Università degli Studi di Sassari
Dipartimento di Scienze umanistiche e sociali
Corso di Dottorato in Culture, Letterature, Diritti, Turismo e Territorio
XXXVIII Ciclo

Artificial Intelligence Standardisation for AI Regulation

A study on the role and limitations of harmonised standards in the context of the European Union legislation on AI

Relatore:
Prof. Massimo Dell'Utri
Prof. Enrico Panai

Candidato:
Alessio Tartaro

Anno Accademico 2025/2026

"Io sono il *coscienzioso dello spirito*, rispose l'interrogato, e nelle cose dello spirito difficilmente vi è qualcuno che le affronti con maggiore severità, scrupolo e durezza di me, se non colui dal quale l'ho imparato, Zarathustra stesso. Meglio non sapere nulla, che molte cose a metà! Meglio essere un folle per propria iniziativa, che un saggio secondo il parere di un altro! Io – vado fino al fondo: – che importa, se è grande o piccolo? Se si chiama cielo o palude? Un palmo di fondo mi basta: purché sia veramente un solido fondo! – un palmo di fondo: basta per starci in piedi. Nella autentica scienza coscienziosa non c'è nulla di grande, nulla di piccolo».
«Allora sei tu forse colui che conosce la sanguisuga? domandò Zarathustra; e insegui la sanguisuga fino agli ultimi fondi, tu coscienzioso?».
«O Zarathustra, rispose il calpestato, questa sarebbe un'enormità, come potrei assumermi un'impresa del genere!
Ma ciò in cui io sono un maestro e un conoscitore, è il *cervello* della sanguisuga: – questo è il mio mondo!"

(Friedrich Wilhelm Nietzsche*, Così parlò Zarathustra, trad. it. M. Montinari*, Milano, Adelphi, 1976, p. 378)

I want to thank me for believing in me, I want to thank me for doing all this hard work. I wanna thank me for having no days off. I wanna thank me for never quitting. I wanna thank me for always being a giver and trying to give more than I receive. I wanna thank me for trying to do more right than wrong. I wanna thank me for being me at all times.

(*Calvin Cordozar Broadus Jr.*)

Table of contents

Acknowledgement

There are many people who, in different roles and in different ways, have been part of the journey that concludes with this thesis. I would like to acknowledge them here, with gratitude for what they have contributed and for the role they have played along the way.

I am deeply grateful to Enrico Panai for giving me the opportunity to embark on this path and for guiding me throughout it as a mentor. I am also sincerely thankful to Professor Massimo Dell'Utri, my supervisor, for granting me the freedom that was so essential to the development of this research. My gratitude also goes to Vins Obnasca, Paola Maieli, Camila Bresolin, Gianpiero Nugheddu, Axelle Crognier, Piercosma Bisconti and Zoe Cocchiaro, my doctoral and work colleagues, as well as to the people at the University of Sassari who made all of this possible, among whom I would especially like to mention Professor Carla Bassu and Dr. Michele Cuccu.

I am also grateful to Professor Stephanie Bijlmakers, Professor Irene Karama, Professor Panos Delimatsis, Tatiana Wenno, and Zuno Verghese, who welcomed me to TILT and allowed me to spend my Erasmus+ traineeship in an environment of genuine academic excellence.

I am grateful to all my colleagues at Accenture, with whom I have begun a new chapter in which I hope to put into practice, and to good use, what I have learned over these years.

My thanks also go to the many colleagues, friends, and acquaintances with whom I had the chance to exchange even just a few words during this journey, often only once, yet always in ways that I found enriching. I will not mention them all by name, as they are too many, but if you are reading these lines and remember a moment of that kind, you may be sure that you would have been included in this list had I chosen to write it.

I am grateful to my family. I left them, and my native land, a long time ago, but a part of my heart has always remained there with them.

I am deeply grateful to Eva, my life partner and flatmate, who has always reminded me that pursuing a PhD is something pretty cool. She has one herself, and she is indeed very cool.

Finally, my deepest gratitude goes to my daughter, Ada, simply for having come into the world.

List of Abbreviations

AFNOR	Association Française de Normalisation (French Standardisation Association)
AI	Artificial Intelligence
ANSI	American National Standards Institute
BSI	British Standards Institution
BTPC	Big Third-Party Certifier
CA	Conformity Assessment
CAB	Conformity Assessment Body
CAP	Conformity Assessment Procedure
CBRN	Chemical, Biological, Radiological, and Nuclear
CE	Conformité Européenne (European Conformity)
CEN	European Committee for Standardisation
CENELEC	European Committee for Electrotechnical Standardisation
CJEU	Court of Justice of the European Union
CoP	Code of Practice
DIN	Deutsches Institut für Normung (German Institute for Standardisation)
DIS	Draft International Standard
EC	European Commission
EEA	European Economic Area
EFTA	European Free Trade Association
EN	European Norm (European Standard)
ESO	European Standardisation Organisation
ESS	European Standardisation System
ETSI	European Telecommunications Standards Institute
EU	European Union
FDIS	Final Draft International Standard
GATT	General Agreement on Tariffs and Trade
GDPR	General Data Protection Regulation
GPAI	General-Purpose AI
hEN	Harmonised Standard
IAF	International Accreditation Forum
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IS	International Standard
ISO	International Organisation for Standardisation
ISO/CS	ISO Central Secretariat
ITU	International Telecommunication Union
JTC	Joint Technical Committee
LLM	Large Language Model
MDR	Medical Device Regulation
MRA	Mutual Recognition Agreement

MSA	Market Surveillance Authority
NAFTA	North American Free Trade Agreement
NB	Notified Body
NIST	National Institute of Standards and Technology
NLF	New Legislative Framework
NSB	National Standardisation Body
NWIP	New Work Item Proposal
OSD	Online Standards Development
PAS	Publicly Available Specification
QMS	Quality Management System
SC	Subcommittee
SDO	Standards Developing Organisation
SDoC	Supplier Declaration of Conformity
SME	Small and Medium-sized Enterprise
SR	Standardisation Request
SSO	Standards-Setting Organisation
STS	Science and Technology Studies
TBT	Technical Barriers to Trade
TC	Technical Committee
TFEU	Treaty on the Functioning of the European Union
TR	Technical Report
TS	Technical Specification
UNI	Unione Normazione Italiana (Italian Standardisation Body)
UNIDO	United Nations Industrial Development Organisation
W3C	World Wide Web Consortium
WD	Working Draft
WG	Working Group
WSC	World Standards Cooperation
WTO	World Trade Organisation

Introduction

1. Problem statement

Artificial intelligence, an idea that once conveyed images of distant, almost fictional, futures, is no longer a far-off prediction. Instead, it is quietly and effectively weaving itself into the very essence of how we live, work, and connect. The ongoing transformation is steadily altering fundamental aspects of our world, compelling us to investigate how we can collectively navigate these profound changes and actively steer the trajectory of AI development in directions that align with our shared values and promote human flourishing.

The remarkable acceleration in AI capabilities witnessed in recent years is not the product of a single breakthrough but of the convergence of three mutually reinforcing factors[1]. First, foundational theoretical work on neural networks, dating back to the 1950s, provided the conceptual basis for machines capable of learning[2]. These early studies of artificial neurons, layered architectures, and learning algorithms were constrained by the computational resources of their time, yet they established the basic models through which machines could discover patterns and learn representations directly from data rather than relying exclusively on hand-coded rules. Second, the availability of data has expanded dramatically, driven by the digitisation of social, economic, and scientific life, the proliferation of connected devices, and the growth of platforms that generate vast quantities of user-produced content. This expanding data environment has supplied the raw material required to train increasingly complex models[3]. Third, advances in semiconductor design, GPU architectures, cloud infrastructure, and specialised accelerators such as TPUs have made the training and deployment of large-scale AI systems economically feasible. Falling computational costs have made it possible to experiment, iterate, and scale AI systems with unprecedented speed[4]. Taken together, these developments in theory, data availability, and computational power have propelled the contemporary expansion of artificial intelligence.

Because of the rapid progress in the field and the escalating power of its applications, the term "AI" has transcended its technical origins to become a pervasive cultural and economic buzzword, signifying both genuine technological prowess and, often, considerable hype. Venture capital floods into "AI-driven" startups[5], established corporations rebrand core offerings with AI features[6], and policymakers

[1] Cf. Michael Wooldridge, *A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going* (Flatiron Books 2021), for an accessible introduction to the history of AI.

[2] The seminal paper introducing this idea is Frank Rosenblatt, 'The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain' (1958) 65 Psychological Review 386.

[3] The idea that vast amounts of data can be surprisingly effective for complex tasks such as machine translation (often termed "The Unreasonable Effectiveness of Data," in a nod to Eugene Wigner's work on *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*) was notably discussed and popularized by Alon Halevy, Peter Norvig and Fernando Pereira, 'The Unreasonable Effectiveness of Data' (2009) 24 IEEE Intelligent Systems 8 <https://doi.org/10.1109/MIS.2009.36>.

