3/2025 anteprima

Rivista di diritto dei media

ISSN 2532-9146

The Brussels Sphinx's Riddle. What is a high-risk AI System?

Andrea Bertolini, Federica Fedorczyk, Marta Mariolina Mollicone, Guilherme Migliora*

Table of Contents

1. Introduction

The Sphinx of Thebes, a creature with the body of a lion, the wings of an eagle, and the face of a woman, is a figure from Greek mythology that has been long discussed by scholars. It is notable for its tendency to pose riddles to passersby, which were rich in symbolism. Those who were unable to answer correctly were killed and devoured by the Sphinx.
In the riddle, the Sphinx inquired as follows:

> «What is the creature that has only one voice, but walks on four legs in the morning, on two at noon, and on three in the evening?»[1]

The answer was a human. Indeed, the morning represents the childhood where a baby crawls on all fours, using four limbs, two hands and two

* Andrea Bertolini, Ph.D., LL.M. (Yale) is Associate Professor of Private Law at Scuola Superiore Sant'Anna (SSSA) and Director of The European Center of Excellence on The Regulation of Robotics And AI (EURA), supervised the structure of the entire article , and is primarily responsible for §§2.2-2.3, 4.2-4.3, 4.5, and 6; Federica Fedorczyk, Ph.D. is post-doctoral research fellow at Oxford University, is primarily responsible for §§2.2-3.4; Marta Mariolina Mollicone, Ph.D. is post-doctoral researcher at SSSA, and is primarily responsible for §§3.2 and 4.5-4.6; Guilherme Migliora, is Ph.D. candidate at SSSA and is primarily responsible for §§4.1-4.5. All websites were accessed no later than August 1st 2025.

[1] Apollodoro (Pseudo), *Biblioteca*, 3, 5, vv. 7-8.

knees; the noon is the adulthood. An adult walks on two legs, upright. The evening refers to old age since an elderly person uses a cane, so they "walk on three legs", two legs plus a cane.

Ultimately that of the Sphinx was a question of classification, not too different from that the European legislator – who assumes the role of the giant regulator of artificial intelligence (henceforth AI) – poses to those who approach the development of a complex AI system (henceforth AIS). Indeed, the riddle they face is the accurate classification of their system pursuant to art. 6 of the AI Act (henceforth AIA)[2], and if the Brussels' Sphynx where to phrase it, it could be as follows:

> "What is the AIS that falls under Annex III, and yet does not fall under the exemption of paragraph 3, or at least performs some profiling function, and is not declared by its provider not to be high-risk pursuant to paragraph 4 or, instead, is itself, or one of its safety components, subject to the product safety legislation enumerated by Annex I, and itself or its safety component, require a third party conformity assessment to be certified for a lawful distribution on the European market?"

Most likely the Sphinx of Thebes would herself be confused when formulating such an articulate question; maybe that of Brussels, instead, is better trained in both tongue- and mind-twisters and would thence find herself at ease laying the question down.

Unlike Oedipus, however, providers and deployers of AI might not find themselves triumphant, and while the repercussions of error are not as dire as in the case of death, they are nevertheless grave.

Indeed, on the one hand, both the classification of a system as high risk and its erroneous designation as such entail the possibility of being subject to the penalties provided for in art. 99 et seq. AIA – in the physiology of technological development and adoption – if a system is qualified as high-risk providers and deployers[3] will need to comply with an articulate series of requirements that apply across all phases of their lifecycle, from design and development to production, deployment, and even the so-called post-contractual phase, after the systems have been distributed and are operating on the market (art. 8-27 AIA). In such a perspective, the more conservative approach of considering a system as high-risk whenever in doubt, would represent a very burdensome strategy, and one that

[2] European Parliament and the Council, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024.

[3] Pursuant to art. 3, para. 3, AIA a provider «means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge». Pursuant to art. 3, para. 4, deployer «means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity».

could come at a relevant economic and strategic cost, in particular with respect to possible competition.

Similarly to the fated heir of the Labdacids, we will attempt to solve the riddle, by deconstructing the articulate definition of High-Risk AI System (henceforth h-AIS) as put forth by art. 6 AIA. The awaited prize is the dominion over the citadel of European AI regulation. Indeed, the AIA horizontal regulatory approach[4] (see §3.1), intended to be "future proof"[5], is partially tempered by a risk-based categorization[6], that causes the very definition of h-AIS to become of paramount importance, for it deter-

[4] The AIA is a horizontal legislation as it creates a supposedly far-reaching framework for AI regulation across various sectors and applications, while attempting to balance the numerous risks and benefits AI can provide. It has the specific objective, among others, to ensure that AIS placed, put into service and used in the Union market are safe and respectful of existing law and fundamental rights. To this end, the AIA explicitly recalls the Ethics Guidelines for Trustworthy AI (EGTAI) developed by the independent High-Level Expert Group on AI (AI HLEG) appointed by the European Commission in 2018. The EGTAI had identified seven non-binding ethical principles, to be taken into account to achieve the design of a "human-centric" AI (i) Human agency and oversight: AIS should be developed and used as a tool that serves people, respects human dignity and personal autonomy, and that is functioning in a way that can be appropriately controlled and overseen by humans; ii) Technical robustness and safety: AIS should be developed and used in a way that allows robustness in the case of problems and resilience against attempts to alter the use or performance of the AI system so as to allow unlawful use by third parties and minimize unintended harm; iv)Privacy and data governance: AIS should be developed and used in accordance with privacy and data protection rules, with processing data that meets high standards in terms of quality and integrity; iv)Transparency: AIS should be developed and used in in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AIS, as well as duly informing deployers of the capabilities and limitations of that AIS and affected persons about their rights; v)Diversity, non-discrimination and fairness: AIS should be developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law; vi)Social and environmental well-being: AIS should be developed and used in a sustainable and environmentally friendly manner as well as in a way to benefit all human beings, while monitoring and assessing the long-term impacts on the individual, society and democracy). Those same principles are today recalled, without prejudice to the legally binding requirement of the AIA itself (and of any relevant applicable Union law), as criteria that are to be considered in the design and use of AI models. However, the requirements laid down by the AIA are substantially more stringent and immediately applicable across all Member States. Indeed, as a regulation, the AIA has direct and immediate applicability across all EU member states, meaning it becomes binding in its entirety without the need for any additional national legislation or acts of implementation. This uniform application ensures consistency in the legal framework governing AI throughout the Union, reducing the risk of fragmentation and promoting harmonized standards. In such a perspective, it certainly reflects a commitment to uphold the digital sovereignty of the European Union. For the announcement of the AI Act refer to: European Commission, *European Artificial Intelligence Act comes into force*, 2021, in *ec.europa.eu*.

[5] European Commission, *Artificial Intelligence – Questions and Answers*, in *ec.europa.eu*.

[6] The pyramid of risks distinguishes unacceptable (art. 5 AIA), high (art. 6 AIA and following), limited (art. 50 AIA), and minimal or no risk, devising radically different regulatory frameworks. A separate section is dedicated to general-purpose AI (GPAI) models, which must comply with specific obligations related to intellectual property protection, transparency, and the mitigation of systemic risks (art. 53–55 AIA) that, however, entirely fall outside the scope of this study.

mines the scope of application of those provisions which also represent the most relevant core of the entire regulation.
From a practical perspective, this will provide providers with the necessary guidance to understand how their AIS may qualify, and subsequently what obligations they must comply with. From a theoretical perspective, it allows for a discussion of the effectiveness of the emergent framework, its ambiguities, and the interpretive challenges that fraught it. Ultimately, it will increase awareness about the complexity of the application of the AIA, and bout the need to collaborate with lawyers, already in the design phase, to ensure compliance with EU regulation.
Much like the original riddle – where the day is divided into three parts to represent the stages of human life: childhood, adulthood, and old age – this analysis is structured in three corresponding sections. "The Morning" sets the stage by outlining the background of the AIA and the principles on which is grounded, with particular attention to its risk-based and horizontal regulatory approaches. It also explores the Act's broad definition of AIS and its commitment to technological neutrality, using concrete examples to illustrate the artificial nature of risk categorization across varied applications.
"The Midday" turns to art. 6, para. 2, examining the criteria for what qualifies as "high-risk" AIS. This section unpacks the structure of art. 6 and its reference to Annex III, revealing inconsistencies in the classification of high-risk applications.
"The Evening" focuses into art. 6, para. 1, analysing the criteria used to identify high-risk systems and highlighting the conceptual and linguistic shortcomings of this provision, along with its implications for both providers and deployers. The paper concludes summarizing the key findings in tables for comparative insights.

2. The Morning – The Regulatory Framework

2.1. The AIA as a New Piece of Product Safety Legislation

On July 12, 2024, the European Union marked a significant milestone by approving the AIA, the world's first comprehensive regulation of artificial intelligence. It is important to note that while the AIA is formally a regulation, it represents a rather distinctive and unconventional type. Although it aims for maximum harmonization across the EU, it simultaneously leaves significant room for interpretation and flexibility. The Act calls for numerous delegated acts[7] and grants substantial discretion to member states,

[7] The power to adopt delegated acts is referred to in art. 6, para. 6 and 7, to add new conditions to those laid down in the art., by amending para. 3; art. 7, para. 1 and 3, allows for the Commission to add, modify or remove use-cases of h-AIS; art. 11, para. 3, refers to the possibility of amending Annex IV in regard to the necessary technical documentation; art. 43, para. 5 and 6 allows for the amendment of conformity assessment requirements in para. 1 and 2 of the same art. and Annexes VI and VII; art. 47, para. 5, allows the Commission to update the content of the EU declaration of conformity

allowing them considerable leeway in interpreting and adapting its provisions. For instance, Member Sates (henceforth MS) can, in exceptional cases (e.g. public security or protection of life and health of natural persons, environmental protection and the protection of key industrial and infrastructural assets), authorize the placing on the market or the putting into service of AI systems, which have not undergone a conformity assessment art. 46 and Recital 130 AIA). This not only creates a high degree of complexity but also raises fundamental questions about the very nature of this legislative instrument. One could even question whether the AIA truly functions as a regulation in the strict sense, or whether it shares characteristics more akin to a directive, ultimately blurring the traditional line between these two forms of EU legislation. Moreover, when we turn to its substance, the AIA is, at its core, primarily a piece of product safety legislation[8].

Despite being a complex and highly articulated discipline, the AIA can, with some approximation, be qualified as a new example of product safety legislation[9], primarily on the basis of its core obligations (see art. 5, 6 and 40-49) that establish a regulatory structure heavily reliant on risk regulation and certification through conformity assessment[10].

To better understand its structure, it is therefore necessary to clarify some fundamental characteristics of European product safety legislation, rooted in the so-called "New Approach" and the subsequent New Legislative Framework (henceforth NLF)[11], expressly referenced to by Recital 9 of the AIA.

European product safety legislation, comprising over 20 directives and regulations[12], only establishes essential safety requirements, specific to different classes of applications, products, devices and components, leaving it to manufacturers to determine how to meet them. An example of such freedom granted to manufacturers is provided by art. 10, para. 1, and Annex III of the forthcoming Machinery Regulation (MR)[13]. Manufacturers

set out in that Annex V; art. 51, para. 3, through which the Commission can amend the threshold for classifying a GPAI as a model with systemic risk; art. 52, para. 4, allows the Commission to update the criteria that justify a qualified alert from the scientific panel (Annex XIII); and art. 53, para. 5 and 6, allows for the Commission to detail measurement and calculation methodologies referred in the art. and to amend Annexes XI and XII.

[8] *Ex multis,* A. Iannuzzi, *Le fonti del diritto dell'Unione europea per la disciplina della società digitale*, in F. Pizzetti – M. Orofino – E. Longo – A. Iannuzzi – S. Calzolaio (eds.), *La regolazione europea della socità digitale*, Torino, 2024, 9 ff.

[9] M. Almada – N. Petit, *The EU AI Act: a medley of product safety and fundamental rights?*, in *Robert Schuman Centre for Advanced Studies Research Paper No. 2020/14*, 2023, 7 ff.

[10] P. Hacker, *AI Regulation in Europe: From the AI Act to Future Regulatory Challenges*, in *arXiv.org*, 2023, 1 ff.

[11] European Commission, *New legislative framework*, in *single-market-economy.ec.europa.eu*.

[12] The AIA, in its Annex 1, contains a list of Union harmonisation legislation based on the New Legislative Framework and a list of List of other Union harmonisation legislation.

[13] European Parliament and the Council, Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on Machinery and Repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive

must comply with the rules provided by the MR, such as that a machinery must «take the necessary protective measures in relation to risks that cannot be eliminated». However, the Regulation leaves room for the manufacturers to decide the means through which this requirement will be achieved. This gives the manufacturers more freedom to innovate.

The regulatory frameworks are then complemented by a very broad, articulate and distinct set of standards, often referred to by engineers as "norms"[14], thus potentially leading to ambiguity about their legal value. Indeed, while standards certainly play an essential role in promoting safety and interoperability, they are never binding. That entails that when developing any application requiring to be certified pursuant to one or more pieces of European product-safety legislation, the party is only bound to comply with the corresponding directive(s) and regulation(s), never with a technical standard, irrespective of the authority having adopted them.

Most technical standards are, in fact, developed at the international level by organizations such as ISO (International Organization for Standardization)[15], IEC (International Electrotechnical Commission)[16], and IEEE (Institute of Electrical and Electronics Engineers)[17]. At the European level, the so-called European Standardization Organizations (henceforth ESOs) – the European Committee for Standardization (CEN), the European Committee for Electrotechnical Standardization (CENELEC) and the European Telecommunications Standards Institute (ETSI) – operate, developing standards applicable within MS. However, relevant efforts are made to ensure alignment with the international framework to avoid market fragmentation, in particular through a series of agreements[18] that ensure coordination between the standardization efforts initiated with a ESO and their corresponding international equivalent[19].

73/361/EEC, OJ L 165, 29.6.2023.

[14] For instance, in Germany and Italy the terminology used in the standardization systems - *normen* in German, issued by the *Deutsches Institut für Normung* (DIN) and *norme* in Italian, issued by the Ente Nazionale Italiano di Normazione (UNI) – does not imply that these are legally binding per se. However, such norms often hold considerable authority in engineering and technical fields, and in some cases, they are formally integrated into legal regulations.

[15] ISO, *ISO - About ISO*, in *iso.org*.

[16] IEC, *IEC Webstore Homepage*, in *iec.ch*.

[17] IEEE Standards Association, *IEEE Standards Association*, in *standards.ieee*

[18] CEN has an agreement for technical co-operation with the International Organization for Standardization (ISO), called the Vienna Agreement, signed in 1991. CENELEC enjoys close cooperation with its international counterpart, the International Electrotechnical Commission (IEC), also thanks to the Frankfurt Agreement of 2016. Unlike CEN and CENELEC, ETSI does not rely on a single equivalent "agreement" akin to the Vienna or Frankfurt Agreements. Instead, ETSI's cooperation is structured through a variety of memorandum of understanding (MoUs) and partnership arrangements, which can be accessed at the following link portal.etsi.org.

