European Journal of Risk Regulation (2026), 1–25
doi:10.1017/err.2025.10077

ARTICLE

Risk, Reasonableness and Residual Harm under the EU AI Act: A Conceptual Framework for Proportional Ex-Ante Controls

Fabian Teichmann

International Relations, The London School of Economics and Political Science, London, UK
Email: teichmann@post.harvard.edu

Abstract

The EU Artificial Intelligence Act (AI Act) establishes a novel risk-based regulatory model for AI systems, categorising uses into four tiers: unacceptable (prohibited), high-risk (tightly regulated), limited-risk (transparency obligations), and minimal-risk (largely unregulated). This article develops a rigorous conceptual framework to analyse the Act's logic of risk, reasonableness, and residual harm. It explains how the principles of precaution and proportionality shape the AI Act's *ex ante* controls, requiring providers to anticipate *reasonably foreseeable misuse* and apply measures that reflect the *state of the art.*[1] We propose criteria for calibrating key requirements (data governance, transparency, human oversight, robustness or cybersecurity) to the severity and uncertainty of risks, drawing on risk-regulation theory (e.g., Baldwin and Black's responsive regulation and Sunstein's cost-benefit rationality). The analysis also situates the EU approach within a comparative context, noting alignments and divergences with US and OECD AI frameworks – for example, the EU's precautionary bans on biometric mass surveillance contrast with the US reliance on voluntary risk management guidelines. Specific high-impact use cases (biometric identification in public spaces, AI in critical infrastructure) illustrate how risk severity triggers stricter controls. The article concludes by discussing policy implications for implementation, including the role of harmonised standards and presumptions of conformity, the interface with parallel cybersecurity regimes (NIS2, DORA) as "risk multipliers," and the need for further guidance and delegated acts to ensure that the AI Act's proportional safeguards remain effective in the face of technological change.

Keywords: EU artificial intelligence act; harmonised standards; proportionality principle; residual risk management; risk-based regulation

I. Introduction

The European Union's Artificial Intelligence (AI) Act (Regulation (EU) 2024/1689) embodies a *risk-based approach* to AI governance, representing the first comprehensive attempt worldwide to impose *ex ante* controls tailored to AI risk levels. Under this approach, described by the European Commission as a "pyramid of criticality," AI systems are classified into four tiers: minimal risk, limited risk, high risk, and unacceptable risk. Each tier attracts a proportionate level of regulatory requirements, ranging from essentially no

[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) 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [2024] OJ L, 2024/1689.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

obligations for minimal-risk uses to outright *prohibitions* on systems deemed to pose "unacceptable" risks to safety or fundamental values. This model seeks to reconcile the *precautionary principle* (preventing serious or irreparable harm under uncertainty) with *innovation-friendly proportionality* (avoiding undue burdens on low-risk AI).[2] The result is a delicate legal balancing act that requires a clear conceptual framework to guide interpretation and implementation.[3]

Thesis and Contribution: This article argues that the AI Act's risk-based logic can be illuminated by integrating concepts from *risk regulation theory* (particularly reasonableness and residual risk) into the legal analysis. We propose a framework that links the Act's core notions – *"reasonably foreseeable misuse," "state-of-the-art,"* and *"acceptable residual risk"* – to the established principles of *precaution*, *proportionality* and the *ALARP* ("as low as reasonably practicable") approach in safety regulation.[4] We thus clarify how *ex ante* obligations should be calibrated to the *severity* and *uncertainty* of AI risks. The article's contribution is threefold: (1) mapping the AI Act's graded risk taxonomy to underlying regulatory principles and ethics (including the protection of fundamental rights as a risk metric)[5]; (2) articulating "conceptual tests" for what constitutes appropriate and state-of-the-art risk mitigation under the Act's requirements; and (3) offering criteria to adjust key control measures (data governance, transparency, human oversight, technical robustness and cybersecurity) in proportion to various risk scenarios.

We adopt a doctrinal and analytical approach that is grounded in EU law and risk regulation scholarship. To ensure its relevance to the enacted provisions, the analysis remains focused on the *final text of the AI Act adopted in 2024*, rather than earlier drafts. We also incorporate a comparative perspective by comparing the EU's approach to emerging frameworks in the United States and the OECD. For example, while the EU mandates explicit risk tiers and binding duties, the US has thus far favoured voluntary guidance, such as NIST's AI Risk Management Framework, and the OECD has developed classification tools and principles rather than enforceable rules.[6] These comparisons help reveal the distinctive features of the EU model (such as its emphasis on fundamental rights risks and use of *presumption of conformity* via standards). Finally, the discussion highlights specific high-stakes use cases – notably biometric identification and AI in critical infrastructure – to illustrate how the abstract principles play out in concrete sectors. Biometric remote identification, for instance, is treated so severely in EU law that it straddles the line between high-risk and prohibited use, triggering stringent controls or bans. Meanwhile, AI applications in energy grids and transportation infrastructure are classified as high-risk due to their potential "serious disruption of critical infrastructure" with life-safety implications, warranting robust oversight and cybersecurity measures.[7]

Roadmap: The article is structured as follows. The following section maps the conceptualisation an operationalisation of "risk" in the AI Act, linking it to the EU principles of precaution, proportionality and the ALARP concept from safety regulation. We then examine the normative yardsticks of the Act: how one determines what controls are "appropriate" and in line with the "state of the art," and how the notion of *residual risk*

[2] Regulation (EU) 2024/1689 (n 1).

[3] Wilson Sonsini, "10 things you should know about the EU Artificial Intelligence Act" (*Wilson Sonsini*, April 2024), <https://www.wsgr.com/a/web/qrkz1SnNzWw6nk7B3oAyDa/10-things-you-should-know-about-the-eu-artificial-intelligence-act_v2.pdf> accessed 31 October 2025.

[4] Gabriella Maselli, Maria Macchiaroli, Antonio Nesticò, "ALARP criteria to estimate acceptability and tolerability thresholds of the investment risk" (2021) 11 Applied Sciences, https://doi.org/10.3390/app11199086.

[5] Carsten Orwat and others, "Normative Challenges of Risk Regulation of Artificial Intelligence" (2024) 18 NanoEthics.

[6] NIST, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)" (*NIST*, January 2023) <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf> accessed 29 October 2025.

[7] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

is managed. Next, we discuss the role of harmonised European standards and the presumption of conformity in providing practical benchmarks for compliance – essentially, how standards translate state-of-the-art techniques into accepted practice. We then explore the interface between AI regulation and cybersecurity law (particularly the NIS2 Directive and DORA Regulation), arguing that cybersecurity threats act as *risk multipliers* for AI systems and that a coordinated regulatory approach is needed. In the penultimate section, we outline the policy implications, including the need for forthcoming guidance (e.g., the Commission's 2026 guidelines on high-risk use cases) and the strategic use of delegated acts to refine the risk framework over time.[8] The Conclusion synthesises how the EU's risk-based AI regime can achieve *proportional yet precautionary* oversight – keeping AI innovation "trusted throughout the world" while also ensuring it is "*as safe as reasonably practicable.*"

II. Mapping "risk" in the AI act – from precaution to proportionality (and ALARP)

Risk tiers and the precautionary logic: The AI Act explicitly takes a "*clearly defined risk-based approach*" to regulation.[9] This approach is codified in a four-level hierarchy of AI systems: unacceptable risk, high risk, limited risk and minimal (or low) risk. Each category triggers a different regulatory treatment:

- Unacceptable-risk AI refers to systems the use of which is *prohibited outright* (Article 5) due to their *unacceptable threats* to safety or fundamental values. This includes AI that deploys subliminal techniques to materially distort behaviour and cause harm, exploits the vulnerabilities of vulnerable groups, implements social scoring by governments, certain kinds of predictive policing, indiscriminate facial recognition database scraping, emotion recognition in sensitive contexts, and (with narrow exceptions) real-time remote biometric identification in public by law enforcement. The EU determined that these uses contravene core values (i.e., human dignity, privacy and non-discrimination) and present harms that "*cannot be tolerated*" – essentially a de facto application of the precautionary principle. Under that principle, if an action may cause severe harm to the public and scientific certainty is lacking, it should be avoided.[10] Banning systems such as social credit scoring and indiscriminate biometric surveillance reflects a precautionary stance: the *potential* for grave societal harm or rights violations is deemed so high that the law prohibits the practice *before* such harm materialises. Notably, even the exceptions for remote biometric ID (limited to serious crimes, threats of terrorism or finding missing persons) are tightly bound by necessity and proportionality tests, reflecting the EU's cautious approach. In effect, the AI Act's top tier is a legal articulation of ALARP's top zone: risks that are intolerable and must be ruled out because no reasonable mitigation could reduce them to an acceptable level.[11]

[8] Ibid.

[9] Tobias Mahler, "Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal" (2022) Nordic Yearbook of Law and Informatics; Regulation (EU) 2024/1689 (n 1).

[10] Bronwyn Howell, "The precautionary principle, safety regulation, and AI: This time, it really is different" (*AEI*, 4 September 2024) <https://www.aei.org/research-products/report/the-precautionary-principle-safety-regulation-and-ai-this-time-it-really-is-different> accessed 30 October 2025; European Union, "Precautionary principle" (European Union, n.d.) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:precautionary_principle> accessed 30 October 2025.