[4] Cf Tim Hwang, 'Computational Power and the Social Impact of Artificial Intelligence' (Social Science Research Network, 23 March 2018) <https://doi.org/10.2139/ssrn.3147971> accessed 23 June 2025.

[5] For an analysis of economic trends in the field of AI as of 2025, see Nestor Maslej and others, 'Artificial Intelligence Index Report 2025' (arXiv, 8 April 2025) <https://doi.org/10.48550/arXiv.2504.07139> accessed 23 June 2025.

[6] A prime example is Microsoft: on January 15, 2025, their entire Microsoft 365 Office suite has been renamed Microsoft 365 Copilot, with "Copilot" serving to highlight the integration of AI within these apps.

globally scramble to position their nations as leaders in the purported "AI race."[7] This intense focus reflects the perceived transformative potential of artificial intelligence; indeed, this potential is frequently articulated in terms of its capacity to fundamentally reorder existing economic paradigms through unprecedented productivity enhancements across diverse industries[8] and the automation of complex tasks[9]. Moreover, the allure of AI is amplified by its prospective contributions to resolving intractable societal challenges, including but not limited to, augmenting diagnostic precision and accelerating therapeutic development within healthcare[10], and fostering more efficient resource management strategies to achieve sustainability goals[11]. The capacity of AI to discern intricate patterns within voluminous datasets further underpins expectations of its role in catalyzing new modalities of scientific inquiry and accelerating the pace of discovery and invention across a multitude of research domains[12].

However, alongside the enthusiasm surrounding AI's potential benefits, a parallel and equally crucial development has been the burgeoning awareness of its inherent risks and potential for societal harm. Concerns encompass a wide spectrum, including the propagation and amplification of biases leading to discriminatory outcomes, threats to individual privacy through pervasive surveillance and data exploitation, the generation and dissemination of convincing disinformation (deepfakes, synthetic media), profound implications for employment markets through automation, the potential development of autonomous weapons systems, and overarching challenges related to transparency, accountability, and the alignment of AI objectives with human values[13]. The recognition of these multifaceted risks has precipitated a powerful and widespread consensus: the trajectory of AI development and deployment must be actively governed to mitigate harms and ensure its alignment with societal well-being and fundamental rights.

Calls for the regulation of AI have consequently proliferated across academic, civil society, and policy spheres. Early initiatives often took the form of ethical guidelines and principles proposed by multi-stakeholder groups, international organisations, and individual corporations, seeking to establish normative frameworks for responsible AI development[14]. The European Union has emerged as a

---

[7] This trend is exemplified by the proliferation of national AI strategies, which are reviewed and discussed in works such as Samar Fatima, Kevin C Desouza and Gregory S Dawson, 'National Strategic Artificial Intelligence Plans: A Multi-Dimensional Analysis' (2020) 67 Economic Analysis and Policy 178; and, from a European perspective, in European Commission. Joint Research Centre. and Organisation for Economic Cooperation and Development., *AI Watch, National Strategies on Artificial Intelligence: A European Perspective.* (Publications Office 2021) <https://data.europa.eu/doi/10.2760/069178> accessed 23 June 2025.

[8] Cf. Mohamed Ali Trabelsi, 'The Impact of Artificial Intelligence on Economic Development' (2024) 3 Journal of Electronic Business & Digital Economics 142.

[9] Cf. Daron Acemoglu and Pascual Restrepo, 'Artificial Intelligence, Automation, and Work' *The economics of artificial intelligence: An agenda* (University of Chicago Press 2018).

[10] Cf. Fei Jiang and others, 'Artificial Intelligence in Healthcare: Past, Present and Future' (2017) 2 Stroke and Vascular Neurology <https://doi.org/10.1136/svn-2017-000101> accessed 23 June 2025.

[11] Cf. Rohit Nishant, Mike Kennedy and Jacqueline Corbett, 'Artificial Intelligence for Sustainability: Challenges, Opportunities, and a Research Agenda' (2020) 53 International Journal of Information Management 102104 <https://doi.org/10.1016/j.ijinfomgt.2020.102104>.

[12] Stefano Bianchini, Moritz Müller and Pierre Pelletier, 'Artificial Intelligence in Science: An Emerging General Method of Invention' (2022) 51 Research Policy 104604 <https://doi.org/10.1016/j.respol.2022.104604>.

[13] For a comprehensive high-level overview of these issues, which are a common focus in the field of AI ethics, refer to Arif Ali Khan and others, 'Ethics of AI: A Systematic Literature Review of Principles and Challenges' *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering* (Association for Computing Machinery 2022) <https://doi.org/10.1145/3530019.3531329> accessed 23 June 2025; Changwu Huang and others, 'An Overview of Artificial Intelligence Ethics' (2023) 4 IEEE Transactions on Artificial Intelligence 799 <https://doi.org/10.1109/TAI.2022.3194503>; Emre Kazim and Adriano Soares Koshiyama, 'A High-Level Overview of AI Ethics' (2021) 2 Patterns <https://doi.org/10.1016/j.patter.2021.100314> accessed 22 November 2024.

[14] A review of key guidelines is available in Anna Jobin, Marcello Ienca and Effy Vayena, 'The Global Landscape of AI Ethics Guidelines' (2019) 1 Nature Machine Intelligence 389; and Nicholas Kluge Corrêa and others, 'Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance' (2023) 4 Patterns 100857.

particularly fertile and proactive ground for translating these governance aspirations into concrete regulatory frameworks. As will be elaborated in Chapter 4 of this dissertation, the EU embarked on a deliberate and multi-stage pathway, beginning with strategic communications and white papers, followed by the development of sector-specific guidelines like the Ethics Guidelines for Trustworthy AI proposed by the High-Level Expert Group. This iterative process, characterised by a growing emphasis on binding rules, ultimately culminated in the landmark proposal and subsequent adoption of the Artificial Intelligence Act, commonly known as the AI Act[15]. This regulation represents a comprehensive, risk-based approach to governing AI applications within the EU market.

A crucial, and indeed defining, aspect of the AI Act, which will receive detailed analysis in this dissertation, is its profound reliance on technical standards for the implementation of the essential requirements for high-risk AI systems. This reliance is not incidental but is deeply rooted in the EU's regulatory approach established with the New Legislative Framework (NLF). As we will discuss in greater details in Chapter 3, the NLF paradigm emphasises defining essential requirements for products or services within legislation itself, while delegating the specification of detailed technical solutions to harmonised standards developed by the European Standardisation Organisations (ESOs), i.e., CEN, CENELEC, and ETSI[16]. As a result of the AI Act's adherence to the NLF's regulatory approach, technical standards have rapidly gained prominence in the discourse surrounding AI governance and regulation.

The AI Act's reliance on harmonised standards for the implementation of its essential requirements has therefore thrust technical standards into a position of central importance within the AI regulation ecosystem. They are poised to become important operational mechanisms through which high-level principles and legislative essential requirements are translated into concrete technical specifications, testing methodologies, and compliance benchmarks for AI systems. This pivotal role as a bridge between abstract principles and legal obligations on the one hand, and tangible technical implementation and operational governance processes on the other hand, consequently renders technical standards an interesting, and at present, arguably underexplored, subject for research within the broader field of AI regulation and governance.

However, this very centrality of technical standards within contemporary regulatory paradigms demands rigorous critical scrutiny. The fundamental issue animating this dissertation, therefore, is whether these tools, as currently conceived, developed, and implemented, possess the required characteristics to effectively fulfil the demanding regulatory role increasingly being ascribed to them. This inquiry is not pursued in a vacuum or as a matter of general abstraction; rather, it is specifically anchored within the normative substratum established by the European approach to Trustworthy AI and legally codified within the AI Act. The investigation undertaken herein consequently pivots on assessing the capacity of technical standards to provide the concrete technical means for the operationalisation of these regulatory requirements. Such an assessment must necessarily confront the complex, and often deeply embedded, value-laden issues that arise when translating the AI Act's high-level legal principles into precise technical specifications. It is precisely this translation process, and the potential for standards to act as de facto arbiters of normative ambiguity, that demands scrutiny. The core of this inquiry, therefore,

[15] For an overview of the legislative path that culminated in the AI Act's adoption, along with the intermediate stages of European approach to Trustworthy AI, see Nathalie A Smuha and Karen Yeung, 'The European Union's AI Act: Beyond Motherhood and Apple Pie?' in Nathalie A Smuha (ed), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (Cambridge University Press 2025) <https://doi.org/10.1017/9781009367783.015> accessed 23 June 2025.

[16] For a high-level overview of the role of standards in the AI Act, see 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/>.

lies in determining whether technical standards can effectively serve as instruments for operationalising the requirements of the AI Act, particularly when these requirements implicate value-laden considerations, and whether they can do so by offering not merely superficial compliance but robust and ethically sound frameworks. The central research question of this dissertation is therefore formulated as follows:

To what extent, and under what conditions, can technical standards developed in support of the AI Act genuinely operationalise the Act's high-level, value-laden requirements for Trustworthy AI, thereby serving as effective and legitimate instruments of techno-legal governance?