[19] Under the Vienna Agreement, if CEN begins developing a standard and it is determined that ISO is already working on the same subject, CEN suspends its development activities. This coordination avoids duplication of efforts and promotes the adoption of international standards at the European level. For example, CEN will stop its drafting work and monitor ISO's progress, potentially adopting the final ISO standard as a European Standard (EN ISO). This procedure is detailed in CEN, *Guidelines for the*

The most distinctive trait of European standardization is instead represented by the so-called harmonized standards (henceforth hENs)[20], whose importance is expected to grow with the AIA, especially since the European Commission (henceforth EC) has begun to promote the adoption of many such standards to address evident gaps related to more advanced applications, falling within the scope of the AIA[21]. European standardization organizations, led by CEN and CENELEC, are in the process of drafting the necessary AI standards, following a request from the EC[22]. The main advantage ensured by this very special category of standards is that adherence to them ensures a rebuttable presumption of conformity with the corresponding piece of product safety legislation.
Overall, a regulatory technique, heavily reliant upon standards, may elicit more than a subtle criticism in light of the lack of democratic control and accountability – primarily – typical of transnational private regulators[23],

implementation of the Vienna Agreement, in *boss.cen.eu*.

[20] See European Parliament and the Council, Regulation (EU) No 1025/2012 of the European Parliament and of the Council on European standardization, OJ L 316, 14.11.2012. Art. 2, para.1, lit. c) defines a «harmonized standard» as «a European standard adopted on the basis of a request made by the Commission for the application of Union harmonization legislation». Among the other relevant pieces of legislation, see European Parliament and the Council, Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products, OJ L 218, 13.8.2008, and European Parliament and the Council, Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, OJ L 218, 13.8.2008. Consider also the EC documents, especially European Commission, Communication of 1 June 2011, *A strategic vision for European standards: Moving forward to enhance and accelerate the sustainable growth of the European economy by 2020,* COM(2011) 311 final; European Commission, Communication of 1 June 2016*, European standards for the 21st century,* COM(2016) 358 final; and European Commission, Communication of 22 November 2018, *The Single Market in a changing world: A unique asset in need of renewed political commitment,* COM(2018) 772 final. In this last document, the EC states that «standardization has played a leading role in developing the Single Market by supporting market-based competition and by helping ensure the interoperability of products and services.
Products which comply with voluntary, harmonized standards endorsed at Union level benefit from a presumption of conformity and can therefore move freely in the Single Market. This has been very beneficial, for instance, in the field of engineering or information technologies. While harmonized standards are developed by European standardization bodies, the Commission initiates, manages and monitors these standards and bears the ultimate responsibility» as stated by the Court of Justice of the European Union (ECJ). See also European commission, Communication of 22 November 2018, *Harmonized standards: Enhancing transparency and legal certainty for a fully functioning Single Market*, COM(2018) 764 final, where the Commission recognized certain shortcomings of the EU standardization process —particularly in the context of rapidly evolving technological developments—and declared its commitment to addressing them.

[21] European Commission, *Commission Implementing Decision C(2023) 3215 final on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on Artificial Intelligence*, Brussels, 2023, in *europa.ue*

[22] The first request for standards was already advanced by the EC to ESOs, see European commission, *Harmonised Standards for the European AI Act*, in *publications.jrc.ec.europa.eu.*

[23] Indeed, many of these standards are developed with a strong influence of private interests often outweighing public accountability, raising worries about the limited

especially when such technical elements may impinge upon the fundamental rights of the individual. Indeed, one of the main aspects differentiating the AIA from other pieces of—pure—product safety legislation is represented by its focus on the potential violation of fundamental rights[24], and on the possible tension that may arise when technical standards interact with the protection of those rights[25].

2.2. The Risk-based Approach as a Way to Offset the Horizontal Approach of the AIA

While the notion of risk is clearly central in the AIA[26], its concept may

democratic oversight in the standard-setting process. See C. Scott – F. Cafaggi – S. Linda, *The Conceptual and Constitutional Challenge of Transnational Private Regulation*, in *Journal of Law and Society*, 38, 2011, 1 ff. Furthermore, in such a perspective, it shall be noted that this reliance on privately developed standards triggers serious concerns about transparency, accountability, and potential conflicts of interest. Because private stakeholders dominate much of the standardization process, crucial product safety rules are shaped largely outside democratic institutions. For this reason, the ECJ addressed this issue in its landmark 2013 ruling (Case C-613/12, *James Elliott Construction Limited v. Irish Asphalt Limited*), asserting that harmonized standards fall within the scope of EU law and are therefore subject to judicial review. This ruling underscored the principle that, although technical standards are developed by private entities, they remain subject to some form of democratic scrutiny. The decision has since been widely discussed in legal scholarship, with many commentators arguing that it represents a necessary check on the growing influence of private standard-setting bodies in public regulatory matters. See, for instance: R. Vallejo, *The Private Administrative Law of Technical Standardization*, in *Yearbook of European Law*, 40, 2021, 172 ff; C. Tovo, *Judicial review of harmonized standards: Changing the paradigms of legality and legitimacy of private rulemaking under EU law*, in *Common Market Law Review*, 55, 2018, 1187 ff.

[24] This focus is also specifically recalled by Recitals 6 and 7 AIA, especially when referring to h-AIS. M. Almada – N. Petit, *The EU AI Act: a medley of product safety and fundamental rights?*, cit., 7 ff.

[25] In this sense, see L. Lessing, *Code Version 2.0*, 2006, in which the author conceptualizes technical architectures as regulatory structures that shape and constraint behaviour in ways comparable to legal rules. When applied to standardization processes, this underscores how technical norms drafted by private or transnational bodies may predetermine how legal rights are exercised in practice. Such *de facto* regulatory effects may generate tensions with fundamental rights guarantees, insofar the design choices embedded in technical norms can restrict or privilege certain rights.

[26] Observing current global regulatory initiatives for AI, it is evident that the risk-based approach appears to be dominant in a global perspective when considering AI regulation. For example, the Bletchley Declaration, signed in November 2023, by 28 countries (including the United States, China, and the European Union) during the UK AI Safety Summit, emphasizes that nations should account for AI-related risks and, where appropriate, adopt «classifications and categorizations of risk based on national circumstances and applicable legal frameworks». Similarly, the Council of Europe's Convention on Artificial Intelligence integrates general principles with a risk-based approach. It mandates actions for «identification, assessment, prevention, and mitigation of risks and impacts to human rights, democracy, and the rule of law» at all stages of the AI lifecycle, from design to decommissioning. Meanwhile, in the United States, President Biden's Executive Order on AI assesses risks associated with dual-use foundation models, especially those with publicly available model weights, and explores potential mechanisms to mitigate risks while maximizing societal benefits.

appear ambiguous. Indeed, risk-based regulation is typically non-European, and relies upon the exact assessment of an existing risk, possibly measured and quantified by scientific research[27]. To the contrary, the regulation of precautionary principle has always characterized European legislation[28], requiring the adoption of measures even when scientific evidence remains inconclusive or uncertainty persists[29], so long as there are reasonable grounds for concerns[30].

In such a perspective, the AIA's risk-based approach emphasizes proactive risk mitigation aiming to prevent potential harm before it materializes, rather than merely reacting to identified risks, and may thence be rightfully framed within the European tradition, rooted in the precautionary principle[31].

---

[27] In contrast, the risk-based approach employed in the United States often operates ex post, addressing risks only after they have been identified and shown to be harmful. This stems from a traditional suspicion of government regulation in U.S. law, which requires an extensive factual record demonstrating "significant risks" to justify regulatory action aimed at protecting public health from environmental contaminants. This foundational norm of U.S. legal culture makes precautionary regulation more difficult, as the government must first assemble substantial evidence to support its actions. In this sense see G. Charnley – E. D. Elliott, *Risk versus Precaution: Environmental Law and Public Health Protection*, in *Envtl. L. Rep. News & Analysis*, 32, 2002, 1036 ff and O. Renn – E. D. Elliott, *Precautionary Regulation of Chemicals in the US and EU*, in J. Hammit – M. Rogers – P. Sand – J. B. Wiener (eds.), *The reality of Precaution. Comparing risk regulation in the United States and Europe*, 2013, ff. 224. Furthermore, for a critique on the European approach see the seminal work of G. Majone, *The Precautionary Principle and Its Policy Implications*, in *Journal of Common Market Studies*, 40, 2022, 89 ff. Majone critically examines the EC's promotion of the precautionary principle, arguing that while it has a role in addressing imminent and irreversible risks, its broader application is fraught with logical, economic, and legal shortcomings. He warns that it can distort regulatory priorities, enable protectionism, and undermine international cooperation, highlighting a sharp contrast to the U.S. risk-based, evidence-driven regulatory framework.

[28] Such as, for instance, the EU Food Safety Regulation. See European Parliament and the Council, Regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety, OJ L 31, 1.2.2002, 1 ff, art. 7.

[29] K. De Smedt – E. Vos, *The Application of the Precautionary Principle in the EU*, in H. A. Mieg (eds.), *The Responsibility of Science. Studies in History and Philosophy of Science*, 2022, 163 ff.

[30] Zander explains that in international law the precautionary principle is often used based on scientific discovery. However, the author points out, that international practice has also permitted the use of the precautionary principle «when the scientific evidence is so scarce that it is not possible to carry out a meaningful assessment of it». J. Zander, *The Application of the Precautionary Principle in Practice. Comparative Dimensions.*, 2010, 1 ff, 73. For an account that does not fully adhere to a sharp distinction between risk-based regulation and the precautionary principle, conceiving the latter as a technique of anticipatory risk governance, see F. De Leonardis, *Il principio di precauzione nell'amministrazione di rischio*, Milano, 2005, 1 ff.

[31] This approach is coherent with the overall European regulatory approach in the technological domain. Indeed, the EU has longstanding reliance on risk-based regulation in various sectors, which has allowed policymakers to strike a balance between innovation and precaution. Risk-based regulation itself is a flexible tool, capable of taking different forms depending on how much of the precautionary principle is incorporated. More recently, this approach has been central to the EU's strategy for regulating the digital

More specifically, the notion of risk is used to offset the horizontal regulatory approach, and – in the intention of the legislator – prevent posing too high a burden on innovation[32], with the concept of risk broadly defined with reference to health, safety and fundamental rights[33]. However, reference to risk represents primarily a narrative used to justify regulatory intervention, rather than the specific assessment of a well-defined concern or potential threat. Said otherwise, the very classification pursuant to different levels of risk is nothing more than a pure conventional and synthetic expression, that does not necessarily correspond to an objective, much less measurable, truth. Indeed, no criterion nor methodology are identified or described by the EC that allows to determine why a given application, and/or class of applications ought to be deemed of greater or lesser concern. One might justifiably wonder why manipulation (art. 5, para. 1, lit. a) AIA) should warrant outright prohibition[34], while retrospective remote biometric identification systems are merely classified as potentially high-risk, and deepfakes[35] are neither prohibited nor even

domain. Beginning with the Digital Single Market Strategy, risk-based regulation has been applied (albeit with varying approaches) across diverse areas such as data protection, online content and platform governance, cybersecurity, and the regulation of digital products and services. The common goal of this regulatory technique is to mitigate risks proactively by guiding technological development and usage as they emerge, rather than merely responding to harms after they occur. In the context of emerging technologies, such as AI, this strategy operates on the normative assumption that the benefits of technological adoption outweigh the potential risks, provided that these risks are effectively managed. In this sense, see M. E. Kaminski, *The Developing Law of AI: A Turn to Risk Regulation*, in *University of Colorado Law Legal Studies Research Paper No. 24-5*. For a more general account of European risk-based regulation in the digital age, and of the multifaced approaches it adopts, see G. De Gregorio – P. Dunn, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Ag*, in *Common Market Law Review*, 52, 2022, 473 ff and for an account of the risk-based approach in the AIA see E. Longo, *La disciplina del "rischio digitale"*, in F. Pizzetti – S. Calzolaio – A. Iannuzzi – E. Longo – M. Orofino (eds.), *La regolazione europea della società digitale*, Torino, 2024, 53 ff, 72.

[32] As Mahler explains, the risk-tiered approach has provided a limitation on the obligations imposed upon non-problematic applications. See T. Mahler, *Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal*, in *Nordic Yearbook of Law and Informatics*, 2022, 246 ff.

[33] N. A. Smuha – K. Yeung, *The European Union's AI Act: beyond motherhood and apple pie?*, in N. A. Smuha (eds.), *The Cambridge Handbook on the Law, Ethics and Policy of Artificial Intelligence*, Cambridge, 2025, 236 ff.

[34] The increasing integration of AI into interpersonal and digital interactions poses serious risks, especially regarding the manipulation of human emotions and trust. Generative AI technologies are increasingly being exploited to commit deception and fraud on a massive scale. One example is the rise of AI-enhanced romance scams, as discussed in the 2024 Alan Turing Institute Report by S. Moseley, *Automating Deception: AI's Evolving Role in Romance Fraud*, 2025, 7 ff, which outlines how LLMs can create convincing fraudulent narratives and synthetic identities. These technologies empower not only individual scammers but also facilitate industrial-scale criminal operations, producing fabricated personas and backstories that evade conventional detection systems. Furthermore, AI-driven psychological profiling allows fraudsters to identify and exploit individual vulnerabilities with unparalleled efficiency, intensifying the impact of these manipulative practices and increasing human vulnerability to digitally enabled threats.

[35] Despite the increasing prevalence and harm of deepfake technologies, particularly non-consensual sexual deepfakes, current EU legislation fails to provide an adequate

considered high-risk.
While in certain cases the solution reached by the European legislator is commendable – e.g. prohibiting manipulatory AISs[36] (despite the limits imposed by the norm may still be less than satisfactory) – there appears to be no objective grounds or methodologies that account for the conclusion reached. There is no ranking of the fundamental rights potentially affected, that would justify such conclusions, nor a description of a balancing technique between the very same rights being at once positively and negatively impinged upon by an identical application.
In such a perspective, the very notion of risk regulation seems to function as a mere synthetic label used to justify a political choice, and its primary purpose is that of offsetting the unacceptable effects of a horizontal approach that attempts to encompass all existing and foreseeable technological applications. This attempt to reduce all technology to a "one size fits all" regulatory framework has been heavily criticized, as it leads to the «[…] substitution of technical complexity with legal complexity, which does not always lead to more straightforward regulation overall»[37].

2.3. The Horizontal Approach and the Importance of Definitions

The coupling of a risk-based categorization with a horizontal regulatory approach was conceived as a solution to the challenge of addressing heterogeneous AISs that can differ significantly from one another, in both function and technical design.
Indeed, the extremely broad, all-encompassing, definition of AIS put forth by art. 3 AIA is a choice precisely intended to ensure maximum

regulatory framework. Not only the AIA, but also the Digital Services Act (DSA) fall short of directly addressing the specific and severe risks posed by such content. Notably, the AIA, despite its human-centric rhetoric, classifies deepfake generation merely as a "limited risk" activity, subject only to minimal transparency obligations, and omits any explicit prohibition of non-consensual sexual deepfakes. This reveals a substantial regulatory gap; as a consequence, serious and harmful applications of AI, such as the creation of non-consensual sexual deepfakes, remain effectively unaddressed within the EU's emerging AI governance framework.