[11] Rostam J. Neuwirth, and Sara Migliorini, "Unacceptable risk in Human-AI collaboration: Legal prohibitions in light of cognition, trust and harm" <https://ceur-ws.org/Vol-3547/paper4.pdf> accessed 31 October 2025.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

- High-risk AI refers to systems that are *allowed* on the market but are subject to extensive *ex ante* requirements and oversight. Article 6 and Annexe III define which AI uses are high-risk, generally covering those that pose a significant threat to health, safety or fundamental rights in specific domains. High-risk categories include (among others) AI used in the safety components of products (such as medical devices or machinery control systems), in critical infrastructure management (energy, transport and the water supply), in education or vocational training (e.g., for grading exams), in employment and worker management (e.g., CV-sifting algorithms), in essential private or public services (credit scoring, welfare eligibility), in law enforcement (e.g., forensic AI, certain types of data analysis for policing), in migration/asylum control (e.g., risk assessment tools), and in the administration of justice (AI assistance of judicial decisions), as well as certain especially sensitive biometric systems such as fingerprint or face recognition (non-real-time). These systems are considered to risk causing a *"significant harmful impact"* if they malfunction or are misused, yet unlike the banned uses, their benefits are considered to outweigh the risks if proper controls are in place. The regulatory logic here is one of proportional risk mitigation rather than prohibition. The Act mandates a suite of risk management, documentation and quality measures (detailed in Chapter III, Section 2) to ensure that high-risk AI systems "do not pose unacceptable risks." This corresponds to the middle band of the ALARP model – risks that are tolerable only if mitigated to the extent that they are *as low as reasonably practicable.* Indeed, Article 9(5) requires providers to reduce *residual risk* to an acceptable level through design or safeguards. If a high-risk system's risk cannot be mitigated to an acceptable residual level even when using state-of-the-art measures, it should not be deployed. Thus, the high-risk tier operationalises ALARP: manufacturers must eliminate or minimise risk until further risk reduction would be grossly disproportionate to the benefits.[12] Notably, the Act does *not* guarantee that compliance equals zero risk; instead, it aims for risks to be *"minimised and acceptable"* – a pragmatic standard that acknowledges that some residual risk will remain, which regulators and society deem tolerable given the system's utility.[13]
- Limited-risk AI refers to systems that are not high-risk but still merit *some transparency* obligations. The Act imposes several requirements in such cases (Article 52). For example, AI systems that interact with humans (such as chatbots or virtual assistants) must disclose that the user is conversing with a machine, "unless this is obvious from the context." Similarly, AI-generated deepfake content must be labelled as such by the creator to prevent deception (with exceptions for authorised research, art or security uses). Particular emotion recognition or biometric categorisation systems (if not outright prohibited) also carry transparency duties to inform affected people. These limited-risk measures reflect proportionality in its most literal sense: lighter rules for lesser risks. They echo the idea in risk governance that regulatory intervention should be commensurate with the risk level. Rather than burdening all AI with heavy compliance, the Act only requires limited-risk systems to implement simple, reasonable precautions, such as informing users – a response calibrated to the lower likelihood or impact of harm. In ALARP parlance, these would be in the broadly acceptable or low-risk region, where only minimal oversight (if any) is needed. The

[12] Gabriella Maselli, Maria Macchiaroli and Antonio Nesticò, "ALARP criteria to estimate acceptability and tolerability thresholds of the investment risk" (2021) 11 Applied Sciences, 9086 https://doi.org/10.3390/app11199086.

[13] Delaram Golpayegani, Harshvardhan J. Pandit, Dave Lewis, "To be high-risk, or not to be – Semantic specifications and implications of the AI Act's high-risk AI applications and harmonized standards" (2023) Proceedings of the 2023 ACM Conference of Fairness, Accountability, and Transparency, 905. https://doi.org/10.1145/3593013.3594050.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

Act's drafters intentionally avoided expanding high-risk obligations to these uses to "not overly stifle innovation" for benign applications. Instead, a light-touch transparency rule addresses specific concerns (such as human dignity or autonomy in human–AI interactions) without implying that the systems pose significant dangers.[14]

- Minimal-risk AI encompasses all other AI systems that do not fall into the above categories. This is essentially an open category comprising the vast majority of AI applications (e.g., AI in video games, entertainment, most business analytics, spam filters, etc.), which are regarded as posing only negligible or routine risks. Such systems face *no mandatory requirements* under the AI Act (other than existing laws). The Commission explicitly noted that most AI systems *"present minimal or no risk"* and thus remain unregulated by the Act. This broad safe harbour is crucial for proportionality and feasibility: regulators did not intend to micromanage low-risk AI, both to conserve enforcement resources and to avoid unnecessary constraints on innovation. The AI Act does encourage voluntary codes of conduct for providers of non-high-risk AI (to promote trustworthy AI principles), but these are non-binding.[15] In risk theory terms, minimal-risk uses occupy the broadly acceptable zone where regulatory costs would far outweigh any benefit, so the rational approach is to leave them be. This aligns with Cass Sunstein's view that regulators should *"not aim low"* with heavy rules for trivial risks, but rather focus their efforts where they matter most.[16] It also resonates with Black and Baldwin's principle of "responsive regulation" – that is, directing regulatory attention in proportion to risk and adjusting stringency as needed.

By mapping the AI Act's provisions to this risk hierarchy, we see how *precautionary logic* and *proportionality* jointly inform its design. On the one hand, the presence of outright bans (unacceptable AI) demonstrates a willingness to take precautionary action in the face of uncertain but deeply problematic AI threats – a stance consistent with the EU's regulatory tradition in health and environmental domains. On the other hand, the stratified obligations for high-, limited-, and minimal-risk AI reflect a nuanced proportional approach, avoiding a one-size-fits-all regime. The Act thus attempts to embody what risk scholars call *"smart regulation"*: stringent measures targeted at the highest risks, and flexible or no measures for lower risks.[17] This approach intends to satisfy the EU Treaties' requirement that legislation respect the principle of proportionality (not exceeding what is necessary to achieve objectives). Indeed, Recital 14 of the Act affirms that the regulation *"does not go beyond what is necessary"* in light of its aims, and Recital 18 underscores the alignment of the rules with international efforts while maintaining flexibility for rapid technological developments.[18]

Risk, rights and residual harm: It is important to recognise that the concept of "risk" in the AI Act is not purely about statistical or safety concerns; it is entwined with fundamental rights and ethical values. This expands the traditional scope of risk regulation. Unlike, say, chemical safety law, which quantifies risk as the probability of physical harm, the AI Act treats *intangible harms* – such as the erosion of privacy, discrimination or loss of autonomy – as risks to be regulated. Scholars have observed that AI's risks often threaten "fundamental societal values" and can be challenging to quantify.

[14] Christine Saloustrou, "The legal framework for low-and minimum-risk AI systems under the AI act" (*SSRN*, 28 March 2025) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5192868> accessed 30 October 2025.

[15] Regulation (EU) 2024/1689 (n 1).

[16] Carsten Orwat and others, "Normative Challenges of Risk Regulation of Artificial Intelligence" (2024) 18 NanoEthics.

[17] Carsten Orwat and others, 'Normative Challenges of Risk Regulation of Artificial Intelligence' (2024) 18 NanoEthics.

[18] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

The Act embraces this by explicitly including risks posed to fundamental rights in the high-risk definition and in the required risk assessment (providers must assess risks to rights such as non-discrimination and privacy, etc., per Article 9(2)). This introduces *normative ambiguity* in risk trade-offs. How much bias or privacy intrusion is an "acceptable" residual risk? The Act does not provide numeric thresholds for such harms, instead relying on broad principles (e.g., the elimination of unlawful bias) and on the "*state-of-the-art*" of mitigation (discussed in the subsequent section). It assigns certain *normative choices* to different actors: developers must make design choices to minimise value harms; standardisation bodies and regulators will flesh out metrics and acceptable levels over time. This dynamic is recognised as a key challenge – balancing AI's benefits against rights-based risks requires political and ethical judgment, not only technical risk analysis.[19] The Act's solution is to embed fundamental rights considerations into the risk management framework (risks to rights are treated on par with safety risks) and to use external reference points (such as the EU Charter of Fundamental Rights and jurisprudence) to guide what is reasonable. For example, if an AI system's residual bias produces systematic discrimination, that would likely be deemed *unacceptable residual risk* because it conflicts with non-discrimination rights. The concept of *residual harm* that remains after mitigation is therefore evaluated not only in quantitative terms but against qualitative legal standards (e.g., is the remaining bias unlawful or significant enough to harm protected groups?). The Act's requirement that residual risk be "*judged to be acceptable*" (Article 9(5)) implies an expectation of reasoned justification: providers should be able to explain why any remaining risks are minor or outweighed by safeguards and benefits, aligning with societal expectations and the "ALARP" principle in safety engineering – i.e., that further risk reduction would require measures grossly disproportionate to any additional risk reduction.[20]

In summary, the AI Act's mapping of AI systems to risk tiers combines the precautionary principle (for truly unacceptable AI practices) with a proportional, risk-based allocation of compliance duties (for the spectrum from high to minimal risk). It seeks an "*as low as reasonably practicable*" level of residual risk for high-risk AI, explicitly requiring providers to eliminate or mitigate risks "as far as technically feasible" and then implement controls or warnings for risks that cannot be designed out.[21] This hierarchy and methodology resonate strongly with classic risk regulation frameworks, such as those found in EU product safety law. In this context, dangerous products are banned or heavily regulated, while low-risk products are left mainly to the market. However, the novelty here is the extension of that logic to the ethical and social risks of AI and the embedding of *reasonableness* (what is foreseeable, what is state-of-the-art, and what is acceptable) as guiding standards. The following sections explore those concepts of reasonableness in more detail, examining how regulators and companies determine what measures are appropriate and sufficient under the Act.

III. Conceptual tests for "appropriate" controls and the "state of the art" – shaping obligations & residual risk

A central challenge in operationalising the AI Act's requirements is understanding the *open-textured* terms that describe the quality and extent of required measures. Throughout

[19] Carsten Orwat and others, "Normative Challenges of Risk Regulation of Artificial Intelligence" (2024) 18 NanoEthics.

[20] Regulation (EU) 2024/1689 (n 1); Gabriella Maselli, Maria Macchiaroli and Antonio Nesticò, "ALARP criteria to estimate acceptability and tolerability thresholds of the investment risk" (2021) 11 Applied Sciences, https://doi.org/10.3390/app11199086.