The answer to this question carries profound implications not only for the successful implementation of the AI Act but for the broader legitimacy of the European Union's model of co-regulation in the digital age. Ultimately, determining the limits and potential of these instruments is a prerequisite for ensuring that the promise of Trustworthy AI remains a substantive reality for European citizens, rather than devolving into a procedural exercise in performative compliance.

2. Objectives and significance of the research

The overarching objective of this dissertation is to provide a comprehensive and critical analysis of the current state of technical standardisation in the field of AI, elucidating its pivotal yet contested role in the implementation of the AI Act, while interrogating its inherent limitations. This inquiry is driven by the recognition that technical standards have ascended to a position of unprecedented prominence, yet the capacity of these instruments to effectively translate high-level legal requirements that inherently involve complex, value-laden matters into precise and actionable technical specifications remains fundamentally underexplored and is, by its very nature, potentially fraught with significant challenges.

In response to the research questions, this research primarily aspires to construct a detailed and holistic overview of the role and limitations of technical standards under the AI Act. Such an undertaking necessitates a preliminary analytical progression that interrogates the fundamental ontology of standards themselves. It is imperative to first establish precisely what constitutes a standard and to articulate the rationale for their pervasive significance in contemporary socio-technical systems. Before assessing their efficacy or legitimacy in the context of algorithmic governance, one must grasp the nature of standards as instruments that are simultaneously technical descriptions and vehicles of regulatory power, capable of shaping market dynamics and technological trajectories.

This foundational understanding serves as a precursor to dissecting the specific modality through which technical standards are integrated into the European Union's broader regulatory strategy for products and technologies. The EU's reliance on standardisation is not a novel phenomenon but rather a cornerstone of its internal market harmonisation, traditionally operationalised through mechanisms that couple essential legislative requirements with voluntary technical specifications. However, this established approach has witnessed a significant expansion and transformation. More recently, this regulatory paradigm has been transposed into the digital sphere, marking a strategic evolution in how the Union governs intangible and complex technologies. This trajectory is most visibly embodied in recent legislative initiatives such as the AI Act and the Cyber Resilience Act, which systematically delegate the

operationalisation of high-level legal mandates—ranging from cybersecurity protocols to fundamental rights protections—to the domain of technical standardisation.

In addition to these analytical objectives, this thesis advances specific arguments in direct response to the core problem statement articulated in the preceding section concerning the fitness of technical standards for AI regulation. The research aims to substantiate and elaborate upon two principal limitations intrinsic to the current paradigm of technical standardisation for AI. The first limitation identified is fundamentally ethical in nature. Technical standards, by their constitution and operational logic, are inherently ill-equipped to resolve substantive questions of value or make complex normative judgments. Their methodologies favour measurability, interoperability, and procedural efficiency, often struggling to engage meaningfully with contested ethical. This incapacity severely undermines their potential contribution to implementing requirements within the AI Act that necessitate nuanced ethical reasoning and value-based determinations, thereby creating a significant gap between regulatory intent and standardised implementation. The second limitation is epistemic. Key areas mandated by the AI Act, including data quality for specific contexts, meaningful human oversight mechanisms for complex autonomous systems, and verifiable robustness against unforeseen adversarial conditions, currently lack established best practices, mature methodologies, and broad scientific consensus. Given that SDOs function as epistemic communities, relying on the collective knowledge and consensus of participating experts, the absence of such mature foundations within these communities directly translates into an inability to codify effective, reliable standards for these critical aspects. The standardisation process, therefore, reflects and amplifies the current state of uncertainty and fragmentation within the underlying scientific and technical field of AI.

Finally, cognisant of these critical limitations, this dissertation does not merely diagnose problems but also aims to contribute constructively to the evolution of AI regulation. Alongside the analytical and critical perspectives, a core objective is to advance concrete proposals for enhancing the effectiveness of technical standards as instruments of socially beneficial AI regulation. These proposals will target the identified shortcomings, suggesting potential pathways for refining standardisation processes, fostering greater interdisciplinary collaboration within SDOs, developing novel methodologies for addressing value-laden requirements, and strengthening the linkages between standardisation activities and broader societal oversight mechanisms. The goal is to outline a framework wherein technical standards can more robustly support regulatory objectives while acknowledging and mitigating their inherent limitations.

This research possesses significant relevance for a diverse range of stakeholders and scholarly domains. It offers crucial insights for policymakers and regulators at the EU level and beyond, grappling with the practical implementation challenges of the AI Act and similar frameworks. For SDOs and their participants, the analysis provides a critical reflection on their evolving role and the practical and conceptual challenges they face in the AI domain. Industry practitioners developing and deploying AI systems will benefit from a clearer understanding of regulatory implications of technical standards. Legal scholars specialising in technology regulation and the evolving interplay between law and technical standards will find the exploration of the NLF's application to AI regulation particularly pertinent. Ethicists and philosophers of technology will engage with the analysis of value conflicts within standardisation. Ultimately, this dissertation aims to contribute meaningfully to the nascent but critical field of AI regulation by providing a rigorous and critically engaged assessment of the promises and perils of relying on technical standards as a cornerstone of AI regulation.

3. Methods, assumptions and materials

3.1. Methods

Given the inherently complex and multifaceted nature of standardisation as a socio-technical phenomenon, particularly within the emergent domain of AI, this research adopts an explicitly interdisciplinary methodological approach. The research is thereby characterised as interdisciplinary, conforming to the framework articulated by Newell[17]. According to Newell, the quintessential feature of an interdisciplinary approach resides in a dual-phase process: initially drawing insights from discrete disciplinary perspectives and subsequently integrating these insights to achieve a more holistic understanding of a given subject.

The initial phase of drawing upon disciplinary perspectives commences with defining the problem broadly. This necessitates the careful determination of relevant disciplines and schools of thought that contribute substantially to explicating the objects being studied. A critical subsequent step involves developing a working command of the salient concepts, theories, and methods relevant to each chosen discipline, to a degree sufficient to illuminate specific features of the subject under investigation. The culmination of this initial phase is the generation of distinct insights into the problem at hand from each of the selected disciplinary perspectives.

The second, and arguably more distinguishing, phase of Newell's interdisciplinary methodology is the synthesis or integration of these disciplinary insights to forge a comprehensive understanding. This integrative process is often elusive yet critical for genuine interdisciplinarity. It begins with the systematic identification of conflicts that may arise among disciplinary insights, frequently achieved by scrutinising the underlying assumptions inherent in each perspective or by comparing the terminological apparatus employed across different disciplines. It is recognised that disciplinary assumptions often reflect the particular facet of reality upon which a discipline concentrates, and specific terminology can carry divergent meanings in disparate contexts. The subsequent evaluative stage involves resolving these identified conflicts, working assiduously towards a common vocabulary and a shared set of assumptions. This is seldom a linear progression but rather a non-linear process, characterised by an oscillation between the discrete disciplinary insights and the emerging, more integrated overall pattern of the phenomenon under study. Techniques instrumental in this endeavor may include the redefinition of terms, the extension of conceptual meanings, the creation of continuums between seemingly disparate ideas, or the transformation of opposing axioms to achieve greater coherence. The overarching objective throughout this process is to minimise alterations to established disciplinary constructs while simultaneously achieving a coherent and robust understanding. This meticulous integration facilitates the construction of a new, more comprehensive understanding of the problem. The ultimate objective, therefore, is to achieve a more comprehensive understanding of a multifaceted subject, an understanding that is qualitatively different from the mere summation of individual disciplinary parts.

According to Newell, for an interdisciplinary approach to be legitimately employed, the primary and indispensable prerequisite is that the object of study must itself constitute a complex phenomenon. Accordingly, the object of study must be multifaceted, presenting various dimensions or aspects that invite inquiry from different disciplinary standpoints. If the phenomenon were monolithic or

[17] William H Newell, 'A Theory of Interdisciplinary Studies' (2021) 19 Issues in Integrative Studies 1. For the purposes of this discussion, Newell's methodology will be simplified to enhance accessibility for readers unfamiliar with the topic. Specifically, references to the concept of complex systems, which is central to Newell's method, will be omitted. Nevertheless, this explanation aims to remain rigorous and provide a clear rationale for why standardization, and particularly the standardization of AI, is well-suited for interdisciplinary inquiry.

unidimensional in its relevant characteristics, a focused investigation within a single discipline would again be the more appropriate and efficient course of action. However, mere multifacetedness is not, in itself, a sufficient condition for true interdisciplinarity. The various facets of the object of study must also cohere; they must be interconnected and interdependent in meaningful ways, forming an integrated whole. In the absence of such coherence, where facets are present but essentially disaggregated or loosely coupled, a multidisciplinary approach, wherein different disciplines offer parallel but unintegrated perspectives, might suffice, as the imperative for genuine synthesis and integration would be diminished. The crucial distinguishing factor necessitating an interdisciplinary, rather than merely multidisciplinary, engagement is the inherent need to understand how these diverse facets interrelate and mutually constitute the phenomenon. In essence, the compelling argument for interdisciplinary studies, as articulated within this theoretical framework, is that they are necessitated by the fundamental structure of the phenomenon under investigation. Such phenomena, by their very nature, unify seemingly divergent analytical approaches, bridging traditional divides between disciplines. This theory thus serves to validate and provide a robust theoretical rationale for determining when and where an interdisciplinary approach is not merely desirable but methodologically indispensable for achieving a profound and actionable understanding of the phenomenon in question.