[36] There are many domains in which the effects of a given technology are inherently ambiguous. For instance, PARO (an advanced interactive therapeutic robot designed to support patients with dementia, Alzheimer's and other cognitive disorders) is generally considered unproblematic, largely due to its simplicity and long-standing presence. However, despite its therapeutic benefits, it can still mislead users by simulating emotional responses, thereby raising ethical concerns about deception and informed interaction, particularly in vulnerable populations. See A. Bertolini, *Human-Robot Interaction and Deception*, in *Osservatorio del diritto civile e commerciale, Rivista semestrale*, 2018, 131 ff; A. Bertolini, *The subtle line between personalization and user manipulation in a European regulatory perspective. A proposal for a technology-assessment methodology for Artificial Intelligence Systems*, in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, 4479 ff.

[37] M. Almada, *Two Dogmas of Technology-neutral Regulation*, in *ssrn.com*. On this point, see also A. Bertolini, *Artificial Intelligence does not exist! Defying the technology-neutrality narrative in the regulation of civil liability for advanced technologies*, in *Europa e diritto privato*, 2022, 369 ff.

technological neutrality[38]. More so, the AIA does not regulate AI technology as such, but rather its use.

That "technology neutrality" is in fact one of the foundational principles of European AI regulation[39], as clearly state by the Commission[40].

However, the very notion of technological neutrality is ambiguous and could lead to contradictory interpretations, as an identical legal provision may be deemed neutral or not according to the chosen perspective. Indeed, a functional understanding requires that legislation should not treat equivalent technologies (those capable of serving similar functions in the same context) in different ways, thereby upholding the principle of non-discrimination[41]. A substantive notion, instead, entails that the rationales that a given piece of legislation pursue should not vary across domains; in such a perspective dedicated norms may be required, differentiating between technologies[42] to ensure an identical treatment – and/or user protection – is granted irrespective of the system used[43]. The two perspectives could

[38] The original definition of AIS in the AIA referred to a specific set of programming techniques, and was thence criticized for lacking sufficient technological neutrality. This approached risked both obsolescence, by failing to capture future AI development, and over-inclusion, by potentially bringing under regulation certain systems not intended to be covered. As AI technologies evolve rapidly, new methodologies may emerge that fall outside the scope of Annex I, thereby escaping regulation despite posing comparable risks. Conversely, some traditional software systems (particularly those using statistical or logic-based operations) might have been unintentionally captured within the regulatory perimeter, despite lacking the autonomous or adaptive capabilities typically associated with AI. In this sense, see M. Almada, *Delegating the Law of Artificial Intelligence. A Procedural Account of Technology-Neutral Regulation*, European University Institute, 2024; 1 ff, 31.

[39] *Ibid,* 73.

[40] «The proposal [of AIA] sets a robust and flexible legal framework. On the one hand, it is comprehensive and future-proof in its fundamental regulatory choices, including the principle-based requirements that AISs should comply with. On the other hand, it puts in place a proportionate regulatory system centred on a well-defined risk-based regulatory approach that does not create unnecessary restrictions to trade, whereby legal intervention is tailored to those concrete situations where there is a justified cause for concern or where such concern can reasonably be anticipated in the near future. At the same time, the legal framework includes flexible mechanisms that enable it to be dynamically adapted as the technology evolves and new concerning situations emerge» in this sense, see European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. COM/2021/206 final, Brussels, 2021. See also art. 12, 14, 15 and 22 GDPR and, on the right to have an explanation, *ex multis*, S. Wachter, *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond*, in *Yale Journal of Law & Technology,*, 26, 2024, , S. Wachter – Brent Mittelstadt – F. Luciano, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 2017, 76 ff.

[41] C. J. Craig, *The Relational Robot: A Normative Lens for AI Legal Neutrality—Commentary on Ryan Abbott, The Reasonable Robot*, in *Jrslm Rev Legal Stud*, 25, 2022, 24 ff, 35.

[42] *Ibid.*, 37.

[43] An example is that of Recital 15 GDPR which ensures that users may obtain an explanation about a given decision, regardless of the technology used to automate it, explicitly discriminating against opaque AI solutions, see M. Almada, *Delegating*, cit., 30; M. Hildebrandt – L. Tielemans, *Data Protection by Design and Technology Neutral Law*, in *Computer Law & Security Review*, 29, 2013, 509 ff.

clearly be at odds[44].

The formulation of the AIA appears at least twofold. On the one hand, it seeks to uphold a principle of non-discrimination between technologies[45], which is why the regulatory approach focuses on uses rather than on specific tools[46]. On the other hand, it aims to establish a future-proof system[47], capable of being relevant despite rapid technological innovation. The latter is characteristic of legal systems, which can adapt, primarily through interpretation, once clear principles are established. However, this flexibility comes with compromises in terms of ex ante legal certainty, foreseeability, and consistency of outcomes, all of which affect market uniformity, particularly across MS (*infra*)[48].

On closer examination, the first perspective appears neither particularly valuable nor coherently implemented in the AIA. Treating all technologies equally is not desirable, as some better promote fundamental rights or support specific policy objectives[49], and therefore should be favoured.

In such sense the AIA is also not fully neutral, providing relevant discrimination between technologies, also through the so-called risk categorization. The latter, in fact, is not truly risk-based (see §2.2) as no criteria are

[44] An example shows the practical implications of distinguishing between functional and substantive neutrality. Scholars have debated whether individuals have the right to an explanation of automated decisions involving their personal data. If AISs are used, technical tools like explainable AI are needed. This meets GDPR Recital 15's idea of neutrality by allowing explanations regardless of the technology. Cfr. M. Almada, *Delegating*, cit., 30.

[45] See D. Kwak, *No More Strategical Neutrality on Technological Neutrality: Technological Neutrality as a Bridge Between the Analogue Trading Regime and Digital Trade*, in *World Trade Review*, 2022, 18 ff; W. J. Maxwell – M. Bourreau, *Technology neutrality in Internet, telecoms and data protection regulation*, in *Computer and Telecommunications Law Review*, 2015, 1 ff.

[46] European Commission, Proposal for a Regulation COM/2021/206 final, cit., 12 ff describes the definition of AI as technology neutral defining it as «software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with…».

[47] *Ibid,* 12 ff mentions that «[t]he definition of AI system in the legal framework aims to be as technology neutral and future proof as possible».

[48] That of ensuring a non-fragmented market is indeed one of the motivating factors for the adoption of a uniform European regulation in the field of AI.

[49] This is the case, for instance, of robotic prostheses that can enhance the rights of people with disabilities. See A. Bertolini, *Robotic prostheses as products enhancing the rights of people with disabilities. Reconsidering the structure of liability rules*, in *International Review of Law, Computers & Technology*, 29, 2015, 116 ff. This stance is indeed recognized also at the international level: for instance, art. 4 of the United Nations Convention on the Rights of Persons with Disabilities (A/RES/61/106) includes, among the general obligations of States, the duty «to undertake or promote research and development of, and to promote the availability and use of new technologies, including information and communications technologies, mobility aids, devices and assistive technologies, suitable for persons with disabilities, giving priority to technologies at an affordable cost». This provision underscores that new technological advancement in enhancing the autonomy, inclusion, and overall quality of life of persons with disabilities are particularly welcomed and can nowadays become a key instrument to close existing gaps in access and opportunity, particularly for the most vulnerable groups in society.

provided to enable an objective assessment of risk. Its categorisation thus does not follow from necessity but reflects a specific policy choice that is, by definition, not neutral. When extremely different applications – such as a lower limb exoskeleton[50] and an algorithm used for decisions affecting employment relationships (such as termination of labour contracts)[51] – end up classified under the same risk level (e.g. high-risk) it seems a questionable outcome, in particular considering the profoundly different technological characteristics and areas of application of each, and subsequently the risks they may give rise to.

At the same time – and as a consequence of problematic and ambiguous definitions such as the one of art. 6 AIA (see §4.2-4.5 and §6) – discrimination may arise between similar applications that differ only in limited technological aspects, and thus may fall either within or outside the high-risk category. Given the complexity of the definition and the numerous factors to be assessed for each AIS, it is also plausible that not all disparities between otherwise comparable technologies are intentional, but rather the result of the room left to interpretation.

As per the possibility of the overall system to preserve its relevance *vis a vis* future innovation because of its horizontal, all-encompassing, and thence neutral, regulatory approach, it is also questionable. More specifically, the framework's future adaptability is not consistently ensured[52], and overly broad constraints that fail to account for technological specificities may, in fact, stifle innovation[53]. The AIA is indicative of the uncertainty and short-circuits that technology neutrality engenders when it lacks precise and robust foundations.

Indeed, while the intent is to prevent regulatory bias, neutrality may give rise to interpretative ambiguities, making consistent and thus foreseeable enforcement difficult. This, in turn, may disproportionately benefit established players, who can navigate vague regulations – and absorb the costs

---

[50] An exoskeleton would be considered a h-AIS according to art. 6, para. 1, which in turns refers to Annex I. In the case of exoskeletons the parameters to be observed would be those set out by the Medical Devices Regulation (EU) 2017/745) and/or the Machinery Directive (2006/42/EC), depending on its application.

[51] Algorithms use in employment relations are considered to be h-AIS according to art. 6, para. 2, which refers to the list in Annex III. Specifically point (4) of the Annex specifies that an AIS will be high-risk if applied to employment, workers management and access to self-employment

[52] See B. A. Greenberg, *Rethinking Technology Neutrality*, in *Minnesota Law Review*, 100, 2016, 1495 ff, 1524, where the author argues that technology neutrality faces a fundamental problem of prediction: legislators cannot adequately foresee how laws should apply to emerging technologies until those technologies are known. As a result, laws intended to be future-prof are often poorly tailored, having been implicitly designed with existing technologies in mind. This predictive limitation might undermine the goal of adapting legal frameworks to paradigm-shifting innovation.

[53] *Ibid*, 1543, Greenberg critiques the illusion of neutrality in lawmaking, arguing that decisions about which technologies are subject to neutral treatment inevitably reflect social and political value judgments. Attempts to pre-emptively include future technologies may inadvertently create discrimination by failing to anticipate novel use cases. He advocates for a model of technological discrimination that combines neutrality with specificity, enabling more tailored, adaptive, and innovation-friendly regulation while enhancing legal clarity and policy alignment.

of *ex ante* compliance as well as the risks of potential *ex post* litigation – far more effectively than smaller innovators[54].

Ultimately, a persuasive account of technological neutrality, that perfectly fits the narrative embedded in the AIA, is that which considers neutrality as a technique to delegate choices to other actors. Indeed, the predominantly top-down nature of the AIA's risk categorization can be highly problematic. The Act itself, together with the EC's power under art. 7 to amend the Annexes, outlines the risk categories and the characteristics that place AISs into each category.

Although this structure is intended to create consistency, it may give raise to ambiguity, stemming primarily from (i) the breadth and fluidity of the defining criteria and (ii) the discretion policymakers retain in determining which systems qualify as high-risks or others[55]. Such a top-down approach can be arbitrary: a system's "risk level" may hinge on interpretive nuances, which can shift if the Commission revises the Annexes to reflect new priorities (art. 7 AIA)[56].

Moreover, unclear or evolving definitions have tangible consequences. A product labelled "high-risk" must meet heightened compliance requirements while a borderline system that avoids high-risk classification might sidestep critical safeguards. The net result is regulatory uncertainty on both ends: providers face unpredictable obligations, and regulators risk missing truly dangerous applications that do not neatly fit the AIA's categories.

Further complexity arises from the partial flexibility granted to providers of h-AIS. As discussed in the following section, art. 6 of the AIA allows

---

[54] W. J. Maxwell – M. Bourreau, *Technology neutrality in Internet*, cit., 2.

[55] In particular, given the concentration of regulatory and administrative powers in the hands of a single entity (the EC) and in the absence of an independent authority tasked with overseeing the matter. In this sense, see, *ex multis*, S. Calzolaio, *Autorità, governo, attuazione dell'AI Act*, in F. Pizzetti – S. Calzolaio – A. Iannuzzi – E. Longo – M. Orofino (eds.), *La regolazione europea dell'intelligenza artificiale nella società digitale*, Torino, 2024, 139 ff.

[56] The power granted to the EC to modify the lists of high-risk systems recalls legislative techniques already present in certain national legal systems, such as the so-called "*norme penali in bianco*" or "*Blankettstrafgesetz*". These are criminal provisions that defer the definition of their prescriptive content to external sources, often of secondary or administrative rank, thus delegating to these sources the concrete determination of criminally relevant conduct. This legislative technique has been widely criticized in legal scholarship, as it entails an extremely broad delegation of normative power, potentially lacking clearly predetermined limits, and potentially undermines the principle of legality, particularly in its aspects of statutory reserve and definiteness. In this sense see, among others, V. Manes, *L'eterointegrazione della fattispecie penale mediante fonti subordinate, tra riserva "politica" e specificazione "tecnica"*. in *Rivista italiana di diritto e procedura penale*, 2010, 84 ff. Similarly, the power attributed to the EC risks resulting in a substantial expansion of its decision-making authority, especially with respect to determining which applications may qualify as high risk and what does not. Such power is exercised outside the bounds of democratic oversight, bypassing both democratic debate and the ordinary legislative process. The criticisms traditionally directed at "*norme penali in bianco*" might therefore also relevant, *mutatis mutandis,* to the power of the EC, insofar as both raise concerns regarding democratic legitimacy and the protection of fundamental guarantees for those subject to the norm. Indeed, as further detailed in paragraph 3 of this Section, the Commission's power rests on the questionable assumption that risk assessment is a purely technical exercise. However, such determinations inevitably involve political choices and removing them from open political debate and legislative process might cause significant democratic *vulnus*.

providers listed in Annex III to argue that their systems do not actually pose the designated risk, effectively introducing a bottom-up mechanism (art. 6, para. 4, AIA). While this provision could, in principle, refine the top-down approach by limiting the Commission's power and inviting a more context-sensitive analysis, it has in practice greatly amplified existing uncertainty. Providers must conduct costly and complex self-assessments, while authorities maintain the latitude to reject or reinterpret claims based on evolving regulatory expectations. Coupled with the Commission's power to revise the Annexes, this results in a moving target for compliance. In this way, the AIA's risk-based classification becomes a conceptual junction, encompassing a broad spectrum of technical systems under one legislative umbrella while simultaneously allowing for case-by-case exceptions.
Rather than offering clarity, the interplay of broad definitions, top-down categories, and partial provider-initiated reclassification create a confusing regulatory environment. Stakeholders struggle to predict their compliance path, and authorities grapple with mitigating risks to health, safety and fundamental rights effectively, as it will be further address in the following Section.