[21] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

the obligations for high-risk AI, the law employs phrases such as "*appropriate* ... measures," "*state of the art*," and "*reasonably foreseeable misuse*," requiring judgments about what residual risk is *"acceptable."* These act as conceptual tests of reasonableness – essentially, criteria that tether the strict letter of the law to evolving technology and context. This section unpacks these concepts:

The phrase "reasonably foreseeable misuse" extends the scope of risk anticipation: High-risk AI providers are obliged not only to consider the AI system's intended use, but also its *misuse* if such misuse is reasonably foreseeable (Article 9(2)(b)).[22] The Act defines "reasonably foreseeable misuse" as the use of an AI system *"in a way that is not in accordance with its intended purpose, but which may result from reasonably foreseeable human behaviour or interaction with other systems."* In plain terms, developers must ask: *How might this AI be mistakenly or wrongfully used in practice, given typical user tendencies or integration into other tools?* This concept is borrowed from product safety law (e.g., machinery safety directives have long required accounting for misuse that is foreseeable by the manufacturer). It closes the gap whereby a provider might claim a system is safe "if used exactly as instructed," while real-world users might predictably do otherwise. Under the AI Act, providers of high-risk AI systems (say, an AI medical diagnosis aid) must anticipate, for example, that a busy doctor might over-rely on the AI's suggestions (a form of misuse through *automation bias*), or that an operator might use the AI on a category of patients outside the intended scope. If such scenarios are reasonably foreseeable, the provider should address them through design safeguards, usage limits, or at least warnings provided in the instructions. Indeed, Article 13(3) requires that the instructions for use include not only the intended purpose and performance information, but also *"any known or foreseeable circumstances ... of reasonably foreseeable misuse, which may lead to risks,"* so that deployers are alerted to those potential hazards. Human oversight measures likewise must aim to *"prevent or minimise the risks ... when [the AI] is used ... under conditions of reasonably foreseeable misuse, in particular where such risks persist despite other requirements"* (Article 14(2)). This ensures a *layered safety net*: even if all design measures are in place, if misuse could still cause harm (e.g., an operator ignoring a safety alert), there should be oversight or a fallback to mitigate it.[23]

The test of foreseeability is fundamentally about reasonableness and knowledge. It asks what a reasonable provider, *armed with current understanding of human behaviour and the deployment context*, ought to predict. This shifts the onus onto AI developers to research and understand the user environment and likely failure modes. For instance, when deploying an AI system in a critical infrastructure setting, it is reasonably foreseeable that human operators might misuse it under pressure or that malicious actors might attempt to manipulate it; thus, the provider should anticipate these risks and build in preventive features (or at least disclose them). By tying obligations to what is "reasonably foreseeable," the Act aligns with the concept of *fault* in tort law, where the foreseeability of harm is a key factor in determining negligence. However, the AI Act's regime is ex ante and does not wait for harm to occur and be litigated; it proactively requires the producer to think like a *"reasonable risk manager."* This aspect could be seen as importing a *negligence standard* into regulatory compliance: failing to address a foreseeable misuse risk could render the AI system non-compliant (and possibly defective under product liability rules).

Notably, the Act clarifies in Recital 65 that identifying mitigation measures for foreseeable misuse *"should not require specific additional training of the AI by the provider,"*

[22] Karen Yeung, "Can risk to fundamental rights arising from AI systems be 'managed' alongside health and safety risks? Implementing Article 9 of the EU AI Act" (*SSRN*, 8 October 2025) ≤https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5560783> accessed 31 October 2025.

[23] Ibid.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

although providers are encouraged to consider it.[24] In other words, a provider is not strictly obliged to retrain an AI model to handle every misuse scenario (which could be an endless task), but they should *document and warn* about such scenarios. The balance here is pragmatic: ensure awareness and *some* mitigation (such as user training or input checks), without mandating impossible perfection. Still, the inclusion of foreseeable misuse underscores that residual risk is partly a function of user interaction. The Act thereby pushes the frontier of manufacturer responsibility closer to the user domain than in many traditional products, reflecting AI's dynamic nature and the fact that human–AI interaction can itself create new risks (e.g., overconfidence in AI recommendations).

"Appropriate" measures refer to proportionality and context-sensitivity: The term "appropriate" recurs in the AI Act's requirement clauses – e.g., *"most appropriate risk-mitigation measures"*, *"appropriate type and degree of transparency"*, *"appropriate human-machine interface"* for oversight, and *"technical solutions… appropriate to the relevant circumstances and the risks"* for cybersecurity. This language injects flexibility, requiring measures to be suited to the specific AI system's context and risk level. It also invokes a proportionality test internally: what is appropriate to address a minor risk might be insufficient for a major risk. For example, Article 14(3) on human oversight says that the oversight measures shall be *"commensurate with the risks, level of autonomy and context of use."* Thus, a high-risk AI with greater autonomy or impact (say an AI triaging emergency patients) demands more stringent oversight mechanisms (perhaps real-time human intervention capabilities or multiple human verifications, etc.), whereas a lower-impact high-risk AI (maybe a resume-screening tool) might justify simpler oversight (periodic audits, override option). The Act explicitly imposes additional oversight requirements for certain use cases: e.g., biometric identification systems in law enforcement must have at least two human operators verify an identification before action is taken, unless deemed disproportionate by law for specific sectors.[25] This illustrates the *calibration of "appropriate" measures to risk severity.*

What counts as appropriate is also linked with the evolving "state of the art" (a concept we address next). Recital 64 states that measures adopted by providers to comply with requirements *"should take into account the generally acknowledged state of the art … be proportionate and effective to meet the objectives"* of the Act. In effect, appropriateness has two dimensions: *effectiveness* (does the measure actually mitigate the risk in light of current tech knowledge?) and *proportionality* (is the burden of the measure justified by the level of risk reduction achieved?). The ALARP principle provides a guide: a measure is appropriate unless its cost or impact is grossly disproportionate to the risk reduction achieved.[26] For instance, if a slight software tweak can prevent a serious failure mode, it is appropriate (and expected) to implement it; but if addressing a very marginal risk would require an enormous expense or fundamentally alter the system's utility, it might be beyond "reasonably practicable" and thus not mandated. Article 9(4) captures this balance, stating that risk mitigation measures should *"achieve an appropriate balance in implementing the measures to fulfill those requirements,"* giving due consideration to combined effects.[27] This suggests that mitigating one risk (e.g., bias) may sometimes affect another aspect (e.g., accuracy), so the provider must balance and find an optimal solution that

[24] Regulation (EU) 2024/1689 (n 1); Yannick Caballero Cuevas, "The principles for interpreting the requirements for high-risk AI systems' (*Centre for Banking and Financial Law*, 7 April 2025) <https://cdbf.ch/en/1406/> accessed 30 October 2025.

[25] Regulation (EU) 2024/1689 (n 1).

[26] Gabriella Maselli, Maria Macchiaroli and Antonio Nesticò, "ALARP criteria to estimate acceptability and tolerability thresholds of the investment risk" (2021) 11 Applied Sciences, https://doi.org/10.3390/app11199086.

[27] Henry L. Fraser, Jose-Miguel Bello y Villarino, "Acceptable risks in Europe's proposed AI Act: Reasonableness and other principles for deciding how much risk management is enough" (2023) 15 European Journal of Risk Regulation 43, https://doi.org/10.1017/err.2023.57.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

appropriately reduces overall risk without overcorrecting one dimension at the expense of another.[28] The concept of "appropriate balance" again echoes reasonableness, requiring a thoughtful, justified trade-off rather than mechanical compliance.

Legally, using a term like "appropriate" means regulators and courts will assess compliance in a *context-specific* manner. They will likely ask: given the nature of this AI system and the foreseeable risks, did the provider implement measures that a competent actor in the field would consider suitable and sufficient? This invites input from standards and best practices: if standards exist (formal or de facto) that describe appropriate safeguards, following them would indicate compliance.[29] Conversely, if a provider did the bare minimum while their peers typically do more to ensure safety, their measures might be judged as inappropriate. This standard also evolves: what is appropriate today may not be in a few years' time if technology advances. That ties directly to the *state of the art*.

"State of the art" represents a dynamic benchmark for safety and mitigation: The state of the art is a critical reference point in the AI Act.[30] It appears in multiple places: Article 8(1) states that high-risk AI must comply with requirements *"taking into account the generally acknowledged state of the art on AI and AI-related technologies"*; Recital 64 and 65 emphasise measures and risk management should implemented in light of the state of the art; and even the definition of harmonised standards notes they are *expected to reflect the state of the art*.[31] *State-of-the-art* essentially means the current level of technological development and knowledge that is *reasonably available* to practitioners. It is a well-established concept in product safety law (e.g., the EU Machinery Directive utilises it) and in standards. The Commission's 2022 "Blue Guide" on product regulation clarifies that references to state of the art serve to provide flexibility for technical progress. In other words, the law does not freeze requirements at a fixed technological solution, but instead expects them to evolve as technology improves.[32]

Practically, for an AI provider, complying with the state of the art means using up-to-date methods and tools for risk mitigation. If safer algorithms, robust training techniques, or improved testing practices have become established in the industry, a provider should incorporate them (or have a compelling reason why not). For example, if the state of the art in adversarial defence (to prevent "tricking" an AI with malicious inputs) advances, future AI systems will be expected to include those improved defences. This creates a dynamic regulatory requirement that can rise over time. It also aligns with the idea of continuous improvement. As new threats emerge or new solutions are invented, the acceptable "residual risk" threshold effectively tightens because more can be done to mitigate risk. Indeed, Recital 65 envisages that the risk management system is *iterative and regularly updated* to remain effective, and explicitly that decisions should be justified in light of the state of the art.[33] If a provider ignores an obvious state-of-the-art solution and a problem occurs, they would likely be found non-compliant (and possibly liable under liability regimes). This incentivises *innovation in safety*: companies have a reason to keep up with the latest research on AI safety and fairness, as the regulatory standard is not static.

From a risk-regulation theory perspective, referencing the *state of the art* is a way to deal with uncertainty and change – hallmarks of AI technology. It is analogous to the FDA's

[28] Regulation (EU) 2024/1689 (n 1).

[29] Regulation (EU) 2024/1689 (n 1).

[30] Claudio Novelli, Federico Casolari, Antonino Rotolo, Mariarosaria Taddeo and Luciano Floridi, "AI risk assessment: A scenario-based, proportional methodology for the AI Act" (2024) Digital Society 3, https://doi.org/10.1007/s44206-024-00095-1.

[31] Regulation (EU) 2024/1689 (n 1).

[32] Yuan Shi, "'State-of-the-art' in new EU medicine device regulations: a review of its development in medical device law, the interpretations from stakeholders, impacts, and possible solutions for implementation" (M.D.R.A thesis, University of Bonn 2022).

[33] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

approach to medical devices, where manufacturers must follow current technical standards or demonstrate equivalent safety. It prevents the regulation from either lagging behind (by locking in old standards) or from becoming quickly obsolete. However, it also means there is some ambiguity: who determines what the state of the art is? In practice, harmonised standards will play a significant role (we cover this in the following section). Standards bodies (CEN/CENELEC, ISO/IEC, etc.) convene experts to codify the state of the art in normative documents. Compliance with such standards then gives a "*presumption of conformity*" – effectively a safe harbour showing that the state-of-the-art requirements have been met. In the absence of a standard, providers might rely on consensus in scientific literature or guidelines from authoritative bodies (such as the High-Level Expert Group's ethics guidelines or technical benchmarks). The AI Act also allows the Commission to publish *common specifications* if standards are delayed, which would similarly capture state-of-the-art practices in a more agile manner.[34]

One consequence of the state-of-the-art clause is that residual risk must shrink over time. For example, today's state-of-the-art technology may not entirely eliminate bias in AI, so some bias may be tolerable as a residual risk if the provider has done everything technically feasible. But if, in a few years, new techniques emerge to virtually eliminate certain biases, continuing to have that bias would no longer be an "acceptable residual risk" because the state-of-the-art offers mitigation. This ties into ALARP: as the frontier of "reasonably practicable" moves with technology, the law demands more risk reduction. On the other hand, the state-of-the-art also guards against impractical demands: regulators should not expect what is not (yet) possible. Article 15's requirement for robustness and accuracy is tempered by "appropriate" levels, which implies not perfection, but as good as currently achievable. Article 15(4) calls for systems that learn post-market to be developed to *"eliminate or reduce as far as possible"* the risk of feedback loops causing performance degradation.[35] *"As far as possible"* is synonymous with ALARP's mandate of reducing risk to the furthest extent reasonable – essentially another way of saying "in line with technical possibility."

