Decision Problems for Regulatees under the EU AI Act: Contested Values, Uncertain Evidence, and the Limits of Standardisation

by **Alessio Tartaro, Arvin Obnasca and Enrico Panai ***

Abstract: Based on the New Legislative Framework (NLF) approach, the AI Act relies on harmonised standards to provide technical specifications for the implementation of its essential requirements for high-risk AI systems. This paper argues that these requirements pose fundamental "decision problems" for regulatees, requiring value judgments and evidence assessment under uncertainty for their implementation. Unlike mature NLF fields with established methodologies and value consensus, the AI domain is characterised by contested values and uncertain evidence, significantly limiting the ability of traditional standardisation to provide clear, universally applicable solutions. This creates uncertainty for regulatees, complicates conformity assessment and enforcement, and risks undermining the overall regulatory effectiveness of the AI Act. In response to these challenges, this paper proposes supplementing standardisation with a procedural approach, such as documented "AI Act Compliance Cases," to compel transparent articulation and justification of regulatees' decisions. This enhances auditability and manageability, bolstering the Act's capacity to achieve its health, safety and fundamental rights objectives despite inherent complexities.

Keywords: AI Standardisation, AI Act, New Legislative Framework, AI Risks, Decision Problems

Recommended citation: Alessio Tartaro, Arvin Obnasca and Enrico Panai, Decision Problems for Regulatees under the EU AI Act: Contested Values, Uncertain Evidence, and the Limits of Standardisation, 17 (2026) JIPITEC 9 para 1.

I. Introduction

1 What constitutes an *acceptable level* of risk for AI systems? When is their *level of accuracy appropriate* for their intended purpose? What *measures of human oversight* are *commensurate* with the characteristics of an AI system, such as its level of autonomy and context of use? With the accelerating integration of AI into myriad aspects of private, social, and economic life, finding robust and actionable answers to these questions is becoming ever more pressing. Indeed, the ability to provide satisfactory responses to these questions may well mark the difference between AI that benefits society and AI that has a detrimental impact.

2 In this context, the AI Act has emerged as the first comprehensive legal framework specifically designed to regulate AI (superscript 1), in order to maximise its benefits and minimise its risks. Consequently, questions like those posed at the opening of this paper are not merely abstract philosophical

* Alessio Tartaro - University of Sassari, a.tartaro@phd.uniss.it
Arvin Obnasca - University of Sassari, a.obnasca@studenti.uniss.it
Enrico Panai - Catholic University of the Sacred Heart, enrico.panai@unicatt.it

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

concerns but are central to the very fabric of the AI Act. They lie at the heart of the European vision of "Trustworthy and human-centric AI," an approach that explicitly aims to align the design, development, and deployment of AI technologies with established EU values and fundamental rights protection[2].

3 The primary burden of translating these high-level aspirations into concrete operational practice and demonstrating compliance with the Act's requirements falls squarely on the shoulders of regulatees, primarily providers of high-risk AI systems. For these actors, effectively navigating the AI Act requires addressing questions such as the acceptable level of risk (Article 9), the appropriate level of accuracy (Article 15), or the commensurability of human oversight measures (Article 14), in order to ensure compliance with the Regulation.

4 However, the AI Act itself does not provide direct, concrete answers to those questions. Regarding the level of risk, for instance, it merely states that the "overall residual risk of the high-risk AI systems is judged to be *acceptable*" (Article 9(5)). Concerning accuracy, Article 15(1) stipulates that "High-risk AI systems shall be designed and developed in such a way that they achieve an *appropriate level of accuracy*" and that "they perform consistently in those respects throughout their life cycle". On the crucial matter of human oversight, Article 14(1) mandates that "the oversight measures shall be *commensurate* with the risks, level of autonomy and context of the use of the high-risk AI system" (Article 14(1))[3]. This openness of the AI Act's core requirements generates significant interpretative, compliance, and enforcement uncertainty.

5 Since the AI Act relies on the New Legislative Framework (NLF)[4], a considerable expectation has arisen that these harmonised standards, developed by the European Standardisation Organisations (ESOs), will "come to the rescue" by providing the essential technical specifications and methodologies needed to concretise the Act's requirements[5]. This expectation is formally codified in the AI Act, where Article 40(1) confers a "presumption of conformity" on systems that comply with such standards. By making adherence to standards the principal means for providers to demonstrate that they meet the Act's legal obligations, the Regulation delegates a crucial role to the standardisation process, "kicking the can down the road" as a result[6]. This reliance on standards to bridge the gap between essential requirements and concrete technical implementation is a cornerstone of the NLF[7]. However, while this NLF model has proven successful in numerous other domains[8], this paper will argue that its application to the field of AI may not yield the anticipated effectiveness.

6 The core reason for this anticipated ineffectiveness lies in the nature of the questions regarding acceptable risk levels, appropriate accuracy levels, and commensurate human oversight measures, questions that regulatees are called upon to address with the support of harmonised standards. These problems can be effectively characterised as "decision problems", that is, problems requiring a choice between alternative courses of action, each with its own set of potential outcomes, benefits, and drawbacks[9]. For example, deciding an acceptable false positive rate for an AI system used in medical diagnosis requires balancing the risk of missing a condition against the risk of unnecessary delays in diagnosis, a choice influenced by varying clinical philosophies, resource constraints, and differing risk tolerance. Similarly, determining the "adequate" level of human oversight for an autonomous vehicle requires balancing the potential efficiency gains and convenience of automation against potential safety risks and the complex moral hazard considerations associated with shifting responsibility from human to machine. Such decisions depend heavily on prevailing societal values regarding safety, human agency, technological trust, and the

2 AI HLEG, 'Ethics Guidelines for Trustworthy AI' <https://ec.europa.eu/futurium/en/ai-alliance-consultation> accessed 4 March 2026.