The inherent complexity of AI standardisation, with its interwoven technical, ethical, legal, and societal strands, strongly indicates its suitability as a phenomenon for rigorous interdisciplinary analysis. As Hesse has cogently argued, standardisation, in general, demands scholarly attention from a confluence of perspectives, including the legal, economic, regulatory, engineering, organisational, sociological, and philosophical[18]. This imperative is demonstrably amplified within the domain of artificial intelligence, where the pace of technological innovation, the profundity of potential societal impacts, and the nascent character of governance frameworks converge to create an analytical challenge of exceptional intricacy. Consequently, attempting to categorise this investigation within a single disciplinary silo would be not only reductive but fundamentally misaligned with the intricate reality of its object of analysis, thereby precluding a sufficiently nuanced appreciation of the subject matter.

As previously adumbrated, the application of an interdisciplinary method, particularly following Newell's framework, commences with the capacity to draw fruitfully upon insights generated from diverse disciplines. This foundational step, however, also delineates a practical limitation regarding the exhaustive application of such a method within the scope of the present research. Indeed, while a comprehensively interdisciplinary investigation into AI standardisation would ideally encompass and integrate all the dimensions identified by Hesse, namely the legal, economic, regulatory, engineering, organisational, sociological, and philosophical, the current study has found it possible to fruitfully integrate insights primarily from the legal, regulatory, and philosophical perspectives. Therefore, while acknowledging the significant insights potentially offered by detailed economic analyses, specific engineering evaluations of technical standards, or organisational studies of standards development bodies and standard implementation with businesses, these particular aspects, which would undoubtedly contribute to an even more holistic, profound, and indeed comprehensively interdisciplinary understanding of AI standardisation, are not explicitly incorporated into the analytical framework of this dissertation. The decision to delimit the scope in this manner stems from the objective to achieve a deep and coherent synthesis within the chosen interdisciplinary nexus, rather than a more superficial survey across an overly broad range of disciplines.

[18] Wilfried Hesser, 'The Need for Interdisciplinary Research on Standardization' [1997] Presentation for the SCANCOR/SCORE Seminar on Standardization September 18-20 1997 Lund, Sweden.

Because of the interdisciplinary character of the present research, it is crucial to render explicit the provenance of the diverse insights that are progressively integrated into the overarching analytical framework. These insights are principally derived from two distinct streams of research which, as this dissertation aims to demonstrate, can be fruitfully interwoven to achieve a more nuanced comprehension of the phenomenon of AI standardisation, the crucial role it plays in AI regulation, and, critically, its inherent limitations.

The first of these intellectual currents consists of the extensive body of existing scholarship concerning the role of standards in EU law. This established corpus has meticulously investigated critical questions pertaining to the legal status of standards supporting EU legislation, alongside persistent challenges concerning their legitimacy, accessibility, representativeness, and related governance issues[19]. While this thesis does not seek to make novel contributions to these well-trodden areas of legal doctrine per se, it performs the vital function of identifying and analysing the transposition of these very same structural and normative problems into the specific context of AI standardisation[20]. The contention here is that the fundamental legal and legitimacy quandaries identified within broader European standardisation processes are replicated, largely unchanged, within the AI domain, thereby forming a persistent background against which newer regulatory and ethical challenges must be understood. This transposition underscores the path-dependent nature of standardisation governance, even in technologically novel fields.

The second primary domain from which insights are systematically drawn is the philosophical field of metaethics[21]. Metaethics, as a branch of moral philosophy, scrutinizes the meaning of ethical terms (such as "good", "right", and "ought to"), the ontological status of moral properties (inquiring, for example, whether they are objective features of reality or subjective constructs), and the logic of moral reasoning. Of particular salience to the present inquiry, metaethics provides sophisticated conceptual tools for analysing terms and concepts that possess both descriptive content, which specifies certain factual features of a situation or entity, and an ineluctable normative or evaluative component, which expresses approval, disapproval, or an obligation. Such terms are commonly described in metaethical discourse as "thick concepts"[22]. Insights derived from this metaethical tradition are employed herein to investigate complex concepts embedded within legal instruments such as the AI Act, including notions of "*acceptable risk*," "*appropriate level of accuracy*," and "*commensurate human oversight measures*." As the thesis will show, these can be considered thick concepts according to the meaning of this term within metaethical discourse, as they inherently conjoin descriptive criteria with evaluative judgments. A metaethical analysis of these concepts, therefore, illuminates the inherent limitations of purely technical standards in operationalising such high-level requirements. As will be argued in the last part of the dissertation, this difficulty arises precisely because these requirements contain an ineluctable evaluative dimension that transcends purely descriptive or technical specification.

---

[19] Mariolina Eliantonio and Caroline Cauffman, *The Legitimacy of Standardisation as a Regulatory Technique: A Cross-Disciplinary and Multi-Level Analysis* (Edward Elgar Publishing 2020) <https://www.elgaronline.com/display/edcoll/9781789902945/9781789902945.xml>.

[20] 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 <https://doi.org/10.57230/EJPLT222AT> accessed 27 November 2023.

[21] Geoff Sayre-McCord, 'Metaethics' in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Spring 2023, Metaphysics Research Lab, Stanford University 2023) <https://plato.stanford.edu/archives/spr2023/entries/metaethics/> accessed 21 August 2025.

[22] Pekka Väyrynen, 'Thick Ethical Concepts' in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Spring 2025, Metaphysics Research Lab, Stanford University 2025) <https://plato.stanford.edu/archives/spr2025/entrieshick-ethical-concepts/> accessed 8 September 2025.

The rigorous integration of these two distinct yet complementary research domains, i.e., the legal scholarship on EU standardisation and the philosophical inquiries of metaethics, yields a more comprehensive and robust framework for understanding both the potential contributions and the significant constraints of technical standardisation in the field of AI regulation. Crucially, these limitations are shown to arise not merely from the legitimacy deficits and governance challenges identified by legal scholars, such as issues of due process, transparency, and stakeholder representation in standards-development organisations. Rather, as a metaethical perspective allows us to appreciate more fully, profound limitations also stem from the very nature of the deeply value-laden questions that instruments like the AI Act raise, questions concerning acceptability of risk levels, appropriateness of accuracy levels etc, which technical standards, by their intrinsic design and purpose, are structurally ill-equipped to comprehensively or definitively operationalise. This inherent mismatch between the technocratic nature of standardisation and the normative complexity of AI regulation highlights a critical tension which will be explored in the research.

3.2. Assumptions

Given the inherently multidisciplinary nature of the research, and the consequent potential for conceptual ambiguity arising from the confluence of distinct disciplinary lexicons, it is imperative to explicitly articulate the core assumptions and conceptual underpinnings guiding the research at this juncture. Several key premises warrant elaboration to delineate the scope and theoretical orientation of this work.

Firstly, the concept of "standards" itself occupies the central position within this investigation and demands preliminary clarification, notwithstanding a more granular analysis reserved for a dedicated analysis in Chapter 2. For the purposes of this dissertation, the term "standard" is employed with a specific, bounded meaning: it refers exclusively to technical specifications and requirements developed and promulgated by formally recognised SDOs. This delimitation deliberately excludes *de facto* standards arising solely from market dominance or informal industry consensus. The focus on SDO-produced standards stems directly from their increasingly formalized role within regulatory frameworks, such as the forthcoming AI Act, where they are envisaged as instrumental tools for operationalising compliance with essential regulatory mandates. Consequently, the analysis herein is circumscribed to standards emanating from these institutionalized bodies possessing established procedures for development, consensus-building, and publication.

Secondly, the research is fundamentally concerned with assessing the regulatory effectiveness of these standards. This necessitates a concurrent explication of the concept of "regulation." Adopting a broad yet analytically useful perspective, this dissertation aligns with the conceptualization advanced by Julia Black, which posits regulation as the "sustained and focused attempt to alter the behaviour of others according to defined standards or purposes with the intention of producing a broadly identified outcome or outcomes, which may involve mechanisms of standard-setting, information-gathering and behaviour-modification."[23] This encompassing view acknowledges the diverse mechanisms and actors involved in contemporary regulatory governance, moving beyond purely state-centric command-and-control models. It is particularly apt for analysing the complex interplay between traditional legislation and delegated technical standardisation, recognising SDOs as actors contributing to the regulatory process within a broader governance ecosystem.

[23] Julia Black, 'Critical Reflection on Regulation' <https://eprints.lse.ac.uk/35985/>.