3. The Midday – Article 6 and the Definition of h-AIS: Annex III and the High-risk Categories.

3.1. The Reference to Annex III

Art. 6 of the AIA identifies the high-risk applications by reference to Annex III (see art. 6, para. 2, AIA) or by imposing two criteria that must be met together (para. 1).
The reference to Annex III is particularly significant, as all AISs listed therein are automatically classified as h-AIS[57], and should thence precisely determine what applications are subject to the most stringent regulation contained in the AIA. However, for each of these eight areas, Annex III identifies specific conditions, objectives or purposes that the AIS needs to respectively meet or pursue to be deemed an h- AIS. The stated rationale is that all those applications negatively impact on the safety, health, or fundamental rights of individuals.
However, the rigid categorization of Annex III raises several concerns. First, it remains unclear what empirical evidence or criteria were used to define these risk categories and compile the list of covered areas and applications. Second, and more fundamentally, the existence of these risk categories demonstrates that the AIA departs from any claim of being truly "technology neutral". By selecting which AI applications warrant stricter oversight, the EU legislator made normative and political decisions about which uses are more problematic, rather than abiding by a purely

[57] Annex III outlines eight areas of application: 1) Biometrics; 2) Critical infrastructure; 3) Education and vocational training; 4) Employment, worker management, and access to self-employment; 5) Access to and enjoyment of essential private services and public services/benefits; 6) Law enforcement; 7) Migration, asylum, and border control management; 8) Administration of justice and democratic processes.

objective or "value-free" process. It is certainly acceptable, and often necessary, for some applications to be subject to stricter regulations, and it is inherently difficult to regulate, under uniform provisions, applications that differ fundamentally. However, in the AIA, the criterion used to assimilate and distinguish these applications is not grounded in a precise or empirically substantiated measurement of risk. Rather, what is referred to as "risk" functions more as a summary expression of a political and normative judgement about potential danger. This kind of judgement, while legitimate, is neither entirely objective nor based on a consistent, transparent evaluation of harm or likelihood. Indeed, no scheme of balancing is either proposed or framed: the classification of AISs into risk levels reflects a political choice rooted in specific priorities and societal values, that however are not narrowly nor clearly defined. For instance, this can be illustrated by the regulation of AIS that involve emotion recognition, particularly in workplace or commercial contexts. While such systems may significantly impact individuals' privacy, their classification as high-risk is not accompanied by an explanation of how these concerns are weighed against other competing interests, such as, for instance, economic freedom or business innovation. The AIA does not articulate the rationale behind prioritizing the right to privacy over companies' freedom to develop and deploy technologies that could, for example, increase productivity or enhance customer service. This reflects a conscious judgement that certain technologies or their uses require heightened scrutiny, while others are relatively less worrisome. However, as noted above, this judgement is not – and cannot be – explicitly motivated and/or objectively defensible.

The risk categorization process, therefore, cannot be purely neutral because it necessarily involves subjective assessments about which harms or rights matter most, a question that depends on shifting legislative, ethical and cultural standpoints. Hence, the AIA's framework embodies political choices that stand in tension with the notion of technology neutrality, revealing that regulators do, in fact, treat distinct AI uses differently in line with societal objectives. Far from being undesirable, these distinctions are often necessary to protect public welfare and fundamental rights, but they underscore why the idea of a fully neutral approach to AI technology is untenable, both in principle and in practice.

In addition to that, and as further evidence of what stated above, many AI applications are considered high-risk solely because they operate within one of these sectors, even if they do not necessarily pose a significant risk of harm. For instance, AISs used for monitoring and detecting prohibited student behaviour during exams in educational and vocational training contexts are automatically categorized as high-risk, irrespective of their technical specifications and of their actual potential to infringe fundamental rights, safety or health[58]. By contrast, certain AI tools deployed in the

[58] On the challenges posed by AI in educational settings, particularly regarding the balance between academic integrity and individual rights, see J. A. Oravec, *AI, biometric analysis, and emerging cheating detection systems: The engineering of academic integrity?*, in *Education Policy Analysis Archives*, 30, 2022, 1 ff and D. R. E. Cotton – P. A. Cotton – S. J. Reuben, *Chatting and cheating: Ensuring academic integrity in the era of ChatGPT*, in *Innovations in Education and Teaching International*, 61, 2024, 228 ff. The authors highlighted how surveillance-

healthcare sector (such as mental health chatbots that interact with users without human oversight) may not automatically fall into the high-risk category, even though their impact can be substantial.
Perhaps for these reasons, art. 6 also introduces exceptions to the automatic classification of all AI applications mentioned in the sectors listed as high-risk. However, these exceptions and counter-exceptions do not clarify the legal framework; rather they complicate it further.

3.2. Exceptions and Future Amendments

The first exception is introduced in art. 6, para. 3, which provides that AISs listed in Annex III may not be considered as high-risk when they do not pose a significant risk of harm to the health, safety or fundamental rights of natural persons. According to the same provision, this applies when at least one of the four criteria outlined is met: when the AIS performs a simple procedural task, supports or refines a human decision that has already been made, identifies patterns without influencing decisions, or merely prepares input for a later human determination (lit. a), b), c) and d))[59]. In such cases, the system, although formally falling under one of the categories in Annex III, should not be classified as high-risk.
This exemption clearly reflects an effort to avoid overregulation by excluding borderline use cases that do not materially influence decision-making outcomes. The underlying rationale is to prevent disproportionate compliance burdens for applications that, while formally included in high-risk areas, play only a limited or supporting role in practice. However, the provision relies on vague and open-ended language, such as "not materially influencing" or "narrow procedural task," which makes it difficult to determine *ex ante* whether a specific system qualifies for the exemption[60].

based cheating detection technologies and tools like ChatGPT raise unresolved tensions between privacy, fairness, and institutional control, often without a clear framework for balancing these competing values.

[59] According to art. 6, para 6 of the AIA, an AIS referred to in Annex 3 should not be considered to be high-risk where any of the following conditions is fulfilled: «(a) the AI system is intended to perform a narrow procedural task; (b) the AI system is intended to improve the result of a previously completed human activity; (c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or (d) the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.»

[60] Scholars have already pointed out that these exceptions can be problematic and seems to add complexity in the AIA's regulatory architecture causing further legal uncertainty. See, for instance, N. A. Smuha, *Nathalie A Smuha, Algorithmic Rule by Law: How Algorithmic Regulation in the Public Sector Erodes the Rule of Law*, 2024, 1 ff, 105; M. Ebers, *Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AI Act*, in *Stanford-Vienna European Union Law Working Paper No. 101*, 2024, 1 ff and S. Wachter, *Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond*, cit., 671 ff. More in general, for a comprehensive overview of the various provisions of the AIA, along with a detailed analysis of the obligations set forth by the Act, see P. Voigt – N. Hullen, *The EU AI Act. Answers to Frequently Asked Questions*, 2024, 1 ff.

The practical implications of the four criteria set out in art. 6, para. 3, are significant and may give rise to several key issues concerning its application, potentially leading to regulatory fragmentation across MS and creating legal uncertainty for providers and deployers.

First, determining when an AIS influences decision-making or is genuinely intended only to detect decision-making patterns, without replacing or unduly influencing prior human assessments, is extremely difficult. For example, in the context of a judicial decision-support tool, it may be difficult to distinguish between an AI that merely highlights trends in sentencing and one that subtly guides judges towards a particular outcome. One of the central challenges lies in defining how an AIS "merely detects" patterns rather than shaping or overriding the human decision-making process. Even when an AIS is designed to provide limited assistance, the way its outputs are generated and presented can significantly affect users' behaviour. In many cases, individuals perceive algorithmic recommendations as objective or expert, which can lead them to rely on these outputs, even if the system is nominally only assisting[61]. This illustrates how system explicability and user perception are interlinked. Indeed, a lack of clarity around how the system arrives at its suggestions often results in humans deferring to, rather than merely consulting, the AI's judgment, striking examples of which can be found, for instance, in the field of medical diagnostics[62].

Consequently, the use of the system (as shaped by its design and presentation) may substantially impact the final decision, calling into question the assumption that human assessments remain unaffected. This practical trend, exacerbated by the growing number of highly accessible applications of generative AI, underscores the fear that users will increasingly accept and over-rely on AI-generated suggestions as authoritative. From a scientific perspective, recent studies, including a comprehensive survey study by Microsoft Research[63], are beginning to empirically document this very phenomenon. These studies show that increased trust in AI correlates with reduced critical thinking and cognitive effort, suggesting that

---

[61] H.-P. Lee – A. Sarkar – L. Tankelevitch – I. Drosos – S. Rintel – R. Banks – N. Wilson, *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*, in *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 2025, 1 ff.

[62] See J. Wei – E. Verona – A. Bertolini – G. Mengaldo, *Explainability matters: The effect of liability rules on the healthcare sector*, in *arXiv.org*, 1 ff. The Authors illustrate that even when an AIS is nominally deployed as a mere "assistant", its perceived objectivity can prompt decision-makers to rely on its recommendations more than intended or warranted. In particular, they note that especially if the AIS functions as an "oracle" (i.e., provides a diagnosis or conclusion with no accompanying explanation), the human user (in the article, the medical practitioner) may strategically conform to the system's output to minimize potential liability. Indeed, when the reasoning is opaque, individuals often defer to the presumed accuracy or authority of the AIS. The authors underline that this behaviour can lead to a form of defensive medicine whereby users adopt the AIS's verdict primarily to shield themselves from potential liability, rather than to serve the best interests of the patient.

[63] H.-P. Lee – A. Sarkar – L. Tankelevitch – I. Drosos – S. Rintel – R. Banks – N. Wilson, *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*, cit.,1 ff.

overreliance is not just anecdotal, but has measurable cognitive and behavioural consequences in real-world knowledge work[64].

Second, ensuring consistency in how such assessments are conducted across different AISs and sectors remains a complex task. As mentioned in Section I, AISs vary greatly depending on their technological design and intended use. The same high-level criteria may yield different results when applied to distinct domains (for example, healthcare versus human resources) because the risk profile, data inputs, and societal implications of these technologies differ substantially.

This diversity underscores the difficulty of using a generalized or "neutral" approach to classification: while technology neutrality suggests regulating at a high level of abstraction, contextual factors can render the same criteria either too lenient or too strict, thereby resulting in uneven or inconsistent enforcement. For instance, a criterion requiring transparency in decision-making may be relatively easy to satisfy in a recruitment algorithm that filters CVs but could be impractically demanding for a diagnostic AIS that relies on complex, non-interpretable models to identify rare diseases. In the former, a failure to meet transparency standards might lead to minor compliance concerns; in the latter, it could result in denying access to potentially life-saving innovations.

Third, questions may arise regarding the extent and effectiveness of external and independent oversight in monitoring these evaluations. It remains unclear who will be responsible for evaluating and enforcing these provisions, and how such oversight mechanisms will operate in practice. In theory, robust external supervision (such as independent auditing bodies or regulatory agencies) could ensure that providers and deployers accurately classify their AISs under art. 6, para. 3. However, the Act does not explicitly define the extent to which such external oversight will be required or how it should be standardized across Member States. Without well-defined guidelines on independent review, there is a risk that evaluations will be inconsistent or that some providers may understate their system's influence on decision-making processes to avoid stricter compliance burdens.

In addition to these issues, it is important to note that another provision contained in art. 6, para. 3, may create further uncertainty. The last sentence of art. 6, para. 3, specifically comma 3, introduces an exception to the exception by adding a further limitation related to profiling[65]. Even if one of the four conditions that would normally exclude an AIS from being classified as high-risk is met, the exemption does not apply if the system engages in profiling, as defined by the GDPR[66]. While this carve-

---

[64] N. Kosmyna – E. Hauptmann – Y. T. Yuan – J. Situ – X.-H. Liao – A. V. Beresnitzky – I. Braunstein – P. Maes, *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task*, in *arXiv.org*, 2025, 1 ff.

[65] «Notwithstanding the first subparagraph, an AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons».

[66] When defining profiling, the AIA (art. 3, para. 52 refers to the definition of profiling provided by art. 4, point (4), of Regulation (EU) 2016/679: «profiling means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health,

out is intended to preserve safeguards against intrusive or discriminatory practices, its scope is not always easy to define. Seemingly benign applications, such as personalized content feeds or basic consumer segmentation, may fall within the definition of profiling and may therefore be deemed h-AIS, while systems used in public sector contexts to detect "deviations from prior decision-making patterns" in areas like law enforcement or welfare eligibility could indirectly steer decisions without being formally classified as profiling or without triggering the high-risk threshold, since they are framed as "mere" procedural or advisory tools.

These unresolved questions, unfortunately, represent only a fraction of the broader interpretative challenges posed by the structure of art. 6, para. 3.

On the one hand, it would appear to establish a presumption *iuris et de iure* of high-risk with regard to the profiling of natural persons, as defined in art. 4, para. 4, of GDPR[67].

On the other hand, it maintains the efficacy of comma 1 of para. 3, specifically the exemption based on the expansive and ambiguous formula previously examined, which states the absence of substantial risks to the health, safety, or fundamental rights of the individual. However, the provisions of comma 2 of para. 3 are not referenced, even though they should constitute a specification of the first.

This prompts numerous inquiries into the interpretation of the text. The exclusion of the initial clause, «without prejudice to the first subparagraph», would result in the classification of all systems engaging in profiling as high risk, thereby eliminating the necessity for additional evaluation. This conclusion would be consistent with the English version of the text, which employs the expression "notwithstanding", and with the German version, which utilizes the term "*ungeachtet*", both of which can be translated as "by way of derogation". By derogating from the initial paragraph, the exception to the rule would be neutralized, thereby classifying any user profiling system as high risk.

However, in the Italian version of the text, the exception referred to in comma 1 is expressly reserved, thus equating profiling with all the other figures indicated in Annex III. The literal interpretation of the text does not appear to allow for divergent interpretations. Consequently, it does not seem possible to consider that profiling applications are subject to a special, worse, or more stringent regime.

In order to attribute meaning to the provision in comma 3 of para. 3, it is necessary to establish a parallel with the regulations concerning unfair terms in consumer contracts[68]. In that context, it is not possible to demon-

---

personal preferences, interests, reliability, behaviour, location or movements».

[67] This is the interpretation given by the Commission, which, to date, is supported by the warning it issues in the EU AI Act Compliance Checker, a novel instrument of the AI Office. This tool has been crafted to clarify the obligations and requirements of the AIA. Once the self-assessment has begun and you have reached the question of whether the AIS can be considered high risk pursuant to art. 6, para. 2, AIA, you must specify whether derogations pursuant to art. 6, para. 3, AIA apply. At this point it is literally warned to note that «an AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons (Article 6 (3) AI Act)». See ai-act-service-desk.ec.europa.eu.

[68] For the parallelism please refer to A. Bertolini, *La definizione di sistemi di Intelligenza*

strate that clauses included in the so-called "black list" (art. 36, para. 2, of the Italian Consumer Code) are not unfair by providing evidence of individual negotiation, without prejudice to other cases of exclusion. In this case, too, it would be necessary to reconstruct a different regime for profiling applications by way of interpretation, i.e., by emphasizing the lack of reference to comma 2 of the same paragraph. Consequently, it can be deduced that, in the context of systems engaged in profiling activities, the criteria delineated in comma 2 from lit. a) to d), are not applicable. This would be equivalent to saying that profiling justifies classifying the system as high risk unless it can be demonstrated that it does not seriously violate the security, health, or other fundamental rights of the individual. Conversely, the developer of a profiling system could exclude its high-risk nature by incorporating the exception cited in the comma 1, without resorting to the cases enumerated in comma 2.
While this interpretation appears more consistent in terms of the formal and literal interpretation of the text, it is not at all satisfactory from a functional and remedial point of view. This raises the question of the very rationality of the provision and its real meaning and scope. Conversely, if the European legislator's intent is to be considered as having been otherwise, it would be necessary to conclude that a significant translation error has been made. This would be analogous to the situation that transpired when the regulations concerning unfair terms were initially introduced[69].
Given the aforementioned points, it is evident that these considerations collectively contribute to the overall ambiguity of the provision and intensify uncertainty regarding its practical application.