3 Italics are ours.

4 Sybe de Vries, Olia Kanevskaia and Rik de Jager, 'Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation' (2023) 8 European Papers - A Journal on Law and Integration 583.

5 Josep Soler Garrido and others, 'Harmonised Standards for the European AI Act' (*JRC Publications Repository*, 2024) <https://publications.jrc.ec.europa.eu/repository/handle/JRC139430> accessed 27 May 2025.

6 Johann Laux, Sandra Wachter and Brent Mittelstadt, 'Three Pathways for Standardisation and Ethical Disclosure by Default under the European Union Artificial Intelligence Act' (2024) 53 Computer Law & Security Review 105957.

7 Stéphane du Boispéan, Markus Mueck and Christophe Gaie, 'Introduction to the European New Legislative Framework' in Markus Mueck and Christophe Gaie (eds), *European Digital Regulations* (Springer Nature Switzerland 2025) <https://doi.org/10.1007/978-3-031-80809-8_1> accessed 28 April 2025.

8 European Commission, 'Commission Staff Working Document: Evaluation of the New Legislative Framework' SWD(2022) 365 Final (16 November 2022) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12654-Industrial-products-evaluation-of-the-new-legislative-framework_en> accessed 18 February 2026.

9 B Fischhoff and others, 'Approaches to Acceptable Risk: A Critical Guide' (1980) NUREG/CR-1614, ORNL/Sub-7656/1, 5045395 <http://www.osti.gov/servlets/purl/5045395/> accessed 6 November 2024.

acceptability of different types of errors. For high-risk AI systems listed in Annex III of the AI Act, the ultimate determination of compliance, including judgments about "acceptable risk" or "appropriate accuracy," rests with the regulatees, primarily AI providers themselves, through internal conformity assessment. For high-risk AI systems referred to in Annex I (i.e., products or safety components of products covered by sectoral legislation), this judgment will often involve notified bodies[10]. Successfully addressing these decision problems is therefore a key element for all regulatees aiming to achieve and demonstrate compliance with the AI Act, for competent authorities enforcing it and for notified bodies assessing conformity.

7 An essential characteristic of such decision problems is their inherent dependence on two intertwined factors: values (what is considered desirable or undesirable, important or trivial) and evidence (information about the likely outcomes of different choices)[11]. In other words, a combination of (often implicit) value judgments and available empirical evidence influences the choice between the different alternatives that define any given decision problem. In the rapidly evolving field of AI, however, a critical complicating factor emerges: the relevant values are frequently contested, pluralistic, and subject to ongoing societal debate (constituting an ethical problem)[12], and the evidence regarding AI system performance, reliability, and broader societal impact is often weak, incomplete, uncertain, or context-dependent (constituting an epistemological problem)[13]. This combination presents a significant obstacle to defining a common, universally accepted way of handling these decision problems in the AI domain, particularly for regulatees seeking clear, predictable, and defensible pathways to compliance.

8 As we shall see, this situation contrasts sharply with decision problems in many other, more mature fields regulated under the NLF. In those sectors, although values and uncertainty invariably play a role, commonly accepted methods, data collection and evaluation practices, and established professional norms are often codified into harmonised standards to guide the identification of the most acceptable alternative and thus finding a broadly satisfactory solution to the decision problem at hand. This disparity has direct and profound implications for what standards can realistically achieve in supporting the implementation of the AI Act, calling into question the anticipated role of harmonised standards in facilitating compliance and supporting the achievement of the Regulation's ambitious objectives.

9 The remainder of this paper will develop arguments to support these claims. While acknowledging the extensive literature already emerging on the AI Act and the role of standards showcased in this special issue, this paper focuses specifically on the "decision problems" faced by regulatees. Section 2 will elaborate on the nature of decision problems and demonstrate how many of the essential requirements for high-risk AI systems are indeed of this type, posing direct and often underappreciated challenges for regulatees. Section 3 will then delve deeper into why these decision problems are particularly difficult to handle in the field of AI, highlighting the pervasive divergences in values and the inherent uncertainty of evidence that characterise this domain, illustrated with a use case. Section 4 will further underscore these challenges by contrasting the AI context with other NLF-regulated fields where decision problems, while present, are often more robustly addressed by standards due to greater epistemic and normative consensus. Section 5 will examine the regulatory implications of these limitations for the AI Act and propose a procedural documentation structure, similar to assurance cases, and clarified roles of technical standards to manage inherent decision problems.

10 See Irene Kamara's paper in this special issue.

11 Baruch Fischhoff, 'Acceptable Risk: A Conceptual Proposal' (1994) 5 RISK: Health, Safety & Environment (1990-2002) <https://scholars.unh.edu/risk/vol5/iss1/3>.

12 Catharina Rudschies, Ingrid Schneider and Judith Simon, 'Value Pluralism in the AI Ethics Debate – Different Actors, Different Priorities' (2020) 29 The International Review of Information Ethics <https://informationethics.ca/index.php/irie/article/view/419> accessed 30 May 2025.