Thirdly, the specific understanding of effectiveness employed herein pertains to the capacity of standards to function as reliable instruments supporting the attainment of explicitly articulated regulatory goals[24]. Within the immediate context of this research, the pertinent goals are those enshrined within the proposed AI Act, and most particularly, the objectives meticulously formulated within its essential requirements. These essential requirements constitute high-level, legally binding obligations that AI systems, particularly those classified as high-risk, must satisfy to be lawfully placed on the EU market or put into service. They address fundamental concerns such as data quality, transparency, human oversight, accuracy, robustness, and cybersecurity, yet are often framed in general terms, necessitating further specification for practical implementation. Effectiveness, therefore, is evaluated against the metric of how proficiently a given standard facilitates concrete progress towards fulfilling these legislatively mandated objectives. This facilitation is primarily achieved by providing the requisite technical means for the operationalisation of the said essential requirements. In essence, standards are assessed based on their ability to translate the abstract, outcome-oriented legal stipulations of the AI Act's essential requirements into verifiable technical specifications, methodologies, processes, or descriptive characteristics that AI providers can implement.

Finally, the analytical trajectory of this dissertation, particularly in its culminating chapters, leverages specific conceptual tools drawn from philosophy that merit preliminary introduction due to their critical role in diagnosing challenges to effectiveness. One pivotal tool derives from philosophy of language, specifically the analysis of thick ethical concepts. Terms central to regulatory discourse, such as "acceptable risk," or "appropriate level of accuracy" in the context of the AI Act, exemplify such concepts. Thick concepts inherently intertwine descriptive and evaluative dimensions; they simultaneously describe a state of affairs and convey a positive or negative valuation. Defining "acceptable risk," for instance, involves not merely identifying a probabilistic threshold of harm but inherently entails a complex normative judgment about what level of potential negative consequence a society is willing to tolerate under specific circumstances, balancing potential benefits against potential harms. Recognising these terms as thick concepts is crucial because it highlights the irreducible normative content embedded within ostensibly technical regulatory criteria. This inherent normativity poses significant challenges for standardisation, which often strives for ostensibly neutral, measurable specifications.

These foundational assumptions related to the specific definition of standards, the adopted conception of regulation and effectiveness, and the critical application of metaethical concepts, collectively establish the conceptual bedrock upon which the subsequent interdisciplinarity analysis rests. They provide the necessary precision to navigate the complex terrain where technology, law, ethics, and governance intersect, while simultaneously equipping the analysis with the theoretical tools required to probe the deeper, often value-laden, challenges inherent in using technical standards for AI regulation.

3.3. Materials

The inherent interdisciplinarity character of this research is substantively reflected in the diversity of sources employed to support its analysis. A comprehensive literature review spanning the legal, regulatory, and philosophical domains provided the essential theoretical scaffolding. This foundational work facilitated a rigorous understanding of the current state of scholarship concerning the role of

[24] 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://doi.org/10.1017/CBO9781139047609.016> accessed 30 May 2025.

standards in EU law and the ethical/epistemological challenges pertinent to standardisation, thereby situating the present inquiry within established academic discourse and identifying key conceptual frameworks. To this theoretical grounding, a multifaceted array of primary and secondary materials was integrated, encompassing formal normative sources (including relevant EU legislation and preparatory documents) and grey literature (such as technical reports, white papers, and policy briefs from standardisation bodies, industry consortia, and regulatory agencies). This combination of sources ensures the analysis is empirically anchored while remaining theoretically informed[25].

4. Structure of the dissertation

The dissertation is structured in five chapters, in addition to the introduction and the conclusion. Chapter 1 reconstructs the principal motivations for regulating AI, the general and AI-specific challenges that regulation encounters, and the principal tools available to policymakers, with particular attention to co-regulatory instruments. Chapter 2 develops a conceptual and historical account of standards and standardisation, examining their forms, functions, production processes, and enforcement through conformity assessment. Chapter 3 then situates these issues within the EU legal order by analysing the evolution from the Old Approach to the New Approach and the New Legislative Framework, as well as the institutional architecture of the European standardisation system. Chapter 4 applies that framework to AI by tracing the emergence of AI as a regulatory object in the EU, examining the core features of the AI Act, and clarifying the role assigned to harmonised standards and conformity assessment. Finally, Chapter 5 presents the dissertation's main critical argument: it evaluates the regulatory effectiveness of technical standards in the field of AI, shows why key AI Act requirements remain vague and contestable, and proposes a more explicitly procedural approach to compliance

[25] The author was involved for two years in standardisation activities relating to artificial intelligence, including in the capacity of editor of a project on competence requirements for AI ethics professionals. While this experience significantly contributed to the author's broader understanding of the subject matter and indirectly supported the development of this dissertation, no information, documentation, or other content originating from that work has been used in the present research.

Chapter 1.
Background: motivations, challenges and tools for the regulation of AI

1. Aim of the Chapter

This chapter provides a background discussion about the contemporary discourse surrounding the regulation of AI, examining the primary motivations, the array of available tools for regulation and the main challenges related to it. It will be argued that, while a spectrum of peripheral justifications exists, the principle of risk minimisation constitutes the most omnipresent, yet profoundly ambiguous, rationale for AI regulation. Having established this powerful, albeit contested, impetus, the analysis will pivot to the challenges inherent in this regulatory endeavour. The discussion will differentiate between general obstacles, as identified in the broader literature on the regulation of technology, and those specific to AI, chief among them the distinctive nature of the risks it engenders. Finally, the chapter will survey the regulatory toolkit available to policymakers and regulators, wherein technical standards emerge as an instrument of particular centrality. This exploration will demonstrate their growing prominence as a privileged mechanism for translating abstract principles into concrete operational practice, thereby setting the stage for the subsequent analysis in this dissertation.

2. Motivations for the Regulation of AI

The question of whether, and to what extent, AI should be regulated elicits a wide spectrum of responses within contemporary discourse. Certain commentators advocate for a largely unregulated environment, suggesting that intervention is premature and risks proving counterproductive by hindering innovation[26]. A more circumspect position posits that, before embarking upon substantive regulatory efforts, a more profound and evidence-based comprehension of the technology's multifaceted risks and opportunities must be achieved[27]. As Redd put it in 2018, "*A 'wait and see' approach is likely to produce better long-term results than hurried regulation based on, at best, a very partial understanding of what needs to be regulated*"[28]. These cautious stances are predicated on several interlocking arguments that question both the feasibility and desirability of imposing a comprehensive regulatory regime upon AI.

A foundational challenge resides in the very act of defining the object of regulation[29]. The ambiguity surrounding what constitutes "artificial intelligence" complicates any legislative effort from its inception. As it is fundamentally a category of algorithmic processes with roots extending back to the 1950s, there

[26] Daniel Castro and Michael McLaughlin, 'Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence | ITIF' [2019] Information Technology and Innovation Foundation (ITIF) <https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence/> accessed 3 August 2023.

[27] Rishi Bommasani and others, 'A Path for Science- and Evidence-Based AI Policy' [2024] https://understanding-ai-safety.org/.

[28] Chris Reed, 'How Should We Regulate Artificial Intelligence?' (2018) 376 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20170360 <https://doi.org/10.1098/rsta.2017.0360>.

[29] Hannah Ruschemeier, 'AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal' (2023) 23 ERA Forum 361 <https://doi.org/10.1007/s12027-022-00725-6>.

is no universally accepted demarcation between AI and other advanced forms of software. This creates a significant risk that legislators, in an attempt to craft a precise legal definition, may codify a conception of AI that diverges sharply from its technical and scientific understanding, potentially leading to regulations that are either over-inclusive or misdirected. The debate thus becomes mired in preliminary questions of whether to regulate software in general, specific algorithmic applications, or indeed if the distinction is meaningful enough to warrant a new regulatory paradigm at all, particularly when "AI" is often deployed more as a marketing term than a descriptor of a discrete technology.

This definitional ambiguity is further compounded by the contention that novel, AI-specific legislation may be superfluous. This critique is not without precedent, recapitulating as it does a foundational debate within early cyberlaw scholarship, famously encapsulated in the "law of the horse" analogy[30]. The central tenet of that argument can be now repurposed for the domain of artificial intelligence to argue that a subject-specific body of law is unnecessary when general legal principles suffice. Applied to the current discourse, this position would maintain that existing legal frameworks already provide a robust and effective regime that could be adapted for AI[31]. To illustrate, proscriptions against unlawful discrimination are not rendered moot by the technological medium; they apply with equal force whether the discriminatory outcome is the product of human prejudice or algorithmic bias. Likewise, one may argue that long-standing principles of tort law requiring remedy for harms caused to others are not fundamentally altered by the involvement of an AI system[32]. With some minor adjustments, the established legal frameworks, which has proven remarkably adaptable over centuries, may possess sufficient elasticity to absorb the challenges posed by AI and other emergent technologies without recourse to new, potentially innovation-stifling statutes. Again, the central fear behind this argument would be that premature or ill-conceived regulation could effectively stall progress not only in AI but across the entire technological ecosystem.

Complementing these theoretical objections, the case against precipitous AI regulation is further reinforced by a series of more practical arguments, presented in Table 1[33]:

Objection	**Rational**
Slower and more expensive AI development	Regulations make the development of AI slower and more expensive because they require firms to establish processes to comply with the requirements. Delays due to conformity can lead to a loss of competitiveness, while costs can be a barrier to entry for SMEs and start-ups.