In summary, the state-of-the-art functions as a *dynamic yardstick for reasonableness*. It ensures that the obligation of "appropriate measures" is anchored to what competent peers would do at present (and not something outdated or purely theoretical). By doing so, it reconciles *innovation and safety*: it does not freeze design, but pushes industry to collectively raise the bar. Legally, it will likely be interpreted such that *compliance is a moving target* – firms will document how their processes reflect current best practices (e.g., utilising the latest secure architectures, bias mitigation libraries, rigorous testing protocols, such as adversarial penetration testing, etc.). Regulators and auditors may consult technical experts or reference documents to determine whether a technology was state-of-the-art at the time of its deployment.[36]

To concretise these ideas, consider a specific high-risk use case: AI in critical infrastructure (e.g., an AI system balancing electricity grid load). The foreseeable misuses could include an operator pushing the AI beyond recommended settings or a hacker injecting false data (cyber misuse). Appropriate measures might involve built-in limits, alarms, and robust authentication. State-of-the-art technology might involve utilising redundant fail-safes and the latest anomaly detection algorithms to detect grid disruptions. The provider must design the AI to be resilient (Article 15(5) actually mandates resilience *against unauthorised third-party manipulation*), using state-of-the-art

[34] Regulation (EU) 2024/1689 (n 1).

[35] Regulation (EU) 2024/1689 (n 1).

[36] Yuan Shi, '"State-of-the-art" in new EU medicine device regulations: a review of its development in medical device law, the interpretations from stakeholders, impacts, and possible solutions for implementation' (M.D.R.A thesis, University of Bonn 2022).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

cybersecurity practices. If, despite all measures, a residual risk remains (for example, a very rare condition that could still cause an outage), the question to ask is whether it is acceptable. If similar systems globally have managed to address that risk, then *not* addressing it likely fails the state-of-the-art test. If no one knows how to solve it yet, the provider might argue the risk is ALARP – minimised to the extent current tech allows – and thus acceptable until new solutions emerge. The continuous monitoring and post-market obligations (such as quality management and incident reporting) further ensure that once state-of-the-art improvements or new evidence of risk arise, the system can be updated or even withdrawn if needed.[37]

Interim conclusion: "Reasonably foreseeable misuse," "appropriate" measures, and "state of the art" collectively operationalise the Act's risk-based philosophy through a *lens of reasonableness and adaptability*. They require AI providers to *think ahead* (about misuse and worst-case scenarios), to *tailor their precautions* to the context and severity of risk, and to *constantly benchmark* their solutions against the cutting edge of science and technology. These concepts are underpinned by both legal tradition (foreseeability and reasonableness echo tort law's duty of care) and engineering approaches (state-of-the-art and ALARP come from safety engineering standards).[38] By embedding them, the AI Act aims to ensure that its high-level mandates (such as "ensure accuracy, robustness, cybersecurity") are neither trivially fulfilled nor impossibly strict but are instead met in a sensible, evidence-based manner that evolves over time.

The following section will examine how harmonised standards and conformity assessment mechanisms help give concrete shape to these flexible concepts, effectively translating "state of the art" and "appropriate measures" into checklists and technical specifications that providers can implement and authorities can verify.

IV. The role of harmonised standards and the presumption of conformity

Implementing the AI Act's requirements in practice will heavily rely on technical standards and conformity assessment procedures. As is common in EU product regulation, the Act follows the New Legislative Framework approach, setting *essential requirements* in legislation to be supplemented by harmonised European standards that provide detailed technical means for compliance. Providers who follow these standards enjoy a *presumption of conformity* with the corresponding legal requirements.[39] This section explores how standards and the presumption mechanism operate under the AI Act and why they are pivotal for proportional ex ante control.[40]

Harmonised standards as benchmarks: Harmonised standards are specifications, typically developed by European standardisation organisations (CEN, CENELEC, ETSI) upon request from the European Commission, then cited in the Official Journal of the EU. When cited, they become recognised ways of meeting the law's requirements. The AI Act explicitly defines "harmonised standard" by referring to Regulation (EU) No 1025/2012. Crucially, Recital 118 of the AI Act states: *"Compliance with harmonised standards . . . which are normally expected to reflect the state of the art, should be a means for providers to demonstrate conformity with the requirements."* In the absence of standards, the Commission can adopt common specifications (essentially technical rules set via an implementing act) as a

[37] Regulation (EU) 2024/1689 (n 1).

[38] Gabriella Maselli, Maria Macchiaroli and Antonio Nesticò, "ALARP criteria to estimate acceptability and tolerability thresholds of the investment risk" (2021) 11 Applied Sciences, https://doi.org/10.3390/app11199086.

[39] Regulation (EU) 2024/1689 (n 1).

[40] Lilian Edwards, "The EU AI Act: a summary of its significance and scope" (*Ada Lovelace Institute*, April 2022) <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf> accessed 31 October 2025.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

fallback.[41] This framework ensures that the abstract obligations (e.g., ensuring data are free of errors, ensuring transparency, etc.) can be translated into testable criteria.[42]

For example, consider the requirement that training data be "relevant, representative, free of errors *as far as possible*, and complete" (Article 10(3)). By itself, this is qualitative. A harmonised standard (perhaps developed by ISO/IEC JTC 1/SC 42 on AI or the CEN-CENELEC Focus Group on AI) could specify quantitative thresholds or procedures – e.g., how to statistically test representativity, how to document data provenance, and metrics for data accuracy. If a provider applies such a standard, they can presume compliance with Article 10's data requirements.[43] Similarly, for robustness and cybersecurity (Article 15), a standard might detail methods for penetration testing AI models and encryption protocols to ensure model integrity, among other measures. The provider following those is presumed to meet the robustness/security obligations. This provides legal certainty: companies prefer clear checklists over fuzzy terms, and standards supply that. It also fosters consistency across the single market, as all players meeting the same standards should be accepted as compliant.

Presumption of conformity: The AI Act explicitly includes a presumption of conformity in certain contexts. Notably, Article 42(1) establishes that if a high-risk AI system is *trained and tested on data* that reflect the specific context of use (i.e., avoiding geographical or demographic bias), it is presumed to comply with the data quality requirement in Article 10(4). This is a specific presumption built into the law, likely to incentivise using localised data to improve accuracy.[44] Article 42(2) provides another presumption: high-risk AI systems certified under a recognised EU cybersecurity scheme (per the Cybersecurity Act 2019/881) are presumed to meet the AI Act's cybersecurity requirements to the extent the cert covers them. This is an interesting cross-regime link (more on cybersecurity interplay later). Beyond these, the general mechanism is that compliance with harmonised standards referenced under Article 40 yields a presumption of conformity with the corresponding requirements. In practice, when the standards are cited, Article 43(1) allows providers to undergo a simpler conformity assessment (self-assessment for certain systems) if they use harmonised standards. If they do *not* use standards, a more rigorous evaluation (often involving a third-party notified body) is required, reflecting that they must then prove, via "other means," that they have met the essential requirements.[45]

The *presumption of conformity* is thus a powerful compliance tool as it shifts the burden of proof. A provider using standards can assume compliance (unless evidence to the contrary emerges), whereas one deviating from standards must substantiate equivalence. Given this, most companies will likely follow standards once available, as it de-risks their compliance process.

Standards embodying the state of the art: The relationship between standards and the *state of the art* is bidirectional. On the one hand, as Recital 118 notes, standards are *expected to reflect the state of the art*. Standardisation committees gather experts and, through consensus, incorporate the latest knowledge (subject to periodic revision). On the other hand, the law's reference to the state of the art means that if standards become outdated (lagging behind technical advances), it could be recognised. The Blue Guide 2022 indicates that if it becomes evident that a harmonised standard no longer represents the state of the

[41] Regulation (EU) 2024/1689 (n 1).

[42] Julio Hernandez, Delaram Golpayegani, and Dave Lewis, "An open knowledge graph-based approach for mapping concepts and requirements between the EU AI act and international standards" <https://arxiv.org/abs/2408.11925> accessed 31 October 2025.

[43] Regulation (EU) 2024/1689 (n 1).

[44] Philipp Hacker, "A legal framework for AI trained data – from first principles to the Artificial Intelligence Act" (2021) 13 Law, Innovation and Technology 257, https://doi.org/10.1080/17579961.2021.1977219.

[45] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

art, it may need updating or even be withdrawn.[46] The AI Act even addresses this: Article 40(2) allows the Commission to publish in OJ the references of harmonised standards that meet its request, implying that if a standard is incomplete or raises fundamental rights concerns, the Commission can choose not to cite it fully. There is also a procedure (outlined in the EU Standardisation Regulation) for the Commission to *object* to a harmonised standard that does not meet the legal requirements. Recital 118 specifically mentions that if standards do not sufficiently address fundamental rights, common specifications could be used. This highlights a vital governance aspect: ensuring that technical standards for AI robustly cover not just technical performance but also ethical and rights aspects (such as bias, transparency, etc.), which are harder to standardise. The Act essentially states: we will use standards, but not blindly – they must truly meet the requirements; otherwise, regulators will intervene.[47]

Conformity assessment and enforcement: The Act provides different *conformity assessment routes* depending on the type of AI system. Many stand-alone high-risk AI systems can undergo internal control assessment (Annexe VI) if standards are applied. However, if no standards exist or the provider chooses not to use them, a third-party evaluation (Annexe VII, involving a notified body) is mandatory. Some systems – specifically those that are safety components of products regulated by other EU laws (such as AI in medical devices, cars, or machinery) – follow the *sectoral conformity assessment* (often involving notified bodies), with the AI requirements integrated into that process. In all cases, the result is an EU Declaration of Conformity and a CE marking on the AI system (or product containing it).[48] The CE mark signals that the system is compliant with the AI Act (and any other applicable regs) and can circulate in the single market.[49]

The presumption of conformity via standards greatly streamlines these conformity assessments – essentially, the notified body (or the provider in self-assessment) can verify compliance if standard clauses are followed, rather than re-deriving test methods themselves. This is vital for efficiency, given the potentially huge range of AI applications. It also provides a common *language* for compliance: for example, a standard might specify that a risk management file should include XYZ analyses, that accuracy be measured in a certain way, and so on, so both companies and regulators know what to expect.[50]

From a risk regulation theory perspective, one can view harmonised standards as a way to implement Sunstein's idea of cost-benefit-sensitive regulation: they implicitly balance effectiveness and practicality, often through industry input. They can also embed "best practices," which represent a consensus on ALARP for specific risks (e.g., the acceptable false-positive rate for an AI in cancer screening might be agreed upon in a standard). By following those, a provider demonstrates they have taken *all reasonable measures* recognised by experts. As one commentary notes, standards serve as *"epistemic authority"* translating broad principles into concrete norms – in AI's case, they will likely cover technical metrics, documentation templates, risk management steps, validation procedures, and so forth.[51]

[46] European Union, "Information and Notices – C247/1" (2022) 65 Official Journal of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:2022:247:FULL> accessed 30 October 2025.