13 Rishi Bommasani and others, 'A Path for Science- and Evidence-Based AI Policy' <https://understanding-ai-safety.org/> accessed 27 March 2026.

II. Essential Requirements and Decision Problems in the AI Act

10 In line with the NLF underpinning its structure[14], the AI Act articulates essential requirements for high-risk AI systems and delegates their technical elaboration to harmonised standards[15], which Article 40 regards as providing a presumption of conformity[16]. These essential requirements are outlined in Articles 8 to 15 of Title III, Chapter 2, which include high-quality data and data governance, comprehensive technical documentation, robust record-keeping, transparency and provision of clear information, effective human oversight mechanisms, and appropriate levels of accuracy, robustness, and cybersecurity. Article 9 is particularly important because it requires a continuous, iterative risk management strategy throughout the entire life cycle. The other requirements in Articles 10 to 15 can be understood as particular risk mitigation measures within the broader framework. The regulatory reasoning, therefore, shifts from isolated technical compliance to organised risk governance. Its ultimate goal is that any residual risks, after all reasonable protections have been implemented, be deemed "acceptable" in line with Article 9(5)[17].

11 Viewed through this analytical lens, it becomes evident that implementing the AI Act's requirements to achieve and demonstrate conformity invariably confronts providers with a series of complex decision problems. Indeed, determinations of risk acceptability are, by their very nature, archetypal decision problems[18]. Such problems necessitate choices among various alternative courses of action, each associated with distinct potential outcomes, benefits, harms, and inherent uncertainties. From this standpoint, an "acceptable risk" is not an objective, pre-existing quantity but rather the outcome associated with the most preferable alternative identified within a specific decision problem, after considering all relevant factors, including societal values, technical feasibility, economic implications, available evidence, and ethical considerations.

12 Consider, for instance, the development of an AI system used in autonomous vehicles, a high-risk application under Annex I. A regulatee (the AI provider) must decide on the trade-offs inherent in its design and performance. Providers must grapple with unavoidable accident scenarios (the "trolley problem" in a new guise)[19]. Should the system be programmed to prioritise the safety of its occupants above all else, or to minimise the total number of potential casualties, even if this entails a greater risk to those within the vehicle? This is a profound ethical decision problem with no single "correct" answer. A further example arises with AI systems intended for content moderation on large online platforms, which, even if they don't fall under the AI Act's high-risk categories, remain highly relevant due to their intersection with the Digital Services Act[20]. Is it more "acceptable" to implement highly aggressive filtering algorithms for potentially harmful content, thereby risking the erroneous removal of legitimate expression and impinging on freedom of speech (a false positive)? Or is it preferable to adopt a more lenient approach that, while safeguarding free speech more broadly, risks the wider proliferation of harmful material, hate speech, or disinformation (a false negative)? These illustrative scenarios underscore the challenging, value-laden trade-offs inherent in determining "acceptable risk" or "appropriate" accuracy, where

14 See, in particular, Article 3 of Decision No 768/2008/EC, according to which, for legislation concerning the marking of products, "Community harmonisation legislation shall restrict itself to setting out the essential requirements determining the level of such protection and shall express those requirements in terms of the results to be achieved," and further, that "where Community harmonisation legislation sets out essential requirements, it shall provide for recourse to be had to harmonised standards, adopted in accordance with Directive 98/34/EC, which shall express those requirements in technical terms."

15 This division of responsibilities, whereby essential requirements are articulated in legislation as high-level objectives, and technical standards, developed by the ESOs, provide detailed technical specifications, forms a foundational principle of European product safety legislation. See, for example, Jacques Pelkmans, 'The New Approach to Technical Harmonization and Standardization' (1987) 25 JCMS: Journal of Common Market Studies 249.

16 Mark McFadden and others, 'Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation' <https://oxil.uk/publications/2021-12-02-oxford-internet-institute-oxil-harmonising-ai/> accessed 11 March 2026; Alessio Tartaro, 'Regulating by Standards: Current Progress and Main Challenges in the Standardisation of Artificial Intelligence in Support of the AI Act' [2023] European Journal of Privacy Law & Technologies.

17 Jonas Schuett, 'Risk Management in the Artificial Intelligence Act' [2023] European Journal of Risk Regulation 1; Henry Fraser and José-Miguel Bello y Villarino, 'Acceptable Risks in Europe's Proposed AI Act: Reasonableness and Other Principles for Deciding How Much Risk Management Is Enough' [2023] European Journal of Risk Regulation 1.

18 Fischhoff and others (n 9).

19 Norbert Paulo, 'The Trolley Problem in the Ethics of Autonomous Vehicles' (2023) 73 The Philosophical Quarterly 1046.

20 Paddy Leerssen, 'An End to Shadow Banning? Transparency Rights in the Digital Services Act between Content Moderation and Curation' (2023) 48 Computer Law & Security Review 105790.

straightforward, universally agreed-upon, or purely technical solutions are often elusive. Regulatees are thus forced to make difficult judgments, balancing competing objectives and values on the basis of uncertain evidence. Even requirements that appear more technical, such as "sufficiently representative" training data (Article 10), resolve into decision problems for regulatees. What constitutes "sufficiently representative" training data to avoid harmful bias related, for example, to different performance for different demographic groups[21]? The regulatee must decide on the relevant demographic categories, the acceptable thresholds for representation within each, and the methods for measuring and mitigating identified biases[22]. These are not purely technical questions; they involve value judgments about fairness, equity, and the potential consequences of data-driven discrimination.