3.3. Amendments to Annex III

In addition to the above, it should be stressed that the list of h-AISs established by the AIA is subject to ongoing updates, as delegated to the EC under art. 7 of the AIA. This delegation grants the Commission the authority to modify the Annex over time, either by adding or removing specific applications or by adjusting the criteria required for systems to qualify as h-AIS. As a result, the regulatory landscape for AI applications is inherently dynamic, with significant implications for providers. However, as already outlined above (§2.3), this power is extremely broad and rests on the fallacious assumption that the definition of risk is the outcome of a purely technical assessment. In reality, such determinations are inherently political, and by removing them from open political debate, the mechanism introduces a potentially significant democratic vulnerability[70].
For instance, applications initially designed to comply with the require-

Artificiale, in G. Cerrina Feroni – S. Orlando – G. Vettori (eds.), *Il Regolamento europeo sull'intelligenza artificiale. Problemi e Sistema*, Bologna, 2025, 183 ff.

[69] *Ibid*, 197.

[70] For an analysis of the vulnerabilities inherent in the digital society and their constitutional implications within the emerging datafied legal order, see, *ex multis,* S. Calzolaio, *Vulnerabilità della società digitale e ordinamento costituzionale dei dati*, in *Rivista italiana di informatica e diritto*, 2, 2023, 14 ff.

ments for h-AISs might later find themselves subject to a more lenient regulatory regime if changes to the Annex or criteria occur. Conversely, applications not originally classified as high-risk might, at a later stage, be added to the Annex if regulators identify emerging risks associated with their deployment or functionality. This adaptability ensures that the regulatory framework can respond to technological advancements and unforeseen developments, but it also introduces a degree of uncertainty for AI providers. For example, a novel application that was not initially anticipated as problematic during its conceptualization and early design stages might later be deemed a source of significant concern. If this occurs, providers may be compelled to revisit and modify the design of their application to ensure adherence to the newly imposed requirements for h-AISs[71]. Such retrospective compliance obligations could be particularly burdensome for applications with extended research, development, and testing timelines. In the field of robotics, for instance, the development cycle for complex applications often spans several years, encompassing stages from initial research and prototyping to rigorous testing and deployment[72]. This extended timeframe increases the likelihood that the regulatory landscape may evolve during the course of development. Consequently, providers may find themselves needing to adapt to new compliance requirements mid-project, which can lead to significant delays, increased costs, and resource reallocation. This is especially challenging for those working on sophisticated, embodied AISs intended for critical or high-stakes environments, where safety and reliability are crucial.

3.4. The "Self-Declaration"

Finally, art. 6, para. 4, allows any provider to conclude that their AIS is not high-risk, even if it falls within the scope of Annex III, provided they document their assessment, register the product and, upon request of national authorities, provide the documentation upon which the assessment is grounded.

This means that a single manufacturer could argue that their device, service, or application should not be classified as high-risk, despite it theoretically falling into that category. To support this claim, manufacturers must document their assessment before placing the system on the market or putting it into service and comply with the registration obligation set out

[71] See on this E. Commission, *Study on the functions and effects of European standards and standardisation in the EU and EFTA Member States and Annexes*, 2021. This document notes that, particularly in fast-moving technology sectors, European standards are often perceived as lagging behind technological developments. Stakeholders report long intervals between the technical finalization and publication of standards, making them less effective for companies needing to adapt rapidly to new regulatory or market requirements (p. 10; see also pp. 158-159). This temporal misalignment reinforces the concern that developers may face significant redesign burdens if compliance conditions shift mid-development.

[72] OECD, *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*, Paris, 2023; N. Alistair, *Artificial intelligence in science: challenges, opportunities and the future of research*, 2024.

in art. 49, para. 2. Only if the national authority requests it, the manufacturer is required to provide the documentation for review.
However, a series of critical ambiguity arises regarding the scope of application of this provision. It remains unclear whether this provision is intended to apply only to systems referred to in Annex III that shall not be considered to be high-risk because they exhibit some of the characteristics outlined in para. 3 (for instance, because they are intended to perform a narrow procedural task, see para. 5) or whether it can be invoked more broadly on the basis of other considerations, eventually more technical in nature, or, again, based on the assessment of the criteria and condition listed for each of the eight categories by Annex III. Furthermore, one might also question whether art. 6, para. 4, could also extend to AISs falling under para. 1 of art. 6.
This lack of clarity may give rise to several criticalities. First, it may lead to divergent interpretations and practices among providers and national authorities, potentially undermining the uniform application of the regulation and creating legal uncertainty for both regulators and providers.
Secondly, the distinct regime under art. 6, para. 4, introduces a criterion and a procedure that may prove insufficiently rational and potentially discriminatory in relation to the scope and the entities covered by art. 6, para. 1. This raises concerns regarding the legal soundness of a potential dual-track system.
Thirdly, the process described in para. 4 lacks any formal decision-making stage or mechanism of official endorsement. There appears to be no principle of tacit approval through administrative silence – *i.e.,* no rule whereby an individual's request is automatically approved if the administration fails to respond within a set period[73] – nor any systematic evaluation procedure conducted by an independent authority. As a result, the provider is left without definitive confirmation that the classification of the AI system as non-high-risk is correct; and, conversely, even if it were to receive express confirmation from a competent authority, the result would no longer constitute a self-certification but rather a certified declaration. This creates legal uncertainty and raises the risk that the system may ultimately fail to meet the requirements for h-AISs as set out in the regulation.
Even more concerning, this gap may lead to situations where the manufacturer is forced to comply with high-risk obligations only at a later stage (when the system is already fully developed) with potentially serious consequences. This could eventually require a complete or partial redesign of the system, resulting in significant costs and delays.
Without an authoritative validation process, both compliance and accountability are compromised, potentially undermining the effectiveness of the regulatory framework intended to safeguard fundamental rights and pub-

[73] A principle of this kind, for example, exists in Italy under art. 20 of Law No. 241/1990, which states that «in proceedings initiated by an application of a private party for the issuance of administrative measures, the silence of the competent administration is equivalent to acceptance of the request, without the need for further applications or warnings, if the administration does not notify the applicant within the time limit set out in art. 2, para 2 or 3». For an overview, see P. G. Lignani, *Silenzio (diritto amministrativo)*, in *Enciclopedia del diritto*, 1999, 978 ff.

lic interests. From a procedural perspective, the absence of a clear "review step" means there is no formal moment in which an authority confirms or disputes a provider's classification. This lack of adjudicative oversight not only deprives the provider of legal certainty but also impedes transparent public accountability, since no authoritative body is required to justify or document its reasoning in approving (or rejecting) the provider's decision. Notably, from a procedural point of view, formal decision-making processes are seen as essential to ensuring uniform application of regulations and enabling due process, yet here those mechanisms are notably absent.
Furthermore, weak *ex ante* oversight can induce moral hazard, since providers may be tempted to classify borderline systems as non-high risk to reduce compliance burdens, all while shifting potential costs (for example harms to users or the public) to a later stage. Without early independent verification, these providers might not internalize the full societal risks of under-compliance. Indeed, rigorous *ex ante* evaluation is often credited with aligning private incentives (profit, speed to market) with broader social welfare goals[74]. If that step is missing, the risk of under-compliance grows, as does the possibility of negative externalities, such as compromised safety or erosion of fundamental rights. In practice, this may erode trust in the market for AISs and weaken the overall regulatory scheme's efficacy. Consequently, while self-declaration might seem efficient at first glance, it creates a fragile balance between promoting innovation and preserving public safeguards, underscoring the need for a more authoritative and transparent validation mechanism.
In conclusion, considering the latitude left to interpretation by the wording of the norm, there is ample scope for discretionary judgments by different authorities. This ultimately suggests that pursuing such an avenue of self-declaration is feasible but highly risky, both for providers (who lack certainty) and for the individuals, whose fundamental interests the regulation should aim to protect.

4. The Evening – Art. 6, para. 1, of the AIA: Defining h-AISs

4.1. The Complexity of para. 1

Para. 1 defines the other, more general conditions, that lead an AIS to be deemed high-risk, by referencing to pre-existing Union Harmonized Legislation (UHL), listed in Annex I, and formulating two distinct criteria, namely:

a. «the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I;
b. the product whose safety component pursuant to point (a) is the AI

[74] For a general perspective on self-assessment mechanisms, see R. A. Epstein, *The Use and Limits of Self-Valuation Systems*, in *University of Chicago Law Review*, 81, 2014, 109 ff and S. Levmore, *Self-Assessed Valuation Systems for Tort and Other Law*, in *Virginia Law Review*, 68, 1982, 771 ff.

> system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I».

Both criteria require complex interpretation. On the one hand, the definition of safety component proffered by art. 3, para. 14 AIA is circular in nature and leaves relevant uncertainty about which component is to be deemed relevant for the analysis, especially in complex and interconnected systems (see §§4.2 and 4.3). On the other hand, the wording chosen when referring to third-party conformity—namely it being "required"—, may radically affect the scope of application of the provision (see §§4.4-4.5), causing it to either encompass a broad spectrum of AISs or, instead, significantly limiting its relevance (see §§4.6-5).

4.2. The Safety Component

The first requirement laid out by lit. a) is met both when the AIS overall, or one of its safety components is subject to UHL. However, while the first instance is easy to verify, the latter rests on the ambiguous notion of "safety component" defined by art. 3, para. 14 AIA as a

> «component of a product or of an AI system which fulfils a safety function for that product or AI system, or the failure or malfunctioning of which endangers the health and safety of persons or property».

The definition is clearly circular – such is the reference to a "safety function" – and fails to provide clear operational criteria that would allow either a legal or technical expert from objectively concluding that a given subsystem is, in fact, the item upon which the remaining analysis – namely that of lit. b) – ought to be carried out.

Indeed, based on the very wording, it is disputable whether the AI-assisted vision system of a mobile robot operating in a human-populated environment is to be deemed a safety component, or solely its braking system (example 1). While it is certain that a malfunctioning in the sensing and elaboration of the information collected could cause an accident, the same is true most likely for most components in the robots itself. This leads to a potentially unbounded interpretation, in which nearly any subsystem could be relevant simply because its malfunction might contribute, however indirectly, to harm. Such reasoning risks an infinite regress of causal attribution, akin to what arises when the *conditio sine qua non* test is applied without normative constraints: if every factual precondition is treated as equally relevant, the scope of responsibility—and here, by analogy, regulatory attention—expands endlessly[75].

[75] Indeed, even the consultation initiated by the EC on June 6th, 2025 identified the notion of safety component as one that could provide a lot of uncertainty to stakeholders. European commission, *Commission Launches Public Consultation on High-Risk AI Systems*, in *digital-strategy.ec.europa.eu*.

This problem is well known in criminal law's causality theory, particularly in the context of the *conditio sine qua non* test. Without normative constraints, this test leads to a *regressus*

At the same time, the AIS could be designed in a way that the braking system is always in condition to intervene, overriding all other inputs provided by the AI-controlled components, perhaps on the basis of the inputs provided by independent proximity sensors. It could thence be questioned whether such a technical solution would suffice in excluding the "safety-nature" of the AI-assisted vision system and therefore allow the entire AIS to escape the subsequent qualification as high-risk, and its regulatory implications.

Uncertainty on the side of manufactures could lead to both over- and under-compliance, also depending on the market positioning of the individual player[76], as well as create an undesirable[77] potential barrier to entering the market.

Moreover, considering how art. 6, para. 1, AIA itself references to UHL—listed in Annex I—that, in turns, adopts individual definitions of "safety component" for the specific purposes of the domain each directive or regulation insists upon[78], a clear coordination between those overlapping frameworks appears necessary, and yet it is missing. The Commission has issued a consultation through which stakeholders were called to present their doubts and opinions about some concepts pertaining to AIA. The survey focuses mostly on the definition of h-AIS according to art. 6, para. 2—as discussed in Section "Midday". Nonetheless, section 1 of the survey opens the space for stakeholders to pose questions related to art. 6, para.

ad infinitum, where any antecedent fact (even remote or incidental ones) could be seen as causally relevant. German legal scholarship addresses this through the doctrine of *objective Zurechnung* which introduces a normative filter to identify legally significant causes. In this sense, see C. G. Roxin, Luis, *Strafrecht Allgemeiner Teil*, 1, 1992. In the Italian legal tradition, a similar concern arises in the distinction between mere physical causation and *rilevanza giuridica*, requiring that the causal link fall within the scope of the protected legal interest. See T. Padovani, *Diritto Penale*, Milano, 2002.

[76] Indeed, it may be anticipated that a larger company who is less concerned about facing ex-post litigation and sanctions, could take the risk of non-complying under uncertainty, whereas a SME or a startup might prefer to pay a certain, higher ex-ante compliance cost that face potential litigation and sanctions at a later day. As explained by Sommers and Cole, «[b]ecause the costs of avoiding requirements are so high, small firms are left with the choice of complying, at very high costs per dollar of sales, or of choosing partial or non-compliance strategies, hoping that regulatory agencies will fail to detect their non-compliance». P. Sommers – R. J. Cole, *Compliance costs of small and larger businesses*, in *Policy Studies Journal*, 1985, 701 ff. Mendoza et al. also point that firm size directly influence the capability of acquiring information and, thus, compliance: «there is a positive and significant association between firm size and knowledge. Acquiring knowledge of complex regulations may be more difficult for smaller firms». J. P. Mendoza – H. C. Dekker – J. L. Wielhouwer, *Firms` Compliance with Complex Regulations*, in *Law and Human Behavior*, 40, 2016, 721 ff.

[77] Indeed, «legal standard is uncertain, even actors who behave 'optimally' in terms of overall social welfare will face some chance of being held liable because of the unpredictability of the legal rule» J. E. Calfee – R. Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, in *Virginia Law Review*, 1984, 965.

[78] These include Machinery Directive (henceforth MD); Machinery Regulation (MR); Toy Safety Directive (henceforth TSD); Lifts and Safety Components Directive (henceforth LD); Radio Equipment Directive (henceforth RED) and the Medical Devices Regulation (henceforth MDR).

1, and Annex I, specifically "on the concept of a safety component"[79]. In itself the consultation does not provide any guidelines on how to interpret the concept of safety component, nor how to balance possible misalignments between the definition of the AIA and those of other UHL, but it shows that the commission recognizes that there might be problem with the concept of safety component.

At the same time, the lack of certainty arising from the AIA could also be compensated by a systematic interpretation of those concepts, deriving guidance from existing definitions of safety components, as recalled in UHL.

