant artificial intelligence: Equinet's plans for 2024", available at https://equineteurope.org/latest-developments-in-ai-equality/ (accessed June 20, 2024).

[122] See in this regard the CENELEC Internal Regulations, Part 3.

[123] See Article 113 of the AI Act, which also lists some exceptions.

actors at national and EU level will need to ramp up their oversight and monitoring capabilities. However, by that time, new AI applications may have found their way to the EU market, which – due to the AI Act's list-based approach – will not fall within the Act, or which the Act may fail to guard against. In addition, since the AI Act aspires a maximum market harmonization for AI systems across Member States, any gaps are in principle *not* addressable through national legislation.

We believe that Europe can rightfully be proud of its acknowledgement that the development and use of AI systems requires mandatory legal obligations, given the individual, collective and societal harms they can engender,[124] and we applaud its aspirations to offer a protective legal framework. What remains to be seen is whether the AI Act will in practice deliver on its laudable objectives, or whether it provides a veneer of legal protection without delivering meaningful safeguards in practice. This depends, crucially, on how its noble aspirations are *operationalized* on the ground, particularly through the institutional mechanism and concepts through which the Act is intended to work.

Based on our analysis, it is difficult to conclude that the AI Act offers much more than "motherhood and apple pie." In other words, although it purports to champion noble principles that command widespread consensus, notably "European values" including the protection of democracy, fundamental rights, and the rule of law, whether it succeeds in giving concrete expression to those principles in its implementation and operation remains to be seen. In our view, given the regulatory approach and enforcement architecture through which it is intended to operate, these principles are likely to remain primarily aspirational.

What we do expect to see, however, is the emergence of flourishing new markets for service-providers across Europe offering various "solutions" intended to satisfy the Act's requirements (including the need for high-risk AI system providers and deployers to establish and maintain a suitable "risk management system" and "quality management system" that purport to comply with the technical standards developed by CEN/CENELEC). Accordingly, we believe it is likely that existing legal frameworks – such as the General Data Protection Regulation, the EU Charter of Fundamental Rights, and the European Convention on Human Rights – will prove even more important and instrumental in seeking to address the erosion and interference with foundational European values as ever more tasks are increasingly delegated to AI systems.

[124] See also Karen Yeung, "Responsibility and AI – A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility within a Human Rights Framework," Council of Europe, 2019, DGI (2019)05; Nathalie A. Smuha, "Beyond the individual: governing AI's societal harm," *Internet Policy Review*, 10(3), 2021.