13 How, then, are such choices among competing alternatives in these decision problems to be made by regulatees in a manner that is defensible, transparent, and compliant with the AI Act's spirit? While the AI Act appears to assign a fundamental role to harmonised standards in this context, the current state of techno-scientific development in the field of AI, characterised fundamentally by contested values and uncertain evidence, makes the codification of unambiguous answers to these crucial questions within harmonised standards truly improbable. The next two sections will provide support for this argument.

III. Contested Values and Uncertain Evidence in the Field of AI

14 While Article 40 is intended to provide a clear compliance pathway via harmonised standards, we argue that the practical application of the essential requirements for high-risk AI systems under the AI Act still necessitates that regulatees navigate a series of complex decision problems which involve contested values and uncertain evidence. To show this, we will proceed by examining a plausible, albeit hypothetical, case study. This illustrative approach will allow us to demonstrate why neither the AI Act's legislative text nor, in all likelihood, the forthcoming harmonised standards can currently provide straightforward, unambiguous, or universally applicable solutions to these deep-seated challenges.

15 Consider the hypothetical case of "DoctorLLM," a sophisticated software system based on a large language model (LLM). This system is designed to process natural language questions and inputs from physicians and other qualified medical personnel. In response, it generates natural language outputs intended to provide recommendations concerning the diagnosis, treatment, or monitoring of human diseases and health conditions. Given its intended purpose, i.e., to provide information used to make decisions with diagnostic or therapeutic purposes, this system clearly falls under the purview of the MDR. According to Annex VIII, Section 6.3, Rule 11 of the MDR, software intended to provide information which is used to make decisions for diagnosis or therapeutic purposes is classified as at least Class IIa. It could be classified as Class IIb or even Class III depending on whether these decisions have an impact that may cause a serious deterioration of a person's state of health or a surgical intervention, or death or an irreversible deterioration of a person's state of health, respectively. For the sake of simplicity, let us assume, for this scenario, that DoctorLLM is classified just as Class IIa under the MDR, which requires challenging third-party conformity assessment[23]. Furthermore, this system would unequivocally be qualified as a high-risk AI system under the AI Act by virtue of Article 6(1a) as it is a product, i.e., a software as medical device, covered by the MDR, which is included in Annex I. Consequently, the provider of DoctorLLM, our regulatee, must ensure compliance not only with the stringent requirements of the MDR but also with all the applicable essential requirements stipulated by the AI Act and applicable to high-risk AI systems.

16 As mentioned, among these AI Act requirements, Article 9 mandates the implementation of a robust and continuous risk management system to ensure that "the overall residual risk of the high-risk AI system is judged to be acceptable." Another pivotal essential requirement, detailed in Article 15, stipulates that DoctorLLM must be designed and developed to achieve a level of accuracy, robustness, and cybersecurity that is "appropriate" to its intended purpose and context of use[24].

21 Laleh Seyyed-Kalantari and others, 'Underdiagnosis Bias of Artificial Intelligence Algorithms Applied to Chest Radiographs in Under-Served Patient Populations' (2021) 27 Nature Medicine 2176.

22 Line H Clemmensen and Rune D Kjærsgaard, 'Data Representativity for Machine Learning and AI Systems' (arXiv, 3 February 2023) <http://arxiv.org/abs/2203.04706> accessed 5 June 2024; FF Liza, 'Challenges of Enforcing Regulations in Artificial Intelligence Act - Analyzing Quantity Requirement in Data and Data Governance' (2022) <https://ceur-ws.org/Vol-3221/IAIL_paper9.pdf> accessed 21 February 2026.

23 Tuomas Granlund, Tommi Mikkonen and Vlad Stirbu, 'On Medical Device Software CE Compliance and Conformity Assessment', *2020 IEEE International Conference on Software Architecture Companion (ICSA-C)* (2020) <https://ieeexplore.ieee.org/abstract/document/9095660> accessed 30 May 2025.

24 Here, 'accuracy' can be broadly understood as functional correctness and reliability in generating clinically relevant

However, determining precisely what constitutes an "acceptable" level of overall residual risk or an "appropriate" level of accuracy for a system like DoctorLLM presents profound and multifaceted challenges for the regulatee. These are quintessential decision problems, characterised by contested values and uncertain evidence.

17 Establishing what constitutes an "acceptable" level of risk for a system such as DoctorLLM is particularly complex. It is not merely a technical calculation but a judgment that must balance the severity and probability of potential harms—such as misdiagnoses, inappropriate treatments, or the erosion of trust in the healthcare system—against the expected benefits. This evaluation is inherently subjective, influenced by ethical values, and made even more daunting by epistemic uncertainties about the system's actual ability to operate safely across all varied and unpredictable real-world clinical contexts, as well as the difficulty of anticipating every possible failure mode or misuse of the system. Furthermore, the definition of "acceptable" can vary significantly depending on the perspectives of the stakeholders involved (patients, physicians, healthcare institutions, and society at large), making consensus an elusive goal.