[30] Frank Easterbrook, 'Cyberspace and the Law of the Horse' (1996) 1996 University of Chicago Legal Forum <https://chicagounbound.uchicago.edu/uclf/vol1996/iss1/7>.

[31] Florian Martin-Bariteau, 'Regulating Autonomous Systems in Canada' (Social Science Research Network, 30 January 2023) <https://doi.org/10.2139/ssrn.5104939> accessed 27 June 2025.

[32] Kholoud Tu'ma Hassan and Raed Siwan Atwan, 'The Civil Liability of Artificial Intelligence Applications: Between the Limitations of Traditional Liability and the Evolution of the Product Concept' (2025) 15 Indian Journal of Information Sources and Services 276 <https://doi.org/10.51983/ijiss-2025.IJISS.15.3.31>.

[33] Adapted from Alessio Tartaro, Adam Leon Smith and Patricia Shaw, 'Assessing the Impact of Regulations and Standards on Innovation in the Field of AI' (arXiv, 8 February 2023) <https://doi.org/10.48550/arXiv.2302.04110> accessed 4 May 2023.

Less innovation	Bans on certain applications of AI will prevent reaping its benefits. For example, a possible ban on autonomous vehicles would lead to the loss of the benefits that would result if they were adopted.
Lower-quality AI	Requirements on explainability may lead to the development of lower quality AI systems due to the trade-off between accuracy and explainability.
Less AI adoption	Requirements on human oversight may reduce the adoption of AI. Requiring a human to review the decisions of an AI system leads to a loss of the benefits (speed, accuracy) that result from adoption.
Less economic growth	Regulations based on the precautionary principle discourage AI adoption and hinder economic growth because they slow down the increase in productivity.
Fewer options for consumers	When certain AI systems are banned, consumers lose the ability to access products or services based on those applications.
Higher prices	Regulations create compliance costs and these are passed on to the consumer. Furthermore, the ban on certain applications prevents the costs of certain processes from being reduced.
Inferior consumer experiences	The ban on certain AI applications such as facial recognition prevents the pro- vision of better services for consumers, e.g., supermarkets without cashiers.
Fewer positive social impacts	The ban on some applications prevents its positive social effects from being realized. For example, autonomous vehicles can use platooning to reduce emissions and bring environmental bene- fits.
Reduced economic competitiveness and national security	Finally, the regulation of AI damages the countries where it is adopted, because it gives an economic and national security advantage to countries where AI is not regulated.

Table 1 - Main arguments against AI regulation

Diametrically opposed to these perspectives are proponents who maintain, with considerable conviction, that AI necessitates stringent regulation due to the clear and present risks it poses to public health, individual safety, and the integrity of fundamental rights. The rest of this section will examine the main motivations put forward by the pro-regulation front. However, the section does not endeavour to adjudicate the ultimate validity of these competing claims or to prescribe a definitive stance on the need for AI regulation. Rather, its principal objective is to conduct a review of the motivations most recurrently invoked within scholarly and public discourse, and policymaking circles to justify the development of regulatory frameworks for AI systems. The following analysis, while not aspiring to be exhaustive and perhaps more illustrative than comprehensive in its coverage, nonetheless draws upon discernible patterns of argumentation to furnish concrete evidence for a survey of the principal rationales advanced.

As will be demonstrated, within this complex tapestry of justifications, the imperative of risk minimisation emerges as the cardinal motivation. However, this central concern is often interwoven with a number of peripheral, yet significant, considerations that also merit careful examination in understanding the contemporary impetus towards AI regulation.

2.1 Promotion of innovation and other peripheral motivations

The promotion of innovation emerges as a notable, albeit potentially counterintuitive, peripheral motivation underpinning the drive towards regulating AI. This perspective may appear paradoxical, as a considerable body of thought analysed in the previous section posits that regulatory frameworks inherently act as impediments to innovation, particularly within dynamic fields such as AI.

Notwithstanding these concerns, the notion that regulation can, under specific circumstances, serve as a catalyst for innovation is well-established within both scholarly discourse and policymaking circles. For example, Blind et al. have demonstrated that, contingent upon certain conditions, regulation can exert a positive influence on innovative activities[34]. Building on this, Tartaro et al. contend that the specific characteristics of the AI market are conducive to triggering the beneficial effects of regulation on innovation as identified by Blind et al.[35]. Subsequent research appears to corroborate these findings[36]. This perspective is also reflected in policy initiatives; for instance, a stated objective of the European Union's AI Act is to "Ensure legal certainty to facilitate investment and innovation in AI"[37].

The primary mechanism through which regulation is posited to foster innovation is the establishment of legal certainty for market operators[38]. In nascent and rapidly evolving technological domains such as AI, a lack of clear rules can create an environment of ambiguity and unpredictability. This uncertainty can deter investment, as firms and investors are hesitant to commit resources without a stable framework defining permissible activities, liability, and standards. By providing clear "rules of the game," regulation can reduce these risks, thereby encouraging long-term investment in research, development, and the deployment of new AI applications. Furthermore, well-designed regulation can stimulate innovation by creating demand for new solutions that meet higher safety, ethical, or performance standards. It can also level the playing field, preventing a race to the bottom where firms might otherwise compete by cutting corners on safety or ethical considerations, and instead incentivizing competition based on technological superiority and responsible innovation. Additionally, by signalling societal acceptance and setting clear expectations, regulation can foster consumer trust, which is crucial for the widespread adoption of new technologies like AI, thereby expanding the market for innovative products and services.

Such innovation-spurring effects of regulation have been observed in other domains. In the realm of data privacy, for example, the enactment of increasingly stringent data protection regimes has

---

[34] Knut Blind, Sören S Petersen and Cesare AF Riillo, 'The Impact of Standards and Regulation on Innovation in Uncertain Markets' (2017) 46 Research Policy 249 <https://doi.org/10.1016/j.respol.2016.11.003>.

[35] Tartaro, Smith and Shaw (n 33).

[36] Dasharathraj K Shetty and others, 'Analyzing AI Regulation through Literature and Current Trends' (2025) 11 Journal of Open Innovation: Technology, Market, and Complexity 100508 <https://doi.org/10.1016/j.joitmc.2025.100508>.

[37] '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.

[38] Maria Weimer and Luisa Marin, 'The Role of Law in Managing the Tension between Risk and Innovation: Introduction to the Special Issue on Regulating New and Emerging Technologies' (2016) 7 European Journal of Risk Regulation 469 <https://doi.org/10.1017/S1867299X00006012>.

demonstrably propelled the development and adoption of privacy-enhancing technologies[39]. Similarly, environmental regulation has, in certain instances, been credited with driving innovation in green technologies and more efficient industrial processes[40]. The so-called Porter Hypothesis for instance, suggests that well-designed environmental regulations can trigger innovation by compelling firms to re-evaluate their production processes and products, leading to efficiency gains and the discovery of new, cleaner technologies that can confer competitive advantages[41]. By internalizing environmental externalities, such regulations can create market opportunities for firms that develop innovative solutions to environmental challenges.

In reality, the impact of regulation on innovation is not monolithic but highly variable, contingent upon a confluence of factors including the specific industry sector, prevailing market structures, the nature of the technology itself, and the design of the regulatory intervention[42]. As a matter of fact, ambivalent results on the impact of regulation on innovation have been recorded in the literature[43]. While this matter cannot be further investigated here, we can speculate that he most crucial determinant, however, remains the substantive content of the regulation itself. We can illustrate this through *reductio ad absurdum* in the case of AI. An AI regulation mandating tax exemptions for companies developing AI products and services would almost certainly stimulate innovation in the field as it would provide it provides a powerful economic incentive for investment and development, albeit potentially with other unintended consequences. Conversely, a regulation imposing a blanket prohibition on all AI applications would unequivocally stifle it. Between these fictional extremes lies a vast spectrum of regulatory approaches, each with potentially differing impacts on innovation. Consequently, asserting a priori that regulation universally hinders or promotes innovation is an oversimplification which is simply unverifiable. There exists no universal answer to this conundrum; rather, the effect must be assessed on a case-by-case basis, meticulously considering the specific regulatory design and its context. Moreover, while a robust evaluation of this relationship is best conducted ex post, drawing upon empirical data, the nascent stage of comprehensive AI regulation necessitates an ex-ante assessment, which is inherently more speculative and conjectural.

2.2 A Stiglerian view on AI regulation motivations

Notwithstanding the ongoing scholarly debate concerning the net positive or negative impact of regulation on innovation, the contemporary landscape of AI development presents a compelling empirical scenario. In an environment largely characterised by the absence of bespoke, AI-specific regulatory frameworks and the persistence of significant legal ambiguities, as shown, for example, by unresolved questions surrounding the utilization of copyrighted material for training AI models, a small

[39] Wanyi Chen and others, 'Can the Establishment of a Personal Data Protection System Promote Corporate Innovation?' (2024) 53 Research Policy 105080 <https://doi.org/10.1016/j.respol.2024.105080>.

[40] Peng Zhou, Frank M Song and Xiaoqi Huang, 'Environmental Regulations and Firms' Green Innovations: Transforming Pressure into Incentives' (2023) 86 International Review of Financial Analysis 102504 <https://doi.org/10.1016/j.irfa.2023.102504>.