[47] Regulation (EU) 2024/1689 (n 1).

[48] Regulation (EU) 2024/1689 (n 1).

[49] Theodoras Karathanasis, "Guidance on classification and conformity assessments for high-risk AI systems under EU AI Act" (*AI-Regulation*, 22 February 2023) <https://ai-regulation.com/wp-content/uploads/2023/02/Article-AI-Classification.pdf> accessed 31 October 2025.

[50] Jonas Schuett, "Risk Management in the Artificial Intelligence Act" (2024) 15 European Journal of Risk Regulation, 367. https://doi.org/10.1017/err.2023.1.

[51] Regine Paul, "European artificial intelligence «trusted throughout the world": Risk-based regulation and the fashioning of a competitive common AI market' (2024) 18 John Wiley & Sons, 1065. https://doi.org/10.1111/rego.12563.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

Limitations and importance of oversight: However, reliance on standards also has limitations. There is a risk that standard-setting may lag behind fast-moving AI innovation or that industry-driven standards may dilute stringent requirements. The Act's governance addresses this partially: it establishes a European AI Office and an EU AI Board (comprising national regulators and the Commission) to oversee implementation and, if necessary, recommend issuing common specifications or updating requirements. The Board can flag standards that are lacking or inadequate. Additionally, under Article 40, if no standards are available or a gap exists, the Commission *must* step in with common specifications for certain requirements to ensure that regulatory expectations do not wait indefinitely on standardisation.[52]

The Act also mandates that technical documentation (Annexe IV) and post-market monitoring records be kept, detailing whether standards were applied and how. This provides transparency. If a problem arises with an AI system that was "compliant," investigators will examine whether the relevant standards were actually sufficient or if there was a shortcoming. Article 71 even outlines a procedure for the Commission to implement if it finds that a given AI system compliant with standards still presents a risk – it can require corrective actions or withdraw the presumption if needed.[53] This is akin to the safeguard clauses in product directives.[54]

In summary, harmonised standards are the linchpin for bridging high-level legal principles and on-the-ground implementation of AI risk controls. They carry the state-of-the-art into compliance by codifying it, and the presumption of conformity rewards those who follow them by simplifying market access. The interplay of standards and law in the AI Act exemplifies the EU's "regulated self-regulation" approach: industry (with other stakeholders) drafts technical rules, which gain legal status through Commission oversight, thus blending flexibility with accountability.[55] For AI developers, engaging in the standards process (or at least following the outcomes) will be crucial, as these standards will effectively determine what design/testing processes are considered *reasonable* and *sufficient*. For regulators, ensuring that standards are robust and up to date will be key to the Act's success, as heavy reliance on standards means that any gaps directly translate to compliance gaps.

Having examined how the AI Act's framework will be operationalised via standards and conformity assessment, we now turn to an important complementary aspect of risk management: cybersecurity. AI systems, especially high-risk ones, need to be not only designed safely under normal conditions, but also secured against malicious attacks or misuse. Weak cybersecurity can turn an AI system from safe to dangerous in an instant, effectively multiplying risks. The following section examines how the AI Act's requirements interact with broader EU cybersecurity laws, such as NIS2 and DORA, and why treating cybersecurity as an integral part of AI regulation is necessary for addressing residual risk.

V. Interface with cybersecurity (NIS2, DORA) as risk multipliers

In the modern threat landscape, cybersecurity vulnerabilities in AI systems can dramatically amplify the risks those systems pose. An AI that is well-behaved during testing can be driven to erratic, harmful behaviour if an attacker manipulates its inputs (adversarial examples), training data (data poisoning), or underlying infrastructure.

[52] Regulation (EU) 2024/1689 (n 1).

[53] Regulation (EU) 2024/1689 (n 1).

[54] Soler Garrido and others, "JRC Technical Report – Analysis of the preliminary AI standardization work plan in support of the AI Act" (*European Commission*, 2023) https://doi.org/10.2760/5847.

[55] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

Conversely, AI systems deployed for critical functions become attractive targets for cyberattacks aimed at causing maximum disruption. The EU legislator recognised this interdependence, making *robustness and cybersecurity* one of the mandatory requirement pillars for high-risk AI (Article 15).[56] This section explores how the AI Act's built-in cybersecurity obligations interact with horizontal cybersecurity regimes, notably the NIS2 Directive (Directive (EU) 2022/2555 on network and information security of critical entities) and the DORA Regulation (Regulation (EU) 2022/2554 on digital operational resilience in finance). These frameworks are not AI-specific but cover many AI deployers, and they reinforce the AI Act's aims by addressing the *operational environment and response* aspects of risk.

Cybersecurity requirements within the AI Act: Article 15 of the AI Act requires high-risk AI systems to be designed and developed to achieve an *"appropriate"* level of cybersecurity, in addition to accuracy and robustness.[57] Specifically, the AI must be as resilient as possible to errors *and* unauthorised third-party attempts to alter its performance. The Act calls for technical solutions "appropriate to the circumstances and risks" to secure the AI. Furthermore, it explicitly lists AI-specific vulnerabilities to address, *"where appropriate,"* including data poisoning attacks (tampering with training data), model poisoning (compromising pre-trained models), adversarial examples (inputs designed to fool the AI), as well as attacks on confidentiality or exploiting model flaws.[58] By enumerating these, the Act essentially requires providers to anticipate and protect against known forms of AI attacks. For example, a provider of an image recognition AI should implement defences against adversarial images if those could cause safety incidents. A provider of a machine learning model that continually learns online should employ measures to detect anomalous data inputs (to prevent poisoning).[59]

These obligations are tied back to the *state of the art*: since AI security is an active research area, what counts as "appropriate technical solutions" will evolve. Initially, adherence to existing standards such as ISO/IEC 27001 (information security management) or following guidance from ENISA (the EU Cybersecurity Agency) on securing AI might suffice. In fact, the AI Act leverages the EU Cybersecurity Act's voluntary certification schemes: if an AI receives a certificate under a relevant *European cybersecurity certification scheme*, it is presumed to comply with Article 15's requirements.[60] The Commission and ENISA are likely to develop such schemes (there is discussion of a scheme for AI that possibly builds upon the AI Act's requirements).[61]

The underlying rationale is clear: even the best algorithm can yield catastrophic outcomes if hacked or misused. For instance, an AI system for controlling traffic lights poses a high risk; while its inherent design might be safe, if an attacker penetrates it and switches lights erroneously, lives will be endangered. Thus, the *residual risk* of an AI system cannot be fully understood without considering cybersecurity. Weak cybersecurity effectively raises the *true* risk level of an AI deployment by exposing it to intentional misuse beyond "foreseeable misuse" by typical users. Therefore, treating cybersecurity as

[56] Regulation (EU) 2024/1689 (n 1).

[57] Henrik Nolte, Miriam Rateike and Michèle Finck, "Robustness and Cybersecurity in the EU Artificial Intelligence Act" (2025) Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency 283, https://doi.org/10.1145/3715275.3732020.

[58] Regulation (EU) 2024/1689 (n 1).

[59] Enisa, "Multilayer framework for good cybersecurity practices for AI" (*Enisa*, June 2023) <https://www.enisa.europa.eu/sites/default/files/publications/Multilayer%20Framework%20for%20Good%20Cybersecurity%20Practices%20for%20AI.pdf> accessed 30 October 2025.

[60] Regulation (EU) 2024/1689 (n 1).

[61] Elias R. Sandström, "Exploring the gap in security standardization for Artificial Intelligence: A qualitative analysis of expert opinions in cybersecurity" (*Stockholm University*, 2024) <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1955652&dswid=2231> accessed 31 October 2025.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

a first-class requirement (not an afterthought) in the AI Act integrates the "security-by-design" principle into AI governance. It ensures that providers assess not only accidental failures but also *malicious scenarios*.

NIS2 Directive – securing critical sectors: The NIS2 Directive (in force since 2023) aims for a high common level of cybersecurity across essential entities in the EU (covering sectors such as energy, transport, health, water, digital infrastructure, banking, and the public sector, etc.).[62] It requires these entities to implement *appropriate cybersecurity risk management measures and incident reporting*.[63] Many high-risk AI systems will be deployed by entities under NIS2's scope. For example, a hospital using an AI diagnostic tool, an energy grid operator using AI for load management, or a railway company using AI for traffic control would all likely qualify as essential entities under NIS2. NIS2 obliges such organisations to assess risks to their network and information systems and take measures, including access controls, business continuity, encryption, etc., as specified by an EU implementing act (Commission IR 2024/2690). ENISA recently provided technical guidance for NIS2 implementation, mapping security measures for digital infrastructure and service providers.[64]

How does this interplay with the AI Act? Essentially, while the AI Act imposes security-by-design on the *product (i.e., the AI system)*, NIS2 imposes security-by-design on the *operator*. For instance, the AI Act would ensure that an AI traffic control system is engineered to resist attacks (with robust authentication, failsafes, etc.), whereas NIS2 would ensure the transport operator has an overall security program, performs regular risk assessments, keeps software up to date, trains staff, and so forth.[65] NIS2 does not explicitly mention AI, but its broad requirements certainly encompass any digital technologies used by the entity, including AI. One explicit difference: NIS2 focuses on continuity and availability of services, whereas the AI Act focuses on preventing AI from causing harm. However, they converge in practice: a cyber incident that takes down an AI-driven service can cause indirect harm; conversely, compromising an AI can lead to service outages or dangerous incidents.