18 At first glance, problems concerning the determination of an "appropriate level of accuracy" for DoctorLLM might appear somewhat less structurally complex than those involving "overall risk acceptability." Nevertheless, they too rest fundamentally on underlying value judgments and assessments of uncertain evidence, which the regulatee must carefully consider. Determining the appropriate level of accuracy for DoctorLLM is not just about achieving the highest possible number in isolation; it involves defining a performance standard that is clinically safe and sufficient for its intended use case, balancing different types of potential errors (e.g., false positives vs. false negatives) based on the severity of the condition and the consequences of misdiagnosis.

19 Furthermore, determining the "acceptability of the evidence" that supports claims about a given level of accuracy is itself a critical and often overlooked part of this decision problem. This relates closely to what philosophers of science have termed "inductive risk", i.e., the risk of error inherent in making inductive inferences from limited evidence, and the value judgments involved in deciding how much evidence is "sufficient" to accept or reject a hypothesis[25]. Ascertaining the true, generalisable accuracy of a complex AI system like DoctorLLM is fraught with challenges for the regulatee. These challenges include limitations in the validation datasets used to test the system (e.g., representativeness, bias), questions about potential performance degradation or "drift" over time as medical knowledge and patient populations evolve, and the inherent difficulty in translating accuracy demonstrated in controlled lab settings to the complex, varied environment of real-world clinical practice. Additionally, there is epistemic uncertainty about the model's behaviour across the full range of potential inputs, particularly regarding novel or edge cases and the risk of generating inaccurate information, commonly known as "hallucinations."

IV. Addressing Decision Problems in NLF-regulated Fields

20 Upon closer inspection, the situation described above, where essential legislative requirements are articulated in broad, outcome-oriented terms, and the detailed technical means for achieving compliance are expected to be provided by harmonised standards, does not appear, on the surface, to be fundamentally different from that encountered in several other well-established fields regulated under the NLF. Precisely because NLF legislation is designed to delegate the specification of technical means for implementing the delineated essential requirements to standardisation bodies, these essential requirements are, by necessity, often couched in somewhat vague or general terms within the legislative text itself. Even in mature NLF sectors such as toys, machinery, or medical devices, regulatees (manufacturers and other economic operators) are continuously confronted with decision problems analogous to those described for AI regulatees.

21 For instance, how does a manufacturer of a high-risk medical device (a regulatee under the MDR) determine that, as an outcome of their comprehensive risk management process, "the residual risk associated with each hazard as well as the overall residual risk is judged acceptable," as mandated by Annex I, Chapter 1, Point 4 of the MDR? This problem of judging risk acceptability is structurally very similar to the challenge concerning the acceptability of risk for high-risk AI systems (Article 9(5) AI Act) faced by AI regulatees like the provider of DoctorLLM. Both involve making a judgment call about safety and risk in the face of uncertainty.

22 There is, however, a crucial set of differences between these established NLF fields and the nascent domain of AI regulation that significantly impacts how regulatees can approach and resolve these inherent decision problems. These differences relate fundamentally to the maturity of the field,

and sound recommendations.

25 Heather Douglas, 'Inductive Risk and Values in Science' (2000) 67 Philosophy of Science 559.

the availability of methodologies for evidence generation and evaluation, and the degree of societal and professional consensus on underlying values and acceptable trade-offs.

23 In contrast to the often-uncharted territory of AI, decisions regarding the acceptability of risk for a traditional medical device, for example, are typically conducted according to rigorous, systematic, and internationally recognised procedures. A cornerstone of this process is clinical investigation and evaluation. Clinical investigations for medical devices are conducted in line with widely accepted and highly detailed international standards, most notably ISO 14155:2020 on "Clinical investigation of medical devices for human subjects – Good clinical practice." This standard, and others like it, provide a comprehensive framework for the systematic and controlled collection of clinical data. During this process, the safety and performance of a medical device can be quantitatively measured through shared, validated, and meticulously documented testing procedures, patient follow-up, and statistical analysis. This ability to generate evidence systematically and rigorously, using methods that are themselves subject to consensus and standardisation, constitutes the first, crucial epistemic element that differentiates established NLF cases like medical devices from the current state of affairs in AI. Such established methodologies provide much clearer pathways for regulatees in those traditional sectors to gather the evidence needed to support their resolution of these decision problems in order to show compliance.

24 Yet, this systematic evidence collection and evaluation, while vital, is only the first differentiating element. As discussed, any collection, interpretation, and application of evidence is also subject to underlying value choices. Philosopher Heather Douglas has compellingly articulated the concept of "inductive risk," which refers to the unavoidable role of non-epistemic (e.g., ethical, social, economic) values in scientific judgment, particularly when deciding how much evidence is sufficient to accept or reject a scientific claim, and in weighing the potential societal consequences of being wrong (e.g., the cost of a false positive versus a false negative in a regulatory context)[26]. This involves value-laden decisions about what constitutes valid evidence, how it should be collected and interpreted, what level of uncertainty is tolerable, and at what point investigation can reasonably cease, balancing the pursuit of greater certainty against practical constraints and societal needs.