[41] Stefan Ambec and others, 'The Porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness?' (2013) 7 Review of Environmental Economics and Policy 2 <https://doi.org/10.1093/reep/res016>.

[42] Pontus Braunerhjelm and others (eds), *Handbook of Innovation and Regulation* (Edward Elgar Publishing 2023) <https://doi.org/10.4337/9781800884472> accessed 3 July 2025.

[43] Knut Blind, 'The Overall Impact of Economic, Social and Institutional Regulation on Innovation: An Update' *Handbook of Innovation and Regulation* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781800884472/b-9781800884472.00020.xml> accessed 3 July 2025.

number of pioneering actors has nonetheless achieved profound innovative breakthroughs and established significant market presence.

This very success, paradoxically, may itself engender a demand for AI regulation emanating from these established actors. Having successfully navigated the nascent, relatively unregulated stages of AI development and secured advantageous positions within both the technological ecosystem and the commercial marketplace, these incumbents may now perceive the introduction of regulatory intervention not as a hindrance, but as a strategic instrument to consolidate and protect their hard-won competitive advantages against emerging challengers.

It is, therefore, a noteworthy observation that prominent calls for the implementation of AI regulation have frequently originated from some of the largest and most influential AI labs. For instance, leadership figures from major technology firms and dedicated AI research organisations, such as OpenAI, Anthropic, and xAI, have publicly advocated for varying degrees of regulation for AI[44]. These appeals are often articulated through public statements, testimonies before legislative bodies, and the dissemination of white papers, and are typically framed in terms of ensuring responsible development, promoting public safety, managing existential risks, and upholding ethical principles. While these pronouncements undoubtedly address legitimate societal concerns, they may also strategically align with the interests of entities that have already invested substantially in establishing a dominant position in AI markets. Such organisations might propose regulatory standards that mirror their existing operational practices or advocate for resource-intensive safety protocols, thereby creating a smoother regulatory transition for themselves while simultaneously imposing potentially significant burdens on smaller or newer market entrants who lack comparable resources or established processes.

For these reasons, the phenomenon of established AI actors advocating for regulation can be productively analysed through the theoretical lens of Stigler's theory of economic regulation[45]. Stigler's seminal contribution to the field of political economy posits that, contrary to the conventional "public interest" view which assumes regulation arises primarily to correct market failures or protect diffuse public welfare, regulatory frameworks are often actively sought, acquired, and shaped by the very industries they are ostensibly intended to govern. He argued that well-organized industry groups, characterised by concentrated interests and superior access to information and resources, can exert disproportionate influence over the legislative and regulatory process. Consequently, regulations may be designed, whether explicitly or implicitly, to serve the interests of incumbent firms by, for example, erecting substantial barriers to entry for potential competitors, restricting price competition, standardizing products in ways that favour existing technologies, or even directly subsidizing established players. In essence, Stigler conceptualised regulation as a valuable, tradable good that firms and industries compete to acquire for their private benefit.

Viewed through this Stiglerian framework, the demand for specific forms of AI regulation by leading industry players could, therefore, function as a sophisticated strategic manoeuvre. By championing regulations that entail complex compliance procedures or high thresholds for safety testing and algorithmic auditing, incumbent AI firms might effectively, and perhaps intentionally, elevate the costs

[44] See, for example, Cecilia Kang, 'OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing' *The New York Times* (16 May 2023) <https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html> accessed 3 July 2025; 'CEO of Anthropic Warns against AI Deregulation' (6 June 2025) <https://www.semafor.com/article/06/06/2025/the-ceo-of-ai-company-anthropic-warns-against-deregulation> accessed 3 July 2025; "'Overwhelming Consensus" on AI Regulation - Musk' (13 September 2023) <https://www.bbc.com/news/technology-66804996> accessed 3 July 2025.

[45] George J Stigler, 'The Theory of Economic Regulation' (1971) 2 The Bell Journal of Economics and Management Science 3 <https://doi.org/10.2307/3003160>.

of market entry. Such measures would disproportionately affect smaller startups and potential disruptors that lack the extensive financial, technical, and human capital resources of their larger counterparts. In this manner, calls for regulation, while presented under the commendable auspices of responsible innovation and public safety, could simultaneously serve to entrench existing market power and insulate established leaders from the full force of future competition.

2.3 It's all about risks?

While a multifaceted array of rationale, spanning economic competitiveness, the promotion of innovation, and the incumbents' will to establish barrier to entry, may contribute to the justification for regulatory intervention in the domain of AI, these motivations, when considered in isolation, appear insufficient to account for the current global momentum towards comprehensive AI regulation. A thorough analysis of the contemporary scholarly discourse, public debate, policymaking documents, and the advocacy efforts of non-governmental organisations reveals that it is rather the imperative to mitigate the manifold risks associated with the development and deployment of AI systems that constitutes the most pervasive and compelling driver for regulation. This focus on risk transcends disciplinary and sectoral boundaries, crystallizing a shared sense of urgency around the potential for AI to produce significant, and at times irreversible, harm to individuals, groups, and societal structures. We argue that it is this shared grammar of risk that provides the foundational logic for a proactive, rather than merely reactive, regulatory posture.

This reliance on risk as the principal catalyst for governance is consonant with the broader academic observation of a "riskification" of regulatory governance, a paradigm wherein complex social and technological challenges are primarily framed, analysed, and managed through the calculus of potential negative outcomes[46]. The discourse surrounding artificial intelligence has become a prominent arena for this phenomenon. In the scholarly discourse, the development of AI regulatory frameworks is predicted on the perceived need to anticipate and minimise risks[47]. Consequently, the dominant narrative compelling regulatory action is one of pre-emption and mitigation, where the primary function of law and policy is to serve as a bulwark against potential future harm.

This theoretical orientation towards risk as the central organizing principle for intervention finds its most salient and concrete expression in contemporary legislative efforts. For instance, the pre-eminence of risk mitigation as the crucial motivation for AI regulation is rendered particularly explicit within the AI Act. The European regulation is structurally predicated on a risk-based methodology, which categorises AI systems according to their potential for harm and imposes commensurate regulatory

---

[46] Karen Yeung and Sofia Ranchordás (eds), 'The Rise of Risk in Regulatory Governance' *An Introduction to Law and Regulation: Text and Materials* (2nd edn, Cambridge University Press 2024) <https://doi.org/10.1017/9781009379007.005> accessed 14 June 2025.

[47] See, for example, Matthew U Scherer, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (Social Science Research Network, 30 May 2015) <https://doi.org/10.2139/ssrn.2609777> accessed 29 June 2025; Pedro Rubim Borges Fortes, Pablo Marcello Baquero and David Restrepo Amariles, 'Artificial Intelligence Risks and Algorithmic Regulation' (2022) 13 European Journal of Risk Regulation 357 <https://doi.org/10.1017/err.2022.14>; Giampiero Lupo, 'Risky Artificial Intelligence: The Role of Incidents in the Path to AI Regulation' (2023) 5 Law, Technology and Humans 133 <https://doi.org/10.3316/informit.139102409659905>; Luciano Floridi and others, 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28 Minds and Machines 689 <https://doi.org/10.1007/s11023-018-9482-5>; Gianclaudio Malgieri and Frank Pasquale, 'Licensing High-Risk Artificial Intelligence: Toward Ex Ante Justification for a Disruptive Technology' (2024) 52 Computer Law & Security Review 105899 <https://doi.org/10.1016/j.clsr.2023.105899>; Margot E Kaminski, 'Regulating the Risks of AI' [2022] SSRN Electronic Journal <https://doi.org/10.2139/ssrn.4195066> accessed 15 October 2024; Karl Manheim and Lyric Kaplan, 'Artificial Intelligence: Risks to Privacy and Democracy' (2019) 21 Yale Journal of Law and Technology 106.

obligations. The original Proposal for the AI Act, for instance, unequivocally frames its purpose through this lens, observing that, the "proposal seeks to ensure a high level of protection for those fundamental rights and aims to address various sources of risks through a clearly defined risk-based approach."[48]. This framing demonstrates that the central objective is not to regulate the technology per se, but rather to govern its societal impact by pre-emptively addressing the dangers it may pose, thereby ensuring that its integration into society aligns with fundamental European values.

This emphasis on risk mitigation is by no means unique to the European Union's legislative efforts; rather, it reflects a growing international consensus on the appropriate framework for AI governance. In the United States, for example, the NIST has developed its AI Risk Management Framework, a voluntary but highly influential guide designed to help organisations identify, measure, and manage risks associated with AI systems throughout their lifecycle[49]. Similarly, Canada's Directive on Automated Decision-Making mandates algorithmic impact assessments for governmental bodies, a tool explicitly designed to evaluate and mitigate the risks to fairness, due process, and transparency before a system is deployed[50], given the significant ethical implications of these systems[51]. Even policy initiatives from international bodies like the OECD, while not legally binding, consistently foreground principles such as accountability, safety, and security, which are fundamentally concerned with the prevention and mitigation of risks and adverse outcomes[52]. The convergence around such risk-centric paradigms underscores a global recognition that the responsible stewardship of AI technology is intrinsically linked to the capacity to manage its inherent uncertainties and potential for negative consequences.