A concrete example: An electricity grid operator utilises an AI system for load balancing (this AI is high-risk, per Annexe III). The AI Act ensures that the system is built securely (perhaps the model has adversarial training to avoid manipulation and is tested against known attacks). NIS2 ensures the operator's entire ICT environment is secure – including the servers that host the AI, the communications, and the authentication of personnel, among other aspects. If a hacker still succeeds and the AI misbehaves, NIS2 also mandates *incident response*: the operator must report incidents to the authorities within a set time and have plans in place to mitigate the impact. This incident reporting could also inform AI regulation enforcement – if a security breach led to the AI causing damage, it might trigger a review of compliance with Article 15 or prompt revisions to standards.[66]

[62] Niels Vandezande, "Cybersecurity in the EU: How the NIS2-directive stack up against its predecessor" (2024) 52 Computer Law & Security Review, https://doi.org/10.1016/j.clsr.2023.105890.

[63] Maria Chiara Meneghetti, and Giulia Zappaterra, "EU: ENISA guidelines on compliance with NIS 2 directive published" (*JD Supra*, 14 August 2025) <https://www.jdsupra.com/legalnews/eu-enisa-guidelines-on-compliance-with-6647884> accessed 30 October 2025.

[64] Enisa, "Technical implementation guidance" (*Enisa*, June 2025) <https://www.enisa.europa.eu/sites/default/files/2025-06/ENISA_Technical_implementation_guidance_on_cybersecurity_risk_management_measures_version_1.0.pdf accessed> 30 October 2025.

[65] Hyperproof, "Understanding the relationship between NIS2 and the Eu Cyber Resilience Act" (*hyperproof*, n.d.) <https://hyperproof.io/understanding-the-relationship-between-nis2-and-the-eu-cyber-resilience-act> accessed 30 October 2025.

[66] Fabian Teichmann, "Cybersecurity of critical infrastructure in Europe: the NIS2 directive in focus" (2025) 6 International Cybersecurity Law Review 207, https://doi.org/10.1365/s43439-025-00154-4.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

DORA – resilience in the financial sector: The Digital Operational Resilience Act (DORA) came into effect in January 2025 and applies to banks, insurers, investment firms, and other financial entities.[67] DORA consolidates and elevates ICT risk management in finance, requiring firms to have robust ICT risk management frameworks, conduct regular testing (including threat-led penetration testing for significant institutions), manage ICT third-party risks, and report major incidents.[68] Financial services are increasingly using AI for credit scoring, fraud detection, algorithmic trading, etc. Some of those AI systems might be classified as high-risk under the AI Act (credit scoring AI is high-risk in Annexe III; algorithmic trading AI might be indirectly covered under critical infrastructure or not, but it is critical for financial stability either way). The DORA will ensure that any AI used by, e.g., a bank, is encompassed in its overall risk controls. A bank must inventorise its ICT assets (which include AI models), ensure they have continuity plans if the AI fails or is attacked, test them under cyber-stress scenarios, and ensure third-party providers (such as an AI vendor) meet security requirements.[69]

One relevant overlap is that DORA requires advanced digital operational resilience testing, which could include testing AI models under attack scenarios (adversarial ML).[70] If a financial firm's AI is critical, it might be expected to simulate attacks on it as part of a threat-led penetration test exercise.[71] If vulnerabilities are found, they must be patched. This complements the AI Act's design-time obligations with ongoing runtime vigilance. Also, the DORA's incident reporting means that if an AI-related outage or incident happens (e.g., a rogue AI trade causes a financial incident possibly due to a data integrity attack), it will be reported to regulators, who can then investigate both under the DORA and potentially under the AI Act if the attack occurred due to non-compliance.

Notably, the AI Act explicitly acknowledges interdependence: Article 42(2) incentivises the use of *cybersecurity certification* under Regulation 2019/881. While not the same as NIS2 or the DORA, it is part of the EU's cybersecurity ecosystem (the Cybersecurity Act allows for voluntary EU-wide certifications of ICT products). If an AI system receives such a certificate (for example, if an EU cybersecurity scheme for AI is created and the system is certified at, say, a "high" assurance level), that *in itself* gives a legal presumption of meeting the AI Act's security requirements.[72] This is a strong alignment of incentives: it encourages AI makers to go through cyber certification (which NIS2 or DORA might indirectly push their customers to demand anyway).

Cyber risks as risk multipliers: We term cybersecurity a *risk multiplier* because a breach or attack can transform a low- or moderate-risk AI scenario into a high-risk or catastrophic

[67] Hal S. Scott, "The E.U.'s Digital Operational Resilience Act: Cloud Services & Financial Companies" <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3904113> accessed 31 October 2025.

[68] Georgia M. P. Karakasilioti, "Supporting the Digital Operational Resilience of the Financial Sector: The EU's DORA Digital Operational Resilience Act" (MSc Dissertation, University of Piraeus 2024, <https://dione.lib.unipi.gr/xmlui/bitstream/handle/unipi/16273/DORA%20-%20MTE2109%20Karakasilioti.pdf?sequence=1> accessed 31 October 2025.

[69] Anna Ribeiro, 'ENISA releases cyber stress testing handbook to boost critical infrastructure resilience under NIS2 directive' (*Industrial Cyber*, 19 May 2025) <https://industrialcyber.co/reports/enisa-releases-cyber-stress-testing-handbook-to-boost-critical-infrastructure-resilience-under-nis2-directive> accessed 30 October 2025; Eiopa, "Digital Operational Resilience Act (DORA)" (*Eiopa*, n.d.) <https://www.eiopa.europa.eu/digital-operational-resilience-act-dora_en> accessed 30 October 2025.

[70] Dirk Clausmeier, "Regulation of the European Parliament and the Council on digital operational resilience for the financial sector (DORA)" (2023) 4 International Cybersecurity Law Review 79, https://doi.org/10.1365/s43439-022-00076-5.

[71] Lena Kontseva, "Digital Operational Testing (part of DORA) – What to expect?" (*Under Defense*, 17 July 2024) <https://underdefense.com/blog/digital-operational-resilience-testing-part-of-dora-what-to-expect> accessed 30 October 2025.

[72] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

one.[73] For example, an AI chatbot is typically limited-risk (it just needs transparency). But if someone injects malicious instructions and it starts giving dangerous advice or phishing, it suddenly poses security and safety risks. Or an AI driving system might usually be safe, but if hacked, it could cause accidents. Therefore, addressing cyber risk is essential to ensure the *true* risk of AI systems stays within the intended category. The regulators understand this: Recital 59 of the DORA notes that ICT risk can spill over across sectors, hence the need for harmonised resilience. Recital 9 of NIS2 emphasises the importance of *supply chain security* and dependencies – AI could be part of that supply chain. The AI Act's requirement for providers to have a *post-market monitoring system* (Article 61 and 72) will inevitably involve monitoring for new vulnerabilities or attacks, much like NIS2 requires monitoring for threats.[74] The AI Act also coordinates with the broader regime by establishing cooperation between the AI Board and the European Cybersecurity Board, perhaps, and by requiring consistency of any new delegated acts with other laws.[75]

In practice, compliance efforts will likely merge: organisations deploying high-risk AI will integrate AI Act compliance into their NIS2 or DORA compliance programs. For instance, under NIS2's risk management, an entity might include a check: "If we use AI, have we received an EU Declaration of Conformity from the provider? Are they certified? Are there known vulnerabilities in that AI?" Likewise, AI providers may advertise compliance with cybersecurity standards (e.g., ISO 27001, ETSI EN 303645 for IoT) and even obtain certification to meet client expectations under NIS2/DORA. The policy implication of this is that regulators should provide guidance on this intersection, perhaps via the AI Office and ENISA. ENISA could issue guidelines on how to secure AI (they already have some work on securing AI and tackling adversarial threats). The AI Act mandates the Commission to develop *"harmonised standards or common specifications"* for security, too, presumably in coordination with ENISA.[76]

A noteworthy challenge is incident handling: The AI Act itself does not create an incident notification duty for the AI regulator (except if the provider later learns of serious incidents; then they have some recall obligations). But NIS2 and DORA do have strict reporting timelines (e.g., report within 24 or 72 hours of major incidents).[77] A cyber incident involving an AI could fall under both, so ideally, the AI regulatory authorities should liaise with cybersecurity authorities. Perhaps in practice, a significant AI incident will prompt market surveillance authorities to coordinate with CSIRTs (Computer Security Incident Response Teams) under NIS2.

In conclusion, the AI Act's success in controlling risk hinges not only on AI-specific measures but also on secure deployment environments. By embedding cybersecurity by design and leveraging regimes such as NIS2 and DORA, the EU creates a multi-layered defence: the AI has to be built securely, and the users (operators) must run it securely. Cyberattacks are a form of *reasonably foreseeable misuse by third parties*, and the combined regulatory framework addresses it from different angles – preventive controls (as outlined in the AI Act and NIS2/DORA), and reactive controls (incident response and crisis management in NIS2/DORA). This holistic approach acknowledges that technology risk is

[73] Angela Gendron, "Cyber threats and multiplier effects: Canada at risk" (2013) 19 Canadian Foreign Policy Journal, https://doi.org/10.1080/11926422.2013.808578.

[74] Enisa, "Technical implementation guidance" (*Enisa*, June 2025) <https://www.enisa.europa.eu/sites/default/files/2025-06/ENISA_Technical_implementation_guidance_on_cybersecurity_risk_management_measures_version_1.0.pdf accessed> 30 October 2025.

[75] Regulation (EU) 2024/1689 (n 1).

[76] Enisa, "Multilayer framework for good cybersecurity practices for AI" (*Enisa*, June 2023) <https://www.enisa.europa.eu/sites/default/files/publications/Multilayer%20Framework%20for%20Good%20Cybersecurity%20Practices%20for%20AI.pdf> accessed 30 October 2025.

[77] George C. Gruia, "Enhancing cybersecurity resilience: An analysis of DORA and NIS2 in the EU digital economy" (2025) 10 Journal of Financial Studies 56.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

systemic in nature. The case of *critical infrastructure AI* is illustrative: a power grid AI failure due to hacking is both an AI safety issue and a cybersecurity issue. The regulatory net – comprising the AI Act and NIS2 – aims to catch it either way.