25 Crucially, in the field of AI, these fundamental value-laden decisions regarding evidence generation, interpretation, and the thresholds for acceptability are often as contested and unsettled as the empirical evidence itself, creating significant ambiguity and burden for regulatees. This is partly because, as acknowledged in the European Commission in the impact assessment accompanying the proposed AI Act, "robust and representative evidence for harms inflicted by the use of AI is scarce due to lack of data and mechanisms to monitor AI as a set of emerging technology"[27]. Under the stringent criteria for evidence-based policymaking and risk assessment typically applied in established domains such as pharmaceutical regulation or environmental risk assessment[28], such an acknowledged scarcity of robust, longitudinal evidence might have raised significant questions about the sufficiency of the evidentiary basis for a comprehensive legislative intervention of the AI Act's scale. This divergence arguably highlights a difference in underlying values guiding the regulatory approach itself: in the AI context, pressing concerns about the potential future threats posed by AI to fundamental rights, democratic values, rule of law, and societal well-being appear to have (perhaps justifiably) prevailed over strict adherence to the rigorous, data-intensive criteria for evidence collection, evaluation, and validation that might be demanded before similar evidence-based regulatory action in other, more data-rich areas[29]. While this proactive stance may be commendable from a precautionary perspective, it leaves AI regulatees to navigate a complex regulatory landscape with far fewer established evidential benchmarks and a less developed societal consensus on how to weigh competing values.

26 Unlike in the field of AI, in mature NLF domains such as medical devices, toys or machinery safety, decision problems regarding risk acceptability and other evidence-based and value-laden problems can often be resolved by regulatees in a shared, relatively uncontested, and predictable manner. This predictability in mature fields is grounded in a convergence of shared epistemic foundations, enabled by validated methodologies for systematic evidence generation, and a higher degree of societal and professional consensus on underlying values, including what constitutes acceptable risk levels

26 Douglas (n 26).

27 'Impact Assessment of the Regulation on Artificial Intelligence' (Shaping Europe's Digital Future, 21 April 2021) <https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence> accessed 3 August 2023.

28 Špela Majcen, 'Evidence Based Policy Making in the European Union: The Role of the Scientific Community' (2017) 24 Environmental Science and Pollution Research 7869.

29 Stephen Casper, David Krueger and Dylan Hadfield-Menell, 'Pitfalls of Evidence-Based AI Policy' (arXiv, 18 April 2025) <http://arxiv.org/abs/2502.09618> accessed 30 May 2025.

and the core principles guiding necessary trade-offs. Technical standards in these sectors effectively codify these shared understandings, translating agreed-upon methods for evidence generation and established criteria for interpreting that evidence into practical guidance that facilitates compliance for regulatees. In contrast, the AI domain currently lacks this robust foundation of shared methodologies and value consensus, which consequently renders the codification of these elements into harmonised standards exceedingly challenging at present.

V. Implications and Potential Solutions

27 This inherent difficulty in translating the AI Act's high-level, often deliberately vague, essential requirements into concrete, readily implementable, and consistently verifiable specifications for providers constitutes a significant hurdle for their compliance efforts and, by extension, for the overall success of the Act[30].

28 A principal implication of this potential misalignment between the NLF's assumptions and the realities of AI governance is the risk of undermining the regulatory effectiveness of the AI Act[31]. In this context, regulatory effectiveness can be understood as the capacity of a regulatory regime to achieve its stated objectives[32], i.e., in the case of the AI Act, ensuring a high level of protection for health, safety, and fundamental rights, fostering legal certainty, promoting innovation, and building public trust in AI. If the core "decision problems" that lie at the heart of the AI Act cannot be robustly and consistently addressed by regulatees through the anticipated guidance of harmonised standards, due to the aforementioned deep-seated challenges with divergent values and uncertain evidence, then the burden of interpretation and practical implementation shifts significantly, and often problematically, onto these individual economic actors.

29 Secondly, without clear, widely accepted benchmarks derived from such standards, conformity assessment procedures, either conducted via internal control for Annex III systems or involving notified bodies for Annex I systems, may lack consistency, comparability, and, in some cases, sufficient rigour. This, in turn, could significantly hamper the oversight capacity and effectiveness of notified bodies[33] (where involved) and national competent authorities. If each regulatee develops idiosyncratic interpretations of "acceptable risk" or "appropriate accuracy," or if different notified bodies apply varying criteria, the goal of a harmonised level of protection and a level playing field across the EU internal market could be jeopardised. Different Member States might also see their national authorities adopt divergent enforcement stances based on their own interpretations of these open-textured terms, leading to regulatory fragmentation rather than harmonisation. This is a particular concern for small and medium-sized enterprises, which constitute a significant portion of the AI development landscape in Europe, but may lack the extensive legal and technical resources of larger corporations to navigate such profound ambiguities and develop bespoke, highly sophisticated internal governance processes for each decision problem.

30 Consequently, the Act's ability to genuinely mitigate risks and protect fundamental rights in a consistent and predictable manner could be diluted, thereby impeding the achievement of its core societal goals. The actions, interpretations, and internal decision-making processes of regulatees are thus absolutely critical to the overall regulatory effectiveness of the AI Act. Where these regulatees face undue ambiguity, intractable value conflicts without clear guidance, or insurmountable evidential burdens, the effectiveness of the entire regulatory edifice suffers[34]. The promise of "presumption of conformity" via harmonised standards, a key pillar of the NLF[35], may prove illusory or difficult to attain in practice for many core AI Act requirements if the harmonised standards themselves cannot adequately capture and resolve these underlying decision problems.

30 Alessio Tartaro, 'Towards European Standards Supporting the AI Act: Alignment Challenges on the Path to Trustworthy AI.', *Proceedings of the AISB Convention 2023* (2023) <https://ssrn.com/abstract=4470766>.