The table below provides a comparative overview of the different notions emerging in the UHL recalled in Annex I. However, please note that most of the UHL referred to in Annex I do not provide a definition of safety component, but refer to concepts that allow the product regulated to comply with the requirements set by that specific directive or regulation. The purpose of the table is also to highlight potential convergences and divergences in the wording and notions technical practitioners are already trained in tackling through their work.

[79] European commission, *Commission Launches Public Consultation*, cit.

Legislation	Definition of "Safety Component" (or note if none)	Safety function	Independently placed on the market	Failure endangers the safety of persons	Unnecessary for functioning
AI Act	«'safety component' means a component of a product or of an AI system which **fulfils a safety function** for that product or AI system, or the failure or malfunctioning of which **endangers the health and safety of persons or property**»				
Directive 2006/42/EC (MD), Art 2, lit. c)	«'Safety component' means a component: (a) **serving a safety function**, (b) **independently placed on the market**, (c) whose **failure endangers the safety of persons**, and (d) which is **not necessary for normal functioning** or can be replaced by standard part»				
Regulation (EU) 2023/1230 (MR), Art 3, para. 17	«'Safety component' means a physical or digital component, including software, designed to fulfil a **safety function**, **independently placed on the market**, whose failure or malfunction **endangers the safety of persons**, and which is **not necessary** in order for the machinery or related product to function, **or can be replaced with a component that does not fulfil a safety function**»				
Regulation 2016/424, (Cableway Installations, CIR)	«safety component' means any component of equipment or any device intended to be incorporated into a subsystem or a cableway installation for the purpose of ensuring a **safety function**, the failure of which **endangers the safety or health** of passengers, operating personnel or third parties»				
Directive 2009/48/EC (TSD)[80]	No definition of "safety component". Manufacturers are required to protect the health and safety of consumers (Art 4, para. 4, and 10 TSD). Annex II provides a list of specific requirements the manufacturer must observe				
Directive 2014/33/EU (LD)[81]	Refers to "safety components for lifts" (art. 8 and Annex III, LD) and provides a list of safety components. Refers also to the application of the MD's Annex I when it comes to essential health and safety requirements				

[80] European Parliament and the Council of the European Union, Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys, OJ L 170, 30.6.2009, 1,1.

[81] European Parliament and the Council of European Union, Directive 2014/33/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to lifts and safety components for lifts, OJ L 96, 29.3.2014, 251.

Directive 2014/53/EU (RED)[82]	No definition of "safety component." Establishes essential requirements at product/system level. The RED refers to the safety requirements established by the LVD	✗	✗	✗	✗
5. Regulation 2017/745 (MDR)[83]	No definition of the concept of "safety component." The regulation uses "accessory" and "component" but without legal definition of safety-specific components	✗	✗	✗	✗

The MD and its successor, the MR represent the exception amongst the UHL listed in Annex I in that they both offer a definition of safety component, based on 4 elements, three of which have what we may consider a technical nature. Indeed, the component needs to have a safety function lit. a), its failure endangers the safety of persons lit. c) but at the same time, it is not necessary for the functioning of the product lit. d). The fourth is, instead, an exception, absent in all other definitions, and refers to independent the placing on the market lit. d), that is to be understood as the possibility that the component is sold onto the market as such. That further reference does not add, however, to the elements that would at least theoretically allow for the identification of those components that indeed play a safety role, and needs not be further commented upon.

The key elements that appear to have inspired the solution enacted by the AIA, as well as that by the CIR, are the reference to both the safety function and the consequence of failure. Those criteria, however, appear again purely circular, incapable of providing any relevant indication. On the one hand, coordination between those definitions appears quite straightforward, in as much as a safety component for the purposes of the AIA will most certainly be considered such also by the MD and MR. The missing reference to the requirement laid down by lit. b) appears, in fact, irrelevant. On the other hand, we may not look at the wording of those definitions to gain a better understanding of what exactly falls under the scope of art. 6, para. 1, lit. a), but maybe the current applications of those, long existing pieces of legislation could indeed provide some guidance.

Radically different approaches are instead adopted by more specific pieces of legislation, such as the LD, that by referring to a very narrowly defined category of homogeneous products – lifts – may replace ample and underdefined concepts with a practical list of components that undoubtedly

[82] European Parliament and the Council, Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC, OJ L 153, 22.5.2014, 62.

[83] European Parliament and the Council, Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, OJ L 117, 5.5.2017, ibid.

serve a safety function.
In contrast, to legislation that provides precise definitions of safety-related terms, certain sector-specific regulatory frameworks—such as the MDR and the RED—adopt a broader and less prescriptive approach to safety terminology. For example, the MDR establishes, in Annex I, a set of "general safety and performance requirements" that apply to medical devices placed on the EU market (art. 5, para. 2, MDR). These requirements encompass a wide array of considerations related to intended use, risk management, and clinical evaluation of medical devices, rather than defining safety in narrow or technical terms.
Similarly, the RED does not offer a specific definition of a "safety component." Instead, it explicitly refers to the safety objectives articulated in the Low Voltage Directive (LVD[84]) as sufficiently comprehensive to address the safety aspects of radio equipment. Recital 7 of the RED states that «the objectives with respect to safety requirements laid down in Directive 2014/35/EU are sufficient to cover radio equipment, and should therefore be the reference and made applicable by virtue of this Directive». This cross-referencing illustrates the EU's regulatory strategy of leveraging existing frameworks to avoid redundancy and promote consistency across sectors.
The LVD itself outlines, in Annex I, general principles of safety, which include protection against hazards originating from the electrical equipment as well as those that may arise due to external influences acting upon the equipment. These principles are intentionally broad, focusing on outcomes rather than prescribing specific technical measures. The use of such generalized language reflects a regulatory emphasis on risk prevention and functional safety outcomes across a wide variety of electrical products.
Particularly relevant is the definition advanced by the MR, both for its greater complexity and sophistication and for the importance of that piece of legislation, applying to possibly the broadest spectrum of products[85] among all those recalled under Annex I. In particular, the definition formulates an "indifference condition", whereby the component is intended for safety purposes if it «[…] is not necessary in order for that product to function or for which normal components may be substituted in order for that product to function»[86]. So intended, a safety component is separate from the product, regardless of its necessity for the product's performance[87]. The concept here is that the component might improve safety, but is not necessary in order for that product to function[88].
If we applied this notion to the "example 1" above, we could doubt wheth-

[84] European Parliament and the Council, Directive 2014/35/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of electrical equipment designed for use within certain voltage limits, OJ L 96, 29.3.2014.

[85] Indeed, pursuant to art. 3, para. 17, MR it applies to all machinery defined by art. 3, para. 1 MR.

[86] European Parliament and the Council, *Regulation (EU) 2023/1230*, cit.

[87] T. De Graaf – G. Veldt, *The AI Act and Its Impact on Product Safety, Contracts and Liability*, in *European Review of Private Law*, 2022, 803.

[88] European Parliament and the Council, *Regulation (EU) 2023/1230*, cit.

er the braking system is replaceable, or the AI-assisted vision system. Most likely both could be deemed essential for the functioning – given that the purpose of the machine is to navigate the environment without colliding with obstacles – and thence could possibly not be deemed safety components for the purposes of art. 2, lit. c) MD. Moreover, even if we were to consider the indicative list of Annex V, MD, it does not seem to directly address «example 1». Obviously, the list is indicative and the argument could be made to consider such systems safety components. That would be reinforced by an analysis of Annex II, MR – the successor of the MD – where the indicative list cites (18) Software ensuring safety functions and (19) Safety components with fully or partially self-evolving behaviour using machine learning approaches ensuring safety functions. While this might be an indicative that AIS would fall into the definition of safety component. This doesn't seem to be an improvement, as it still rely on the circular definitions of safety components and safety functions.

In relation to the theories regarding safety, authors have suggested that system safety in general is intertwined with some factors, including the redundancy of components that can perform each critical function and the existence of fail-safes[89]. Elevator brakes are a classic example of a fail-safe feature. They are attached to the outside of the cabin and are held open only by the tension in the cables that the cabin is suspended on. If tension is lost in the cables, the brakes automatically clamp shut onto the rails in the elevator shaft. This means that, even if the cables break, the brakes should prevent the cabin from falling; even if the system fails in its function, it should at least be safe.

While easy cases certainly exist[90], relevant stakeholders[91] urged the EC to clarify this notion, so as to prevent an expansive interpretation of the h-AIS category (see §3.1 *supra*), and feedback is expected on this very point to be found in the forthcoming guidelines on art. 6, due on February 2nd, 2026[92].

4.3. Safety Components and Complex Systems

Finally, even if the notion of safety component were clear, it would be most complicated to understand when, indeed, a given application is the safety component of another, in all those systems that comprise more

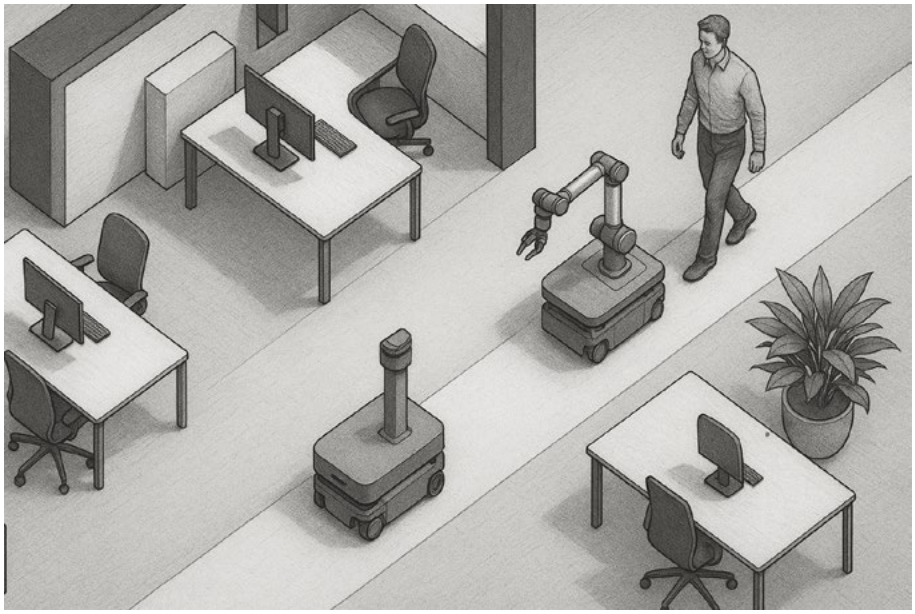
[89] D. Hendrycks, *Introduction to AI Safety, Ethics, and Society*, 2024, 188.

[90] A prime example Microsoft' project InnerEye, which uses machine learning for automatic delineation of healthy anatomy and tumors. Microsoft Research, *Project InnerEye - Democratizing Medical Imaging AI*, in *https://www.microsoft.com/en-us/research/project/medical-image-analysis/*; For an overview on the academic production on the use of various AIS on medical diagnosis refer to S. Kaur – others, *Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives*, in *IEEE*, 2020.

[91] Industries regulated by the RED have manifested their discontent with the expansive interpretation provided by the EC. L. Bertuzzi, *Device Makers Using AI in Safety Features Could Have Strict AI Act Obligations, EU Commission Says*, in *mlex.com*.

[92] L. Bertuzzi, *Product Makers to See AI Act Interplay with EU Safety Rules in Future Guidelines*, in *mlex.com*.

than one machine in themselves, or anyway see the collaboration between multiple machines. Indeed, Industry 4.0 scenarios[93], based in the adoption of embedded and non-embedded applications, are going to heavily rely on those forms of technological solutions, heavily relying on integration.
To exemplify this concept, please consider the following case where more co-bots (collaborative robots) are allowed to roam in an office space or factory, while connected to a radar that uses an algorithm to detect and categorize potential obstacles, thence ultimately allowing the robots to avoid collision (see Figure 1).



This setup raises important questions: should the radar and processing unit alone be classified as the h-AIS? Should the robots also be included in this classification? Or should the entire integrated system—robots and radar together—be evaluated under art. 6 of the AIA? The outcome of this classification has significant implications, particularly in identifying which parties are subject to the obligations outlined in art. 8 and beyond of AIA.
One conservative way of interpreting it is that only the external unit, if sold separately, is the safety component. Everything else, even if it implements safety functions, is machinery. However, the opposite interpretation is also plausible.
A separate entity, known as the system integrator, might then combine these components to meet a client's specific needs. If each robot were considered a high-risk system, the number of responsible parties would grow considerably. If, on the other hand, only the radar and processing system were deemed high-risk, responsibility would fall solely on their developers, providers, and deployers. In a third scenario, where only the integrated system as a whole is treated as high-risk, liability would rest primarily with the system integrator. This ambiguity once again highlights the serious challenges posed by the current regulatory uncertainty.

[93] H. Lasi – Others, *Industry 4.0*, in *Business & Information Systems Engineering*, 2014, 239 ff.

4.4. The Certification of the AIS or of its Safety Component: an Overview

Even if the system using AI was indeed a safety component, for the AIS to be deemed high risk the AIA sets an additional requirement, namely that of that very component—or the overall AIS—being subject to a TPCA under one of the listed pieces of European product safety legislation (UHL, see Annex I).

However, in order to dive into the interpretation of art. 6, para. 1, lit. b) it is necessary to clarify some essential concepts about product certification, a requirement under the so called the New Approach to Market Integration of the European Union, intended to ensure the essential safety of all products before they are distributed within the EU[94].

The process begins with identifying the applicable directive or regulation for a given product. The EU has established multiple directives and regulations that define specific requirements for different categories of goods. For example, the MD applies to industrial and consumer machinery, LVD governs electrical equipment, and the MDR ensures the safety and performance of medical devices. Other directives include those related to toys, electromagnetic compatibility, and radio equipment. A manufacturer must first determine which piece of European legislation applies to their product before proceeding further[95].

Then the manufacturer must assess the conformity route required to demonstrate compliance. Depending on the product type and its associated risk, certification may be achieved through self-assessment or require the intervention of a Notified Body (henceforth NB)[96]. Products such as a LED lamp under the LVD can often be self-certified (art. 12-14 and Annex I, LVD). However, items with critical safety requirements, like a life-supporting medical ventilator, require third-party verification to ensure their safety and reliability (art. 52, para. 1, and Annex VIII MDR).

Following this, a comprehensive conformity assessment must be carried out. This process involves conducting a risk assessment to identify potential hazards, testing the product—most commonly, even if not necessari-

[94] Your Europe, *CE Marking*, in *europa.eu*.

[95] G. Ballor, *CE Marking, Business, and European Market Integration*, in *Business History Review*, 2022, 77 ff.

[96] NBs are independent organizations designated by EU member states to assess the conformity of certain products before they enter the EU market, ensuring compliance with safety, health, and environmental standards. These bodies are authorized under specific EU directives and regulations, often for high-risk products like medical devices (art. 35 MDR) and machinery (art. 30 MR). NBs play a key role in the CE marking process, which signifies that a product meets EU regulatory requirements. Their primary function is to perform conformity assessments which may involve testing, inspecting, and auditing products or manufacturing processes. The AIA has followed previous EU regulations and directives and introduced the role of NB in its Section 4. They are responsible for assessing «the conformity of high-risk AI systems in accordance with the conformity assessment procedures set out in art. 43» (art. 34, para. 1 AIA). The AIA also reproduces the regulatory standard of linking the CE mark to the certification performed by the NBs (art. 48, para. 4, AIA). See also J.-P. Galland, *The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies*, in *European Journal of Risk Regulation*, 2013, 365 ff.

ly—against harmonized standards, and compiling the necessary technical documentation[97]. The technical file should include a detailed product description, information about its design and manufacturing process[98], results from risk analysis, and test reports proving compliance with the relevant EU standards[99]. For example, a children's toy must undergo rigorous mechanical and chemical safety assessments to ensure it does not present a choking hazard and is devoid of toxic substances[100].