As we move forward, continuous updates to both AI security standards and cybersecurity policies will be necessary. The Commission may adopt a delegated act to mandate that specific AI categories also comply with future cybersecurity schemes (especially once a relevant ENISA scheme exists). There is also the upcoming Cyber Resilience Act (CRA) proposal, which would require cybersecurity for all products with digital elements. If passed, it will overlap with the AI Act for AI systems (since AI systems are software). Ensuring coherence between these laws will be crucial (the CRA proposal currently exempts products regulated by other sectoral laws to avoid duplication – presumably, the AI Act could be considered such a law if it covers similar requirements).[78]

Having covered the substantive regime of the AI Act and its interplay with cybersecurity and standards, we finally consider how this framework will be *fine-tuned and implemented* in practice. The following section addresses policy implications: what guidance, delegated legislation, and further measures are needed to operationalise the Act's proportional risk controls, and how the EU can ensure the regime stays up-to-date through soft law (e.g., guidelines, codes of conduct) and hard law (delegated/implementing acts, amendments).

VI. Policy implications for guidance and delegated acts

Designing the AI Act's risk-based framework in legislation is only the beginning. The real-world effectiveness and proportionality of the regime will depend on *how it is implemented and updated* over time. Several mechanisms are built into the Act for this: the issuance of guidelines, the development of standards (discussed earlier), and the adoption of delegated acts to adjust the scope and details in response to new information. There are also broader policy decisions regarding the support of compliance (through an AI Office, regulatory sandboxes, etc.) and aligning interpretations across Member States. This section highlights key implementation steps and their significance:

Commission guidelines (soft law) to clarify concepts and use-cases: The AI Act explicitly mandates the Commission to publish guidelines on the practical implementation of certain provisions.[79] For example, Article 6(5) (now Article 6(5) or Article 4(5) in final numbering) requires the Commission, by February 2026, to issue guidelines with *"a comprehensive list of practical examples of use cases of AI systems that are high-risk and not high-risk"*. This guidance will be extremely valuable for clarifying the grey zones. For instance, it might illustrate borderline cases in education or human resources: e.g., is a university admissions algorithm high-risk (likely yes, affecting access to education) vs. a tutoring app (likely no)? It might also clarify what kinds of "biometric identification" fall under the high-risk category beyond the obvious cases. The guidance cannot alter the law, but it can provide an interpretation, giving both providers and regulators a shared understanding. Crucially, it is to be developed in line with Article 96 (which likely involves consultation with the European AI Board and stakeholders).[80]

Such examples of guidance serve proportionality by ensuring the high-risk classification is applied consistently and only where intended. Without it, Member

[78] Hikvision, "What you need to know about NIS2, AI Act and CRA" (*Hikvision*, 4 July 2025) <https://www.hikvision.com/europe/newsroom/blog/what-you-need-to-know-about-nis2–ai-act-and-cra> accessed 30 October 2025.

[79] Claudio Novelli, Philipp Hacker, Jessica Morley, Jarle Trondal, and Luciano Floridi, "A robust governance for the AI Act: AI office, AI board, scientific panel. And national authorities" (2024) 16 European Journal of Risk Regulation 566, https://doi.org/10.1017/err.2024.57.

[80] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

State authorities might diverge – one might treat an AI as high-risk while another does not, causing fragmentation or over-regulation. The list of examples can also evolve; although it is not binding law, it will carry weight as a Commission publication (similar to how the GDPR had WP29/EDPB guidelines with examples to illustrate ambiguous terms like "legitimate interests").

Another area ripe for guidance is the interpretation of *"foreseeable misuse," "state of the art,"* and *"acceptable residual risk."* While these are defined or described in the recitals, practical guidance can be helpful. For instance, a guidance document might provide a methodology for conducting the risk management outlined in Article 9, including how to document foreseeable misuse scenarios and how to determine whether something is reasonably foreseeable (perhaps referencing standards such as ISO 31000 on risk management).[81] It could also tie into how to perform *risk-benefit analyses* for Article 7 (when assessing new high-risk uses to add or remove via delegated act, as the criteria in Annexe III and Article 7(2) lay out a quasi-risk–benefit test including benefits to individuals or society).[82]

Delegated acts adjusting scope (Annexe III) and detail (Annexes IV, V, etc.): The AI Act acknowledges that what is considered "high-risk" can change. Technology and its impacts evolve, and the law includes a mechanism for recalibration. Article 7 empowers the Commission to adopt delegated acts to *amend Annexe III* – the list of high-risk AI use-cases. The Commission can add new use cases or modify existing ones if certain conditions are met, or remove use cases that no longer pose a significant risk. These conditions ensure evidence-based changes: there must be *"concrete and reliable evidence,"* and the risk must be equivalent to those already listed to add, or evidence that a listed use is no longer high-risk to remove. Also, changes must not reduce the overall protection level.[83] This is effectively a precautionary safeguard: you cannot remove something from high-risk if doing so lowers protection, even if that system individually might appear safe, because it could set a precedent or introduce cumulative risks.

Annexe III criteria: The law provides detailed *criteria* in Annexe III (or Article 7(2) as above) for assessing whether an AI use is high-risk, including factors such as the number of people affected, the severity of harm, reversibility, the dependence of people on the outcome, power imbalances, etc. These read like a risk assessment framework for regulators. When, in the future, the Commission considers, for example, generative AI or large language models as potentially high-risk, it will apply these criteria. Indeed, the final Act introduced the concept of "general purpose AI with systemic risks" (GP AI models that may be designated as having a broad impact) along with obligations (Articles 51–5). The Act asks the Commission to potentially develop codes of practice or standards for these, and possibly intervene if systemic risk emerges from them.[84] This demonstrates foresight: earlier drafts did not explicitly cover general AI, but the final text does, primarily through soft measures and monitoring.

Other delegated acts: Article 8 allows delegated acts to update technical documentation requirements (Annexe IV) in line with technological progress. Article 40(6) and others permit the updating of standards references and common specifications if issues arise. Article 50 (transparency for certain AI) might also be updated. And crucially, Article 70+ likely empowers guidelines for *sandboxes*. The Act encourages the use of AI regulatory

[81] Henry L. Fraser, Jose-Miguel Bello y Villarino, "Where residual risks reside: a comparative approach to Art 9(4) of the European Union's proposed AI regulation" <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3960461> accessed 31 October 2025.

[82] Regulation (EU) 2024/1689 (n 1).

[83] Ibid.

[84] Regulation (EU) 2024/1689 (n 1).

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

sandboxes (Article 57) to foster experimentation under regulatory guidance.[85] These sandboxes can yield insights that might inform future delegated acts or standards, especially for innovative uses that do not fit neatly into current risk categories.

Ensuring uniform interpretation – the European AI Board and AI Office: The Act establishes a governance structure (Chapter V or VI). A European Artificial Intelligence Board is created to facilitate uniform administration (similar to the GDPR's EDPB).[86] It will include representatives of each national AI regulator and the Commission. The Board can issue opinions, recommendations and share best practices. For instance, if one authority encounters a new type of AI that might be high-risk, they can bring it to the Board to discuss a common approach (possibly recommending the Commission use Article 7 to update Annexe III). The Board can also help coordinate market surveillance, ensuring the consistency of enforcement against non-compliant AI (such as those that slipped through without conformity assessment or present serious risks).

The EU AI Office (essentially the Commission services acting as a central coordinator) performs tasks such as managing the EU database of high-risk AI systems (Article 60 requires providers to register their high-risk AI in an EU database before deployment, to increase transparency). This Office will also support the Board and could issue guidance or coordinate sandboxes. The interplay with other agencies (ENISA for cybersecurity, EDPS for fundamental rights in EU institutions' AI, etc.) is implied. The AI Office should ensure that delegated acts under the AI Act consider related law updates (e.g., if a new Cybersecurity Act scheme arises, to integrate it in Article 42's presumption list, which Article 42(2) already does for existing schemes).[87]

Guidance for specific sectors and use cases: The question of *sector-specific guidance* is crucial. While the AI Act is horizontal, many sectors (healthcare, transport, etc.) have their own regulators and guidelines. Aligning those with AI Act obligations will avoid duplication. For instance, the European Medicines Agency might issue guidance on the use of AI in drug development or clinical decision support, to complement the AI Act by addressing domain-specific risk considerations (like the reliability of AI in clinical trials). The AI Act explicitly amends some sectoral legislation in its annexe (it modifies certain references in product regulations to include AI aspects).[88] But further soft law will likely emerge; for example, an *ETHICS GUIDELINE* for the use of biometrics by police (to navigate between the outright ban of real-time ID except for exceptions, and the allowed uses under high-risk forensic scenarios). Europol or the European Data Protection Board might weigh in on how to apply Article 5 prohibitions in practice.

International and comparative alignment: The user specifically asked for a comparative perspective. From a policy perspective, the EU is likely to issue guidance that aligns the AI Act with the OECD AI Principles and any future *global AI governance frameworks*. For example, the Commission might clarify how compliance with the AI Act also ensures adherence to the OECD AI Recommendation principles (such as transparency and fairness), allowing firms to streamline their efforts.[89] There may also be *mutual recognition* issues: if a US company complies with the NIST AI RMF, how does that align with AI Act compliance? Guidance or bilateral agreements may address partial recognition of frameworks. The Act allows for the possibility of equivalence decisions (though not explicit, but, e.g., Article 5(h) on biometric ID exceptions states that if law enforcement cooperation with a third

[85] Ibid.

[86] Ibid.

[87] Regulation (EU) 2024/1689 (n 1).

[88] Ibid.

[89] David Krause, "The EU AI Act and the future of Ai governance: Implications for U.S. firms and policymakers" <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5181797> accessed 31 October 2025.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

country has adequate fundamental rights safeguards, it might be allowed, demonstrating some openness to foreign frameworks if they are protective).[90]

Supporting SMEs and innovation: The EU is keen not to smother AI startups. The Act already has SME considerations (e.g., simplified technical documentation forms for small enterprises). The Commission can provide further guidance or even tools, such as templates for risk management or an "AI Act compliance sandbox," where startups can test their systems with regulators' feedback without incurring enforcement penalties. Articles 55–6 encourage *codes of conduct* for non-high-risk AI and for general-purpose AI providers to comply with requirements voluntarily.[91] The Commission and AI Office will likely facilitate these codes (analogous to how it supported codes in the GDPR context).[92] Such codes can provide tailored guidance: e.g., a code for AI in healthcare might detail best practices that go beyond the AI Act's minimal requirements, or a code for general-purpose AI (foundation models) could list steps to ensure downstream compliance by users (such as providing documentation or allowing model monitoring). If the Board approves these codes, they could even lead to reduced regulatory scrutiny for adherents – an incentive to industry.