31 Alessio Tartaro, 'Value-Laden Challenges for Technical Standards Supporting Regulation in the Field of AI' (2024) 26 Ethics and Information Technology 72.

32 Morag Goodwin and Roger Brownsword (eds), 'Regulatory Effectiveness I', *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://www.cambridge.org/core/books/law-and-the-technologies-of-the-twentyfirst-century/regulatory-effectiveness-i/F8062987DBCD0CE416B062C31FB7B992> accessed 30 May 2025.

33 See the contribution of Kamara and others in this special issue.

34 Morag Goodwin and Roger Brownsword (eds), 'Regulatory Effectiveness III: Resistance by Regulatees', *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://www.cambridge.org/core/books/law-and-the-technologies-of-the-twentyfirst-century/regulatory-effectiveness-iii-resistance-by-regulatees/418FCE986492BF2447F589AD61102678> accessed 30 May 2025.

35 Philippe Portalier, 'Myths and Realities of the Presumption of Conformity' <https://orgalim.eu/insights/myths-and-reality-presumption-conformity> accessed 6 April 2026.

31 To address the profound difficulties in operationalising the AI Act's essential requirements for high-risk systems, the solution proposed in the rest of this section aims to bolster the procedural and evidentiary foundation of conformity assessments, compelling regulatees to transparently articulate how they have navigated these complex decision problems, thereby shifting some emphasis from elusive substantive harmonisation via standards to robust procedural harmonisation and documented reasoning. Instead of relying solely on standards to deliver substantive solutions, a complementary approach that focuses on procedural rigour and transparency in regulatees' decision-making processes offers a more promising path forward. This approach aims to make the internal deliberation processes for these decision problems more manageable for regulatees and, consequently, more transparent and auditable for notified bodies and competent authorities.

32 Given the recent debate on the potential revision of the AI Act[36], the core proposal here involves amending the documentation requirements in Article 11 and Annex IV of the AI Act to explicitly require providers of high-risk AI systems to develop, document, and maintain what can be termed an "AI Act Compliance Case." This concept draws inspiration from established practices of "safety cases" in critical industries (e.g., aviation, nuclear power, railway signalling) and "assurance cases" increasingly discussed in the context of AI safety and ethics[37]. An "AI Act Compliance Case" would be a structured body of evidence and argumentation, forming a central part of the technical documentation, that explicitly demonstrates how the regulatee has identified, analysed, deliberated upon, and resolved the key decision problems pertinent to their specific high-risk AI system in order to meet the Act's essential requirements.

33 Specifically, this enhanced documentation would mandate that regulatees (primarily providers, but with implications for deployers who may need to contribute or verify parts of it) undertake and record the following:

34 Identification and framing of decision problems: regulatees would be required to explicitly identify the specific decision problems they encountered in seeking to comply with each relevant essential requirement. This includes articulating the alternatives considered and the core trade-offs involved for their particular AI system and its intended context of use.

35 Articulation of relevant values: For each significant decision problem, the regulatee must document the relevant values that were considered pertinent. This involves acknowledging potential conflicts between these values. Standards such as the IEEE 7000 offer a structured process to carry out these activities[38], and dedicated professionals can support the process[39].

36 Evidential basis and uncertainty assessment: The regulatee must comprehensively outline the evidential basis used to inform their decisions. This includes detailing the data, metrics, testing methodologies, and validation processes employed to generate evidence regarding the system's performance, reliability, and potential impacts. Crucially, this section must also include a transparent assessment of the limitations, uncertainties, and potential biases inherent in this evidence (e.g., gaps in training data, limitations of testing environments, and generalisability concerns).

37 Deliberation and justification of trade-offs: Given that many "decision problems" in AI involve balancing competing concerns, regulatees must provide a thorough explanation of how trade-offs between potentially conflicting values or objectives were deliberated upon and ultimately resolved. This would require them to document the decision-making process, the alternatives considered, the rationale for the chosen approach, and a robust justification for why this approach is deemed to lead to compliance with the AI Act (e.g., achieving an "acceptable" risk level or "appropriate" accuracy) in light of the system's intended purpose, context of use, and the acknowledged values and evidence.

38 While this proposal undoubtedly introduces a significant documentation requirement, it aims to make the underlying decision problems more manageable for regulatees in several ways,

36 'EU Commission Opens Door for "Targeted Changes" to AI Act' (POLITICO) <https://www.politico.eu/article/gpai-code-of-practice-to-come-in-weeks-ai-office-says/> accessed 30 May 2025.

37 Rasmus Adler and Michael Klaes, 'Assurance Cases as Foundation Stone for Auditing AI-Enabled and Autonomous Systems: Workshop Results and Political Recommendations for Action from the ExamAI Project' in Matthias Rauterberg and others (eds), *HCI International 2022 – Late Breaking Papers: HCI for Today's Community and Economy*, vol 13520 (Springer Nature Switzerland 2022) <https://link.springer.com/10.1007/978-3-031-18158-0_21> accessed 5 November 2023.

38 IEEE, 'IEEE Std 7000™-2021. IEEE Standard Model Process for Addressing Ethical Concerns during System Design' <https://standards.ieee.org/ieee/7000/6781/>; Sarah Spiekermann, 'From Value-Lists to Value-Based Engineering with IEEE 7000™', *2021 IEEE International Symposium on Technology and Society (ISTAS)* (IEEE 2021) <https://ieeexplore.ieee.org/document/9629134/> accessed 6 June 2022.