A TPCA will differ from the self-certification in the involvement of a NB, that will be responsible for «one or several of the following activities: calibration, testing, certification and inspection»[101]. The extent to which a notified body will be involved depends on the module of certification that is being used—as per the modules defined in Decision No 768/2008/EC. Furthermore, each piece of legislation can adopt changes to those modules, in order to better reflect the necessities of the regulated product. For instance, an EU-type examination for the RED has taken the following shape: «a conformity assessment procedure in which a notified body examines the technical design of the radio equipment and verifies and attests that the technical design of the radio equipment meets the essential requirements set out in Art. 3» (Annex III, RED). The MDR adopts a similar definition, but specifies, amongst other things, that «review the clinical evidence presented by the manufacturer in the clinical evaluation report» (Annex X, para. 3, lit. c), MDR). In sum, for the execution of a TPCA, a notified body will follow the guidelines provided in the specific legislation, that reflect the Decision No 768/2008/EC.

Once compliance has been established, the manufacturer must prepare and sign a European Declaration of Conformity[102]. This formal document affirms that the product meets all applicable EU directives and/or regulations, and must be retained for at least ten years[103]. The next step is affixing the CE marking on the product, which signifies that it conforms to legal safety requirements[104]. For instance, a power drill that meets the requirements of the MR must display the CE marking on its body or packaging before being placed on the market (art. 24, para. 2, MR).

After a product enters the market, manufacturers must ensure ongoing compliance. This includes post-market surveillance, addressing any safety concerns that arise, and making necessary updates to the technical documentation in case of design modifications[105]. If a defect is identified in a product, such as a malfunction in a medical syringe pump covered under the MDR, the manufacturer is responsible for initiating a recall and noti-

---

97 Your Europe, *Conformity Assessment*, in *europa.eu*.

98 S. Rajput, *EU MDR and IVDR Conformity Assessment Guide*, Celegence, 2024.

99 Your Europe, *Technical Documentation and EU Declaration of Conformity*, in *europa.eu*.

100 European Council, *Toy Safety*, in *consilium.europa.eu*.

101 European Commission, *The "Blue Guide" on the implementation of EU products rules*, Brussels, 2016.

102 Your Europe, *Technical Documentation*, cit.

103 *Ibid.*

104 Your Europe, *Conformity Assessment*, cit.

105 Your Europe, *Technical Documentation*, cit.

fying regulatory authorities (art. 10, para. 12 MDR).
Despite the comprehensive nature of the certification process, certain products are exempt from CE marking requirements. This might be the case both for products that are not covered by EU directives and may require national approvals instead, as well as custom-made equipment, designed solely for in-house use within a company, which is not made available on the open market. Furthermore, military equipment and other specialized items used exclusively for national defence purposes are often exempt from CE marking requirements.

4.5. When is a TPCA "Required"?

Art. 6, para. 1, lit. b), AIA, formulating the second condition for an AIS to be deemed high-risk, recites:

> «the product […] is required to undergo a third-party conformity assessment […] pursuant to the Union harmonisation legislation listed in Annex I».

The choice of the verb "required"[106] appears sufficiently clear in restricting the scope of h-AIS to those instances where a TPCA is mandated, and not merely optional, among alternative certification techniques.
The Commission, however, hinted that it might adopt an expansive interpretation to encompass all such cases where a TPCA is merely an option for compliance, but the choice is left with the producer, who may resort to alternative solutions to certify its product, such as self-certification[107]. Such cases are, indeed, all but uncommon[108].
Debates already occurred with respect to the RED and TSD respectively. As per the former, the Commission maintained that demonstrating com-

[106] Similar considerations may be drawn taking other official translations into account, such as the Italian—where «*è soggetto a una valutazione della conformità da parte di terzi*»—, the French—«*est soumis à une évaluation de conformité par un tiers*»—, and the German—«*muss einer Konformitätsbewertung durch Dritte*»—where it is clear that the TPCA is not optional for the manufacturer to choose among potential alternative certification procedures. In other instances, instead, (See for instance the Italian translation of art. 6, para. 3, co. 3), divergences in the translations and their potential meaning may be observed, that have a relevant baring on the interpretation and possible applications of the norm. For a discussion please allow reference to A. Bertolini, *La definizione di sistemi di intelligenza artificiale*, in S. Orlando – G. Cerrina Feroni (eds.), *Il regolamento europeo sull'intelligenza artificiale: problemi e sistema*, forthcoming 2025. It shall also be noted that the same interpretation was reached by stakeholders regulated by the RED. L. Bertuzzi, *Device Makers Using AI in Safety Features Could Have Strict AI Act Obligations, EU Commission Says*, cit.

[107] The EC has elucidated its stance on the matter through the EU AI Act Compliance Checker, the novel instrument of the AI Office. Once the self-assessment has begun and you have reached the question of whether the AIS can be considered high risk pursuant to art. 6, para. 1 AIA, you must specify whether the product, or the product in which the AIS is intended to be used as a safety component, is required to undergo a TPCA under sectoral law. Well, it is precisely at this point that the EC expressly warns that this also includes products for which you can opt out of a TPCA when harmonized standards are fully applied. See ai-act-service-desk.ec.europa.eu.

[108] See, for instance, art. 25, para. 3, lit. a) MR; art. 20, para. 2, and 22, para. 2, TSD and art. 17, para. 3, RED.

pliance through harmonized standards, when conceived as an alternative to a TPCA (art. 17, para. 2, lit. 3), RED), would not *per se* exclude the qualification of the system as h-AIS[109]. As per the latter, the proposed Toy Safety Regulation (TSR), expressly clarifies that «The choice by the manufacturer of the conformity-assessment procedures for such toys, where there is a possibility to opt out of third-party conformity assessment where harmonized standards have been applied, should not affect this classification as high-risk AI system», causing manufacturers to be raise concerns about the impact of the regulation[110].

The Commission has also spoken regarding the interactions of the AIA, the MDR and the IVDR. It has issued a Q&A document[111] aimed at clarifying how those three pieces of EU law should interact. The document presented the table below, which seems to indicate that most of the products regulated by those legislations are to be considered h-AIS. The Q&A document also explicitly notes that, according to Recital 51 AIA, the fact that a product is considered a h-AIS, does not mean it will be considered high- risk for the purposes of the MDR and IVDR[112]. For instance, for the MDR devices are classified according to their level of risk (Class I, Class IIa, Class IIb and Class III)[113]. Class IIa devices are considered to be medium risk, however they still need to undergo a third party conformity assessment[114], thus being subject to h-AIS regime.

Classification	Notified Body Involved?	AIA High-Risk (Art. 6(1)) conditions fulfilled?
MDR Class I (non-sterile, non-measuring, non-reusable surgical)	❌ No	❌ No
MDR Class I (sterile, measuring, reusable surgical)	✅ Yes	✅ Yes
MDR Class IIa, IIb, III	✅ Yes	✅ Yes
MDR Annex XVI[10]	✅ Yes	✅ Yes
IVDR Class A (non-sterile)	❌ No	❌ No
IVDR Class A)	✅ Yes	✅ Yes
IVDR Class B, C, D	✅ Yes	✅ Yes
In-house device according to Art. 5(5) MDR/IVDR	❌ No	❌ No

[109] L. Bertuzzi, *Device Makers Using AI in Safety Features Could Have Strict AI Act Obligations, EU Commission Says*, cit.

[110] L. Bertuzzi, *Manufacturers of Toys Using AI Set to Face Strict AI Act Regime (Correct*)*, in *mlex.com.*

[111] Directorate-General for Health and Food Safety, *MDCG 2025-6 - FAQ on Interplay between the Medical Devices Regulation & In vitro Diagnostic Medical Devices Regulation and the Artificial Intelligence Act (June 2025)*, in *health.ec.europa.eu.*

[112] MLEX, *EU Q&A Document on AI Act, Medical-Device Rules Interplay up for Final Approval*, in *mlex.com.*

[113] European Commission, *Medical device classification*, in *webgate.ec.europa.eu.*

[114] Art. 52, para. 6, MDR determines that «Manufacturers of class IIa devices, other than custom-made or investigational devices, shall be subject to a conformity assessment as specified in Chapters I and III of Annex IX». On its turn Annex IX, Chapter I and III require the involvement of a notified body. Which, as we have seen, triggers the application of art. 6, para. 1.

If required were to mean an inescapable legal obligation—as the most logical interpretation of the AIA would suggest—, the result would be that only a limited amount of products regulated under UHL would be considered high-risk, as those directives and regulations provide different avenues through which the conformity of a product can be assessed, some of which, do not involve the use of notified body, thence a third-party conformity assessment. Instead, if we were to consider the term required to include instances where resorting to NB is among the possibilities, then every product under a UHL would have to be considered high-risk. As a matter of example, let's look into the MR and how it has employed third-party conformity assessments.

Art. 25 MR establishes that products that fall under its scope will be regulated according to procedures (called modules) designated in its annexes:

> «2. Where the category of machinery or related product is listed in Annex I, Part A, the manufacturer or the natural or legal person referred to in art. 18 shall apply one of the following procedures:
> (a) EU type-examination (module B) set out in Annex VII, followed by conformity to type based on internal production control (module C) set out in Annex VIII;
> (b) conformity based on full quality assurance (module H) set out in Annex IX;
> (c) conformity based on unit verification (module G) set out in Annex X.
> 3. Where the category of machinery or related product is listed in Annex I, Part B, the manufacturer or the natural or legal person referred to in art. 18 shall apply one of the following procedures:
> (a) internal production control (module A) set out in Annex VI;
> (b) EU type-examination (module B) set out in Annex VII, followed by conformity to type based on internal production control (module C) set out in Annex VIII;
> (c) conformity based on full quality assurance (module H) set out in Annex IX;
> (d) conformity based on unit verification (module G) set out in Annex X.»

Clearly, art. 25 MR allows an option to the manufacture to choose between the modules, while still being able to demonstrate compliance with the regulation. Among the different modules recalled, only B, H and G unequivocally require the intervention of a notified body, whereas A establishes a procedure that is entirely based on the assessment of the manufacturers themselves[115]. Taking that into account, we can observe how all the options enumerated under para. 2 require the intervention of a NB, whilst para. 3 ensures the manufacture the possibility to resort to self-certification, while still maintaining the choice to opt for a TPCA. Provisions like that appear throughout all the UHL[116]. If we were to argue that the TCPA

[115] Art. 25 MD determines that, regarding Module A, «the manufacturer fulfils the obligations laid down in points 2, 3 and 4, and ensures and declares on its sole responsibility that the machinery or related product concerned satisfies the applicable requirements of this Regulation».

[116] Art. 12 MD; art 25, MR; art. 19 TSD; art. 20, para. 2, and 22, para. 2, Watercraft

is required only when the manufacturer has no option but to conduct such a procedure, this interpretation might be considered too restrictive and that it would leave effectively high-risk products outside the application the AIA. However, such criticism to the literal interpretation of art. 6, para. 1, would appear disproportionate, in as much as those products are already heavily regulated under UHL.

4.6. Case Studies: on the Marginal Discrepancies

Building on the analysis carried out above (see §§ 4.1-4.5), we have conceived a series of three examples of AI-based Systems that are already existent or currently under research. We have construed those examples taking into account the respective technologies, and carried out individual interviews with experts to allow us to determine the correct functioning of those applications, possibly identify their respective application and help us identify what may be deemed a safety component in the specific examples at hand, and determine whether it would be subject to the UHL indicated in Annex I.

Those examples include, a security patrol robot manufactured according to hENs, the humanoid robot with comparable characteristics to that of Boston Dynamics's Atlas[117], and a drone docking station technology in personal watercrafts.

Firstly, a security patrol robot is deemed to be a (semi-)autonomous mobile robot that repeatedly visits or patrols specified areas, equipped with sensors, cameras, and decision-making software, to detect anomalies—such as motion, intrusions, or environmental hazards—and either respond directly or transmit alerts for further action[118].

These robots are typically equipped with a range of advanced sensors, including but not limited to RGB[119] and infrared cameras, microphones, ultrasonic range finders, and environmental sensors (e.g. for temperature, smoke, motion, and light), which enable them to perceive their surroundings, identify suspicious behaviour, and detect intrusions or hazardous conditions.

From a technological standpoint, such robots are capable of autonomous navigation, relying on GPS and RTK[120] systems in outdoor environments

Directive; art. 13, para. 1, Directive 2014/34/EU; art. 17, para. 2, lit. a) RED; art. 14, para. 2, lit. a) and b) Pressure Equipment Directive; art. 19, para. 1 PPE Regulation.

[117] None of the considerations here drawn are to be intended as a binding legal analysis of the specific robot by Boston Dynamics which is, instead, intended as a plausible and realistic example of a technology that could fall under the scope of application of the AIA.

[118] N. Basilico, *Recent Trends in Robotic Patrolling*, in *Current Robotics Reports*, 2022, 65–76.

[119] RGB stands for Red, Green, Blue—the three primary colors of light used in digital imaging and display technologies. In the context of RGB cameras, this term refers to standard digital cameras that capture visual information by detecting the intensity of red, green, and blue light in a scene.

[120] RTK stands for Real-Time Kinematic positioning. It is a satellite navigation technique used to enhance the precision of position data derived from global navigation satellite

and on Simultaneous Localization and Mapping (SLAM) algorithms or track-based navigation for indoor deployment. The decision-making component of these systems is often based on AI techniques, such as convolutional neural networks for facial recognition and object detection, and fuzzy logic or machine learning algorithms for route optimization, obstacle avoidance, and behaviour analysis[121].

In order to determine the classification of the AIS as high risk or not, as outlined in art. 6, para. 1, AIA, it was necessary to legally isolate the safety component.

In accordance with the insights provided by subject matter experts, it is imperative to conceptualize the safety system in its entirety, rather than focusing on its constituent components.

At this juncture, it was necessary to ascertain whether any of the UHLs enumerated in Annex I pertained to this particular robot and whether it was subject to conformity assessment by a notified body (TPCA).

It is our opinion, based on the available evidence, that the MD apply to this robot by virtue of para. 19, Annex IV MD ("protective devices designed to detect the presence of persons").

The aforementioned directive enumerates a series of cases in which a TPCA is requested but only art. 12, para. 4, MD relates to h-AIS in the sense of our perspective (option A) – in other words, the term "required" indicates something necessary and indispensable. Therefore, only AIS that are mandatorily subject to TPCA can be considered high risk. Consequently, a security patrol robot that has not been manufactured in accordance with the hENs and that is subject to EC type-examination or full quality assurance procedure can be classified as high risk (see art. 12, para. 4, MD).