Continuous improvement and review: Finally, the Act will be reviewed periodically (likely a review clause ~ five years). The Commission will collect data from the national authorities and the database to examine how the risk-based approach is working. Are too many systems being incorrectly categorised? Are incidents occurring that suggest some "limited-risk" AI should be classified as "high-risk"? Are obligations like transparency actually effective? This empirical feedback loop is crucial. For instance, if manipulative AI practices (Article 5(a),(b)) proliferate despite the ban, perhaps enforcement needs to be tightened or clarified. If an entirely new technology (such as artificial general intelligence applications or advanced chatbots that have not been envisioned) arises, posing systemic risks, the Commission may need to adapt the Act via delegated acts or propose an amendment.

One emerging issue is AI liability. While this falls outside the Act's scope, the EU has proposed an AI Liability Directive in parallel. If either this or the revised Product Liability Directive is adopted, it will interplay: strong compliance with the AI Act might serve as evidence to rebut fault in liability cases. Guidance may clarify that connection, motivating companies to comply not just for regulatory approval but to shield themselves from lawsuits (e.g., "compliance with harmonised standards under the AI Act can demonstrate due care in liability claims" – a likely implicit but nevertheless important link).

In summary, the AI Act's *proportional ex ante controls* will only remain proportional and effective if guided by clear examples, kept current through delegated acts, and supported by a governance ecosystem that learns and adapts. The Commission's ability to swiftly update Annexe III (within defined bounds) is a significant tool for handling emerging risks without waiting for full legislative revision, which is necessary given the pace of AI innovation. Meanwhile, non-binding guidance and codes can refine the application of requirements in various contexts, preventing both over-compliance (undertaking unnecessary measures for low risks) and under-compliance (failing to meet obligations for high risks). The involvement of various EU agencies (ENISA on cybersecurity, EDPS on fundamental rights, etc.) via the AI Board will help ensure a coherent approach so that, for example, cybersecurity guidelines from ENISA and AI Act standards dovetail nicely.

[90] Regulation (EU) 2024/1689 (n 1).

[91] Regulation (EU) 2024/1689 (n 1).

[92] Tomasz Hollanek, Yulu Pi, Dorian Peter, Selen Yakar and Eleanor Drage, "The EU AI Act in development practice: A pro-justice approach" ≤https://arxiv.org/abs/2504.20075> accessed 31 October 2025.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

VII. Conclusion

The EU's Artificial Intelligence Act pioneers a *risk-based and proportionate* model of AI regulation that seeks to protect health, safety and fundamental rights ex ante, without stifling beneficial innovation. This article has constructed a conceptual framework that elucidates how the Act's logic of "risk, reasonableness, and residual harm" operates. The AI Act effectively creates a regulatory pyramid, prohibiting uses of AI deemed *unacceptably risky* to core values, imposing stringent requirements on *high-risk systems* to pre-emptively mitigate significant risks, applying transparency obligations to *limited-risk tools* to address lesser concerns, and leaving *minimal-risk applications* free for innovation. This stratification embodies the principles of precaution and ALARP: eliminate intolerable risks, and reduce other risks to as low as reasonably practicable through appropriate safeguards.[93]

Central to this regime are the notions of "reasonably foreseeable misuse" and "state of the art," which infuse a *dynamic reasonableness* standard into AI governance. Providers of high-risk AI must adopt a forward-looking mindset, anticipating how their systems might be misused or fail in real-world conditions and implementing preventive or protective measures accordingly. They are expected to continuously update risk controls as technology advances, aligning with the state of the art in science and engineering.[94] What is considered an "acceptable" residual risk is thus not static: it narrows as better techniques for eliminating risks emerge. This ensures that the regulation remains *proportionate* to current capabilities and threats, neither demanding the impossible nor permitting the negligent. In effect, the AI Act leverages familiar regulatory concepts from product safety and tort law (foreseeability, reasonableness, best available techniques), embedding them in a novel AI context and thereby operationalising *ethical principles* (such as fairness, transparency, human oversight) in a manner that can be audited and enforced.[95]

We have also highlighted how harmonised standards and conformity assessment procedures will serve as the backbone for implementing these flexible requirements. Standards translate abstract obligations into concrete technical rules, providing clarity and facilitating the "presumption of conformity" for compliant AI systems. This mechanism is vital for maintaining a competitive common market for AI, as it offers legal certainty and a single compliance regime across Europe. At the same time, the EU's oversight of standards (with the possibility of common specifications and objections to inadequate standards) ensures that industry consensus does not dilute fundamental rights protections.[96] The interplay of regulation and standardisation in the AI Act exemplifies a *co-regulatory approach*: the law sets the goals and guardrails, while technical experts define the means – all under public supervision to safeguard the public interest.[97]

Importantly, our analysis of *specific use cases* illustrates the AI Act's nuanced approach. Biometric identification systems, especially those used for law enforcement, are recognised as posing such acute risks to civil liberties that the Act broadly prohibits them (real-time remote ID in public is banned, barring very narrow exceptions). Even where allowed, they are subject to strict oversight requirements (e.g., mandatory human

[93] Regulation (EU) 2024/1689 (n 1); Gabriella Maselli, Maria Macchiaroli, Antonio Nesticò, "ALARP criteria to estimate acceptability and tolerability thresholds of the investment risk" (2021) 11 Applied Sciences, https://doi.org/10.3390/app11199086.

[94] Yuan Shi, "'State-of-the-art' in new EU medicine device regulations: a review of its development in medical device law, the interpretations from stakeholders, impacts, and possible solutions for implementation" (M.D.R.A thesis, University of Bonn 2022).

[95] Regulation (EU) 2024/1689 (n 1).

[96] Ibid.

[97] Regine Paul, "European artificial intelligence 'trusted throughout the world': Risk-based regulation and the fashioning of a competitive common AI market" (2024) 18 John Wiley & Sons, 1065. https://doi.org/10.1111/rego.12563.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

verification to prevent false identifications). This reflects a precautionary stance toward a technology with high potential for misuse, aligning with Europe's historical aversion to mass surveillance. In contrast, AI in critical infrastructure (such as energy or transportation) is permitted but deemed high-risk; here, the emphasis is on robust design and failsafes to avert catastrophic failures. The Act requires providers to incorporate safety redundancies and resilience, and operators to maintain human oversight, ensuring that AI does not operate unchecked in safety-critical roles.[98] These case studies demonstrate how the severity and context of risk guide the stringency of controls, representing *proportionality in practice*.

From a comparative perspective, the EU's approach demonstrates both convergence with and divergence from other frameworks. It aligns with the OECD's emphasis on *risk-based and trustworthy AI* by institutionalising principles like transparency, accountability and fairness. However, it uniquely gives them binding force through *ex ante* rules. Compared to the more laissez-faire or piecemeal approach in the US, the EU model is more comprehensive and preventive – banning certain AI outright on ethical grounds (something the US has not done at the federal level), and requiring pre-market scrutiny akin to safety certification.[99] The US NIST AI Risk Management Framework shares similar values regarding context-based risk mitigation and is voluntary, but the EU Act makes such practices mandatory for high-stakes AI, enforced with penalties.[100] In essence, the EU is treating AI somewhat like the US treats medical devices or aviation – requiring evidence of safety before deployment. In contrast, the US currently relies on post-hoc accountability (liability, sectoral enforcement) and voluntary guidelines. Over time, these approaches may converge; indeed, the Act's risk categories and emphasis on "reasonable" measures could provide a blueprint for other jurisdictions. The EU's leadership in this area may well set de facto global standards: companies aiming to deploy globally might adopt the EU's requirements as their baseline ("the Brussels effect"). The Act also explicitly seeks international regulatory cooperation, ensuring, for example, that if AI is developed elsewhere but used in the EU, it must meet these standards, and promoting alignment with OECD and other international AI initiatives.[101]

Looking ahead, the successful implementation of the AI Act will require diligent governance and adaptability. Regulators must issue clear guidance (as mandated by 2026) to resolve ambiguities and support compliance, especially among SMEs and startups, as well as train enforcement personnel in this new domain. The European AI Board will need to coordinate a *learning regulatory community* across Member States to handle novel cases consistently. The Commission's use of delegated acts will be crucial for keeping the framework up to date: if emerging AI applications (such as advanced general AI or new biometric technologies) present new risks, the high-risk list can be expanded swiftly, rather than waiting years for a legislative revision.[102] Conversely, if certain requirements prove to be excessively burdensome without adding value, adjustments can be made (without lowering protections). This agility is a key feature of the Act's design, acknowledging that AI technology evolves rapidly and that regulation must keep pace.

Additionally, coherence with parallel EU initiatives, such as the Cyber Resilience Act (for IoT/software security), the Data Act, and sector-specific laws, will need to be managed so that developers face a unified set of expectations. The AI Act's synergy with NIS2 and

[98] Regulation (EU) 2024/1689 (n 1).

[99] Regine Paul, "European artificial intelligence 'trusted throughout the world': Risk-based regulation and the fashioning of a competitive common AI market" (2024) 18 John Wiley & Sons, 1065. https://doi.org/10.1111/rego.12563.

[100] NIST, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)" (*NIST*, January 2023) <https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf> accessed 29 October 2025.

[101] Regulation (EU) 2024/1689 (n 1).

[102] Ibid.

https://doi.org/10.1017/err.2025.10077 Published online by Cambridge University Press

the DORA regarding cybersecurity is a positive example: together, they create a defence-in-depth for critical AI systems.[103] Continued cross-sector collaboration (e.g., involving data protection authorities for AI's impacts on privacy, or competition authorities regarding AI transparency in platform algorithms) will reinforce the Act's objectives.

In conclusion, the EU's AI Act represents a milestone in tech regulation, one that attempts to tame a transformative technology through a calibrated, principle-driven approach. By focusing on risks and reasonable controls, it avoids one-size-fits-all rules and instead tailors obligations to where they matter most. Its success will depend on effective operationalisation by turning legal mandates into practical standards, audits and controls that AI developers and users can realistically implement. If done well, the Act will strike the intended balance of *minimising the residual harms of AI to society "as far as reasonably practicable" while still enabling beneficial AI innovation to flourish in the European Union and beyond*. This balancing act – between risk and innovation, precaution and proportionality – will undoubtedly be refined as we learn from implementation, but it provides a pioneering framework that other jurisdictions are already examining as they craft their own AI governance regimes in the quest for *trustworthy AI.*

[103] Ibid.

Cite this article: F Teichmann, "Risk, Reasonableness and Residual Harm under the EU AI Act: A Conceptual Framework for Proportional Ex-Ante Controls". *European Journal of Risk Regulation*. https://doi.org/10.1017/err.2025.10077