39 Mariangela Zoe Cocchiaro and others, 'Who Is an AI Ethicist? An Empirical Study of Expertise, Skills, and Profiles to Build a Competency Framework' [2025] AI and Ethics <https://doi.org/10.1007/s43681-024-00643-y> accessed 2 June 2025.

rather than simply adding to their burden. The proposed approach aims to enhance manageability for regulatees by providing a clear, structured framework for tackling complex, ill-defined problems, thereby replacing vague requirements with a defined process of analysis and justification. It shifts the focus from searching for elusive "correct" answers to developing robust, evidence-informed justifications for choices, which is often more achievable and fosters better internal understanding and risk management within the regulatees' organisation. This method also offers procedural certainty on how to demonstrate due diligence, even if substantive uncertainty about acceptable risk persists. In this revised framework, harmonised standards still play a crucial, though reoriented, role by supporting the process of reasoned justification rather than dictating specific outcomes. Standards could define the structure and methodology for preparing Compliance Cases, specify accepted methods for generating evidence, like bias testing, propose common metrics, or offer sector-specific guidance for typical decision problems. However, the ultimate judgment, documented in the Compliance Case, would remain with the regulatee, subject to assessment by conformity bodies and enforcement by competent authorities whenever appropriate. This, in turn, would bolster the AI Act's overall regulatory effectiveness by empowering regulatees with a clearer and more manageable framework for demonstrating compliance, facilitating more effective and consistent oversight, and fostering a culture of responsible AI development and deployment. It moves the AI Act towards a model where compliance is demonstrated not just by ticking boxes against standards, but by providing a compelling, evidence-backed narrative of due diligence and reasoned judgment in the face of inherent complexity.

39 However, it is crucial to acknowledge that this proposed shift from substantive, standards-based conformity to procedural, argument-based compliance presents its own challenges, particularly for the controlling authorities. The evaluation of a bespoke compliance case is inherently more resource-intensive and time-consuming than verifying conformity with a harmonised standard. Unlike a technical specification, a compliance case is a qualitative argument about navigating trade-offs and values. Its adequacy is not a matter of objective measurement but of reasoned judgment, making the assessment process itself inherently contestable. Each case would require deep, bespoke analysis by competent authorities, potentially leading to regulatory bottlenecks. The very process designed to enhance transparency for regulatees could therefore become a source of profound practical strain and disagreement for the authorities tasked with its assessment.

40 Nevertheless, this challenge does not invalidate the proposal; rather, it exposes the true and unavoidable cost of effectively regulating artificial intelligence. The strain on authorities is not a flaw in the compliance case model but a symptom of the inherent complexity of AI governance—a complexity that the standard-centric approach merely obscures rather than solves. Relying solely on standards risks creating a veneer of simplicity over a reality of inconsistent application and superficial compliance. The compliance case, by contrast, forces this difficult and resource-intensive work into a documented, auditable, and transparent format. The resulting political and financial challenge of equipping authorities to handle these assessments is therefore not a reason to retreat to a failing model, but an essential precondition for the AI Act to achieve its stated goals. In essence, it trades the false comfort of a simple but ineffective process for the demanding reality of a complex but ultimately more robust regulatory regime.

VI. Conclusion

41 This paper has explored the fundamental challenge posed by the AI Act's approach to regulating high-risk AI systems, specifically how its reliance on the NLF confronts significant friction points when applied to the inherent nature of several core essential requirements. We have argued that requirements such as ensuring "acceptable risk," "appropriate accuracy," and "commensurate human oversight" do not readily lend themselves to straightforward, universally applicable technical specifications capable of being fully codified within harmonised standards. Instead, these requirements compel us to grapple with intricate "decision problems" that necessitate balancing competing values and making judgments in the face of often profound epistemic uncertainty.

42 In contrast to sectors with established, widely accepted methodologies for systematic evidence generation and a higher degree of consensus on the values guiding the interpretation of that evidence and the determination of acceptable thresholds, the field of AI currently presents a landscape marked by divergent values, uncertain and evolving evidence, and a lack of established benchmarks for navigating complex trade-offs. While standards in mature domains can effectively codify shared epistemic and normative understandings into practical compliance guidance, the same mechanism faces substantial limitations in fully addressing the deep-seated value conflicts and pervasive uncertainties inherent in many AI applications.

43 The significant implication of this divergence is the potential diminishment of the AI Act's regulatory

effectiveness. In light of these challenges, this paper has finally proposed a complementary approach focused on enhancing the procedural rigour and transparency of regulatees' internal decision-making processes. The concept of an "AI Act Compliance Case," drawing inspiration from established safety and assurance case methodologies, serves as the cornerstone of this proposal. Within this proposal, harmonised standards would retain a crucial, albeit reoriented, role: supporting the process of reasoned justification by defining methodologies for evidence generation, structuring compliance documentation, and providing guidance on specific aspects, rather than attempting to provide definitive substantive answers to value-laden decision problems.

44 Ultimately, navigating the inherent complexities of AI governance requires moving beyond the sole expectation of static technical standards providing all the answers. By fostering a culture of transparent, accountable decision-making encapsulated in Compliance Cases, the AI Act can more effectively translate its high-level aspirations into concrete, verifiable, and trustworthy AI systems that genuinely benefit society while upholding fundamental rights and Union values.