Nevertheless, an issue emerges when the security patrol robot is produced in compliance with the hENs and the latter encompass all pertinent essential health and safety requirements, as stipulated in art. 12 MD. In this particular instance, manufacturers are not obliged to follow a TPCA and, in accordance with our interpretation of the term "required," the robot is not deemed to pose a significant risk.

This perspective is not shared by those who contend that the term "required" signifies a TPCA that is not essential but rather one of the potential methods of conformity assessment that can be implemented (option B).

Secondly, we analysed a humanoid robot with comparable characteristics to Atlas. As that of Boston Dynamics it is a bipedal humanoid robot primarily designed as a research platform for studying agile mobility, human-like movement, and robot-environment interaction in unstructured environments. The robot exemplifies advanced capabilities in dynamic locomotion, real-time perception, and adaptive control. Its design aims to approximate the morphology and movement of the human body, there-

systems (GNSS), such as GPS.

[121] For an overview on security patrolling robot from a technical point of view please see H. Liu – L. Yu – W. Meng – Q. Zhu – Y. Wang, *Patrol planning for a security robot using a spatio-temporal model*, in *Robotics and Autonomous Systems*, 92, 2017, 16–26; J. Choi – D. Kim, *Design and Implementation of an Indoor Patrol Robot System*, in *International Journal of Advanced Robotic Systems*, 17, 2020, 1-11.

by enabling it to perform complex tasks in unstructured environments. The robot is approximately 1.5 meters tall and weighs around 80 to 90 kilograms, dimensions that make it suitable for navigating human-centric spaces such as buildings, staircases, or rough terrain.

Besides the mechanical architecture, which includes 28 degrees of freedom (DoF), the robot integrates an advanced sensor suite to perceive and interpret its environment. This includes stereo vision cameras, LiDAR sensors, and an inertial measurement unit (IMU). These sensors feed into a real-time perception system that constructs a three-dimensional model of the surrounding environment, which is essential for obstacle negotiation, path planning, and terrain adaptation. Visual-inertial odometry enables Atlas to estimate its position and orientation even in the absence of GPS or external tracking systems. Another defining feature is its control architecture, which combines classical control techniques with modern machine learning and model predictive control (MPC). The robot uses, indeed, AI-based algorithms for real-time perception, which are essential for interpreting complex environments[122].

The classification of robots as high risk is contingent upon the delineation of the safety component's perimeter. In light of the expert advice received, the entire system was identified.

At this juncture, the pertinent UHL was ascertained. Consequently, the MD were deemed applicable. In this instance as well, reference should be made to art. 12, para. 4, of the MD by virtue of para. 17 of Annex IV MD ([d]evices for the lifting of persons or of persons and goods involving a hazard of falling from a vertical height of more than three meters).

Given the robot's status as an American robot—as that of Boston Dynamics—for which European production standards are not obligatory nether suggested, its noncompliance with these standards would, in our estimation (option A), readily categorize it as a high-risk Ais if undergoing a TPCA.

Conversely, if even a subset of the European standards were to be observed, the conclusion would be reversed. In accordance with the provisions stipulated in the relevant legislation, Atlas would not be classified as a high-risk entity unless the adoption of (option B) were to be implemented.

Lastly, we took into account a drone docking station technology in personal watercrafts. A drone docking station technology in personal watercrafts (PWCs) refers to an integrated system that allows a drone (usually an aerial drone) to autonomously land, recharge, and possibly transmit data directly from the watercraft. This innovation combines unmanned aerial vehicle (UAV) capabilities with marine vehicles, enhancing their functionality, especially for recreational, rescue, and surveillance applications. It includes a stabilized landing platform designed to compensate for the motion of the watercraft, ensuring safe and accurate autonomous landings. The station integrates a wireless or contact-based charging mechanism powered by the PWC's onboard electrical system, allowing the drone to recharge

[122] Since the robot's characteristics are comparable to those of Atlas, for further technical details, please refer to Boston Dynamics, Atlas® and beyond: the world's most dynamic robots, in bostondynamics.com/atlas/.

without manual intervention. A data transfer module, typically utilizing Wi-Fi, Bluetooth, or USB-C, enables seamless uploading of aerial footage or sensor data to onboard storage or remote servers. The system relies on GPS and AI-assisted navigation algorithms to support autonomous take-off, tracking, and return-to-dock functions, even while the watercraft is in motion. Additionally, the docking station is housed within a marine-grade protective enclosure to resist water, salt, and vibrations, ensuring durability and operational reliability in harsh aquatic environments[123].

The safety component—even here—can be deemed to be the whole system while the applicable UHL is the Directive 2013/53 on recreational craft and personal watercraft (RCPW). Being the example a personal watercraft, the referral is to art. 20, para. 2, RCPW which provides various type of conformity assessment including TPCA.

According to our reasoning, this technology cannot be qualified as h-AISs given that it does not stipulate TPCA as the sole option for conformity assessment (option A).

Nevertheless, the alternative interpretation (option B) posits that the drone self-docking station in personal watercraft could be classified as h-AIS. This is predicated on the premise that the legislation in effect allows for the subjecting of the product to TPCA.

As illustrated by the provided examples, the differentiation between h-AISs and AISs is predicated on marginal discrepancies. In the initial two scenarios, the distinction is determined by adherence to heN, which are not mandatory. In the latter scenario, the distinction primarily lies on the quality of personal or recreational watercraft.

[123] For an overview on technical features on drone self-docking station in personal watercrafts see C. G. Grlj – N. Krznar – M. Pranjić, *A Decade of UAV Docking Stations: A Brief Overview of Mobile and Fixed Landing Platforms*, in *Drones*, 2022.

AI System example	Interpretation of «required» - option A	Interpretation of «required» - option B
Security patrol robot* Para. 19, Annex IV, Directive 2006/42/EC[124] as recalled in art. 12, para. 4[125] same Directive	NOT high-risk	high-risk
Humanoid robot* Para. 17, Annex IV, Directive 2006/42/EC[126]	NOT high-risk	high-risk
Drone docking station technology in personal watercrafts (eg. GODO A170 Dock) Art. 20, para. 2, Directive 2013/53/EU[127]	NOT high-risk	high-risk

Legenda:
Option A: «required» is intended in a strict way. It stands for "TPCA is compulsory".
Option B: «required» is intended in a broad way. It stands for "TPCA is possible".
Method: Examples are extracted from each recalled directive.
*The product is equipped with AI capabilities. In the context of h-AISs qualification, components are regarded as an integrated whole.

[124] Para. 19, Annex IV Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC, OJ L 157, 9.6.2006: «Protective devices designed to detect the presence of persons».

[125] art. 12, para. 4, *Ibid*: «Where the machinery is referred to in Annex IV and has not been manufactured in accordance with the harmonized standards referred to in art. 7, para. 2, or only partly in accordance with such standards, or if the harmonised standards do not cover all the relevant essential health and safety requirements or if no harmonised standards exist for the machinery in question, the manufacturer or his authorised representative shall apply one of the following procedures: (a) the EC type-examination procedure provided for in Annex IX, plus the internal checks on the manufacture of machinery provided for in Annex VIII, point 3; (b) the full quality assurance procedure provided for in Annex X».

[126] Para. 17, Annex IV Directive 2006/42/EC cit.: «Devices for the lifting of persons or of persons and goods involving a hazard of falling from a vertical height of more than three meters».

[127] European Parliament and the Council, Directive 2013/53/EU of the European Parliament and of the Council of 20 November 2013 on recreational craft and personal watercraft and repealing Directive 94/25/EC, OJ L 354, 28.12.2013, 90-131, art. 20, para. 2.

5. A Critical Perspective

In conclusion, the interpretation of art. 6, para. 1, lit. b) of the AIA plays a pivotal role in determining the scope of h-AIS under the new regulatory regime. The crux of the issue lies in the understanding of the term "required" in relation to TPCAs under UHL. A literal reading confines the AIA's applicability to those instances where a TPCA is legally mandatory. This interpretation maintains regulatory clarity and respects the division of responsibilities between the AIA and sector-specific UHL regimes, which have long governed conformity assessment procedures.
However, the Commission's more expansive approach raises significant concerns. Such an interpretation may lead to the overclassification of AI products as high-risk, even where manufacturers legally opt for less stringent conformity pathways. Examples drawn from the RED, TSD, MDR, IVDR, and the MR reveal how prevalent such optional structures are, and how the Commission's broad reading could inadvertently result in an overly burdensome regulatory landscape for AI developers and manufacturers. Moreover, expanding the definition of "required" to encompass optional conformity routes risks undermining legal certainty and predictability. As already seen in the case of stakeholders regulated by the RED, manufacturers are left uncertain as to whether choosing a less burdensome, yet lawful, certification method might still expose them to additional regulatory obligations under the AIA.

6. The Unsolved Riddle of Art. 6 AIA

In the myth of Oedipus, the riddle posed by the Sphinx was ultimately resolved by identifying a coherent metaphor for the stages of human life. By contrast, the riddle embedded in art. 6 of AIA resists such elegant resolution. Rather than culminating in a unifying interpretation, the regulation's provisions surrounding h-AIS remain fragmented and opaque. Like the multiple heads of a hydra or the distorted echoes in a labyrinth, each interpretative effort to define what constitutes a h-AIS under art. 6 gives rise to new ambiguities and regulatory contradictions. What emerges is not a clarified standard, but a spectrum of possible readings that fail to converge on a single, stable answer.
The core of this ambiguity lies in the complex and sometimes contradictory architecture of art. 6 itself, which is divided into distinct yet overlapping prongs: the automatic classification via Annex III (para. 2), the exemptions framed in vague and elastic language (para. 3), the peculiar mechanism of provider self-declaration (para. 4), and the reference to harmonised product legislation (para. 1). Far from offering a harmonised framework, these provisions create a regulatory palimpsest—a layering of legal instruments, exceptions, and interpretative leeway that renders the classification of AIS an exercise in probabilistic reasoning rather than legal certainty. One cannot, *ex ante*, determine with confidence whether a given AIS will be deemed high-risk without navigating a thicket of technical, functional, and semantic variables, often contingent on discretionary as-

sessments by both private and public actors.

The legislative intent to adopt a technology-neutral approach further compounds this uncertainty. The AIA aims to regulate the use of AI regardless of its technical design, foregrounding the notion of risk to fundamental rights, health, and safety as the primary classificatory axis. Yet, by grounding the high-risk classification in Annex III—a list of sector-specific use cases defined not by demonstrable metrics of harm but by broad socio-political considerations—the regulation departs from genuine technology neutrality. It institutes a *de facto* taxonomy of "acceptable" versus "concerning" AI applications, without articulating a transparent risk assessment methodology or a coherent theory of harm. The very notion of "risk" becomes a rhetorical device rather than a measurable criterion. As such, systems are not categorised on the basis of demonstrated or anticipated risk levels, but through a normative framework lacking analytical precision or empirical grounding.

This conceptual fragility is further evidenced in the self-declaration mechanism of art. 6, para. 4, which ostensibly empowers providers to argue that their systems are not high-risk despite falling within Annex III. On the surface, this provision introduces a welcome degree of flexibility and contextual nuance. In practice, however, it exacerbates legal uncertainty. It does so by failing to establish clear criteria, procedures, or authoritative review mechanisms that could validate such claims ex ante. Providers are left to navigate a compliance landscape without a reliable compass, while regulators retain discretion to override these determinations retroactively. This asymmetry not only undermines legal certainty but also risks eroding trust in the regulatory framework as a whole.

Equally problematic is the vague terminology embedded in art. 6, para. 3, which outlines exemptions based on the nature and influence of an AIS's function. Phrases such as «narrow procedural task», «merely prepares input», or «not materially influencing» are inherently contestable and do not benefit from operational definitions. Their application requires context-sensitive judgments that are unlikely to yield uniform outcomes across Member States. This is particularly true in domains where the boundary between assistance and decision-making is blurred by the persuasive authority of algorithmic output—such as in healthcare diagnostics or judicial support tools. The regulation's reliance on such open-textured terms invites divergent national interpretations and fragmented enforcement practices, running counter to the harmonisation objectives of an EU regulation.

Finally, para. 1, which is intended to be the most technical provision, directly linking qualification as an h-AIS to the intrinsic characteristics of the system rather than embedded policy evaluations, still leaves considerable room for interpretation. Indeed, it refers to a potentially ambiguous concept - that of a safety component – and adopts wording that could either lead to an extremely broad application of the h-AIS category or, on the contrary, to a practical abrogation (at least as per para. 1).

The ambiguity of art. 6 is not incidental; it is the systemic by-product of an attempt to graft risk-based regulation onto a technology-neutral legislative foundation. The resulting synthesis, while politically palatable,

is doctrinally unstable and its implementation remains uncertain. A true risk-based framework would necessitate the identification of risk vectors, quantification of potential harm, and an evidence-based methodology for ranking and mitigating those risks. Conversely, a genuinely technology-neutral regulation would require abstract, function-agnostic rules of general applicability, applicable regardless of context. The AIA accomplishes neither. Its hybrid architecture results in a definitional impasse: a system that aspires to universality but defaults to categorical exceptions; that aims to mitigate risk but lacks evaluative tools to determine its thresholds; that seeks harmonisation but breeds interpretative fragmentation.

Thus, unlike Oedipus, the reader of art. 6 cannot solve the riddle by grasping its inner logic—for it has none that is both coherent and complete. The Brussels Sphinx offers no closure but rather ultimately functions as a standing executive oracle for the provisions of the AIA, which carries the potential for bureaucratic proliferation and ambiguity in its implementation. Indeed, it leaves us suspended in a state of radical uncertainty, where developers, regulators, and stakeholders must perpetually renegotiate the boundaries of compliance. This condition is not merely an epistemic inconvenience; it is a structural vulnerability. It jeopardises the rule of law values that EU legislation is meant to uphold – legal certainty, foreseeability, and uniform application – while simultaneously disincentivising innovation and chilling market entry. Until these ambiguities are addressed through interpretative guidance, delegated acts, or eventual judicial review, art. 6 will remain less a cornerstone of AI regulation than its most impenetrable enigma.

Abstract

The article examines the conceptual and normative "riddle" posed by art. 6 of the EU Artificial Intelligence Act (AIA) in defining "high-risk" AI systems (h-AISs). It argues that the combination of a horizontal, technology-neutral framework with a risk-based classification generates significant interpretative uncertainty and undermines legal certainty. After situating the AIA within the broader EU product-safety regime and the New Legislative Framework, the contribution meticulously examines in detail the critical issues arising under Art. 6 AIA. These range from para. 2 recalling the Annex III list of high-risk AI systems, which does not rest on an objective assessment of risk, to the exceptions in paras. 3 and 4, and the cross-reference to Union harmonisation legislation in Annex I. Particular attention is paid to contested notions such as "safety component" and "third-party conformity assessment required", illustrated through case studies (e.g. security mobile robots, humanoid robots, drone docking stations). The article concludes that this unstable definitional architecture undermines consistent application, equal treatment across sectors, and effective incentives for innovation.

Keywords

Artificial intelligence – AI Act – risk-based approach – high-risk AI systems – interpretative uncertainties