Complementing these institutional perspectives, the calls for regulation emanating from civil society are also predominantly articulated through the language of risk, with a particular and urgent focus on the threats posed to fundamental rights. Advocacy groups, academic institutions, and non-governmental organisations such as AlgorithmWatch, the AI Now Institute, and Access Now have been instrumental in documenting and publicizing instances where AI systems have led to discriminatory outcomes, violations of privacy, and potential violation of human rights[53]. Their analyses highlight how automated systems deployed in critical domains, including hiring, credit scoring, criminal justice, and social welfare, can codify and amplify existing societal biases, thereby infringing upon the right to equality and non-discrimination. Furthermore, they raise alarms about the use of AI for mass surveillance and social scoring, which directly implicates the rights to privacy, freedom of association, and freedom of expression. By grounding their advocacy in established international human rights frameworks, these civil society actors frame AI regulation not merely as a matter of market regulation, but as a fundamental imperative for the preservation of human dignity, democratic integrity, and social justice in an increasingly automated world.

---

[48] 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.' <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.

[49] NIST, 'AI Risk Management Framework' <https://www.nist.gov/itl/ai-risk-management-framework>.

[50] Treasury Board of Canada Secretariat, 'Directive on Automated Decision-Making' (2 July 2024) <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> accessed 3 July 2025.

[51] Simona Tiribelli and Benedetta Giovanola, 'Weapons of Moral Construction? On the Value of Fairness in Algorithmic Decision-Making' (2022) 24 Ethics and Information Technology 1 <https://doi.org/10.1007/s10676-022-09622-5>.

[52] OECD, 'The OECD Artificial Intelligence (AI) Principles' <https://oecd.ai/en/ai-principles>.

[53] See, for example, Algorithm Watch, 'Automating Society Report 2020' (*Automating Society Report 2020*, 2020) <https://automatingsociety.algorithmwatch.org> accessed 3 October 2023; AI Now Institute, 'Artificial Power: 2025 Landscape Report' (*AI Now Institute*, 3 June 2025) <https://ainowinstitute.org/publications/research/ai-now-2025-landscape-report> accessed 3 July 2025; Daniel Leufer, 'Why Human Rights Must Be at the Core of AI Governance' (*Access Now*, 24 September 2024) <https://www.accessnow.org/human-rights-and-ai-governance/> accessed 3 July 2025.

Despite the convergence around this motivation, the concept of risk mitigation possesses a paradoxical quality, functioning simultaneously as a powerful unifying force and a subtle occluding mechanism within the discourse on AI regulation. On the one hand, its capacity to consolidate diverse political and social interests under a single, compelling rationale is undeniable. The language of risk provides a common ground for policymakers, scholars, and civil society advocates, framing regulatory intervention not as an ideological project but as a prudent and necessary measure to ensure public safety and societal well-being. Yet, on the other hand, the apparent simplicity and consensus surrounding the need to manage "AI risk" obscures a far more complex and poorly delineated conceptual terrain than is often acknowledged. The term itself operates as a discursive shortcut, masking profound ambiguities, normative disagreements, and fundamental epistemic uncertainties that lie at the heart of the regulatory challenge, as further explored in the present Chapter.

Upon closer examination, the notion of "AI risk" reveals itself not as a monolithic, objectively identifiable category, but as a heterogeneous portmanteau concept that encompasses a vast and divergent array of potential harms. These range from immediate and quantifiable technical failures, such as algorithmic errors leading to tangible economic loss, to more diffuse and insidious societal harms, including the amplification of systemic bias, the erosion of trust in the information ecosystem, and the degradation of democratic institutions. Furthermore, this category extends to encompass long-term, speculative, and even existential risks associated with the prospective development of artificial general intelligence. The overarching rhetoric of risk mitigation tends to conflate these distinct phenomena, treating them as commensurate problems solvable through a unified regulatory framework. This conflation is problematic, as the nature, temporality, and tractability of these varied risks demand equally varied and context-specific regulatory responses.

Moreover, the focus on risk management can depoliticize what are, in essence, deeply political and normative questions about the kind of society we wish to construct. The technical-managerial language of risk management, borrowed from fields like engineering and finance, creates an illusion of objectivity, suggesting that the identification and prioritisation of risks is a neutral, evidence-based exercise[54]. In reality, the determination of what constitutes an "acceptable" or "unacceptable" risk, especially in the field of AI, is an inherently value-laden judgment[55]. It involves weighing competing interests and making profound choices about which societal values to prioritise. By framing the debate primarily in terms of managing a technical problem, the discourse on risk can obscure these underlying value conflicts and marginalize a more fundamental public deliberation about the desirable ends and limits of technological development. Consequently, while the imperative of risk mitigation has effectively catalysed the global drive for AI regulation, a truly robust and legitimate regulatory regime will necessitate a more critical and nuanced deconstruction of the very concept of risk itself, moving beyond its unifying veneer to confront the complex realities it conceals.

The issue is complex, necessitating a clear distinction between perceived risks and those that are empirically demonstrable. Within the latter category, a further differentiation is required between risks characterised by high uncertainty due to a lack of evidence, and those that are more definitively established. The discourse surrounding AI is currently dominated by a perception of risk that appears sufficiently potent to justify, in and of itself, significant regulatory intervention. This perception is fuelled by narratives of existential threats, from autonomous systems evading human control to apocalyptic scenarios of mass job displacement and the subtle erosion of fundamental rights through automated

[54] Mary Douglas, 'The Depoliticization of Risk' *Culture Matters* (Routledge 1997).

[55] Alessio Tartaro, 'Value-Laden Challenges for Technical Standards Supporting Regulation in the Field of AI' (2024) 26 Ethics and Information Technology 72 <https://doi.org/10.1007/s10676-024-09809-y>.

decision-making. Such anxieties are not confined to the general populace; among experts in the field, a notable contingent posits that advanced AI could pose a genuine existential threat to humanity[56].

This heightened sense of risk is amplified and disseminated through various channels[57]. Media outlets frequently highlight catastrophic or dystopian potential, framing AI development in terms of an uncontrolled race with perilous consequences for social stability and individual autonomy. This narrative of imminent danger is paradoxically reinforced by prominent figures within the AI industry itself. As has been noted, the chief executive officers of leading AI development companies, including firms such as Anthropic and xAI, have publicly articulated the necessity for, and even championed the urgent implementation of, robust regulatory guardrails. These discursive interventions from industry leaders, while potentially serving complex strategic interests, function as a powerful, implicit validation of the premise that the technology they are creating carries profound and potentially unmanageable risks. By advocating for external state-led oversight, these technologists tacitly concede that the trajectory of AI development may outstrip the sector's own capacity for safe and effective self-governance, thereby reinforcing the logic that the risks are not merely theoretical but tangible and imminent.

This process of validation and legitimation, originating from within the technological sphere, has been mirrored and amplified within the political realm. The concerns articulated by industry insiders have been readily absorbed into mainstream political discourse, conferring a powerful sanction upon the risk-centric narrative. A particularly salient example of this convergence can be found in the 2023 State of the Union address by the President of the European Commission, who explicitly invoked the specter of existential threats posed by AI[58]. The enunciation of such grave concerns in a major policy speech by a high-ranking political leader served to institutionalize the discourse of AI-related risk at the highest level of governance. This act of political framing elevates the issue beyond a technical or economic problem, repositioning it as a fundamental challenge to societal stability and security, and thereby solidifying the political will necessary for pre-emptive regulatory action.

However, a significant perception gap often exists between public and even expert perception of risk and its objective, evidence-based reality[59]. This discrepancy between what is feared and what can be proven is a central challenge in formulating sound policy.

A salient example of this gap can be found in the widely discussed risk of algorithmic bias in recruitment processes. The concern that automated hiring tools may systematically disadvantage candidates from protected groups is a cornerstone of the critical literature on AI ethics[60]. Upon closer scrutiny, however, the empirical evidence for the widespread adoption of AI systems in hiring is remarkably thin. It is exceedingly difficult to obtain reliable data on the use and impact of AI in hiring decisions. Existing sources present wildly divergent figures, with some reporting that 24% of firms utilize

---

[56] Hugo Neri and Fabio Cozman, 'The Role of Experts in the Public Perception of Risk of Artificial Intelligence' (2020) 35 AI & SOCIETY 663 <https://doi.org/10.1007/s00146-019-00924-9>.

[57] Hugo Neri, *The Risk Perception of Artificial Intelligence* (Bloomsbury Publishing PLC 2020).

[58] Alessio Tartaro, 'Alessio Tartaro on the "State of the European Union Address"' (*Critical AI*, 10 October 2023) <https://criticalai.org/2023/10/10/alessio-tartaro-on-the-state-of-the-european-union-address-9-13-2023/> accessed 3 July 2025.

[59] David Ropeik, 'The Perception Gap: Recognizing and Managing the Risks That Arise When We Get Risk Wrong' (2012) 50 Food and Chemical Toxicology 1222 <https://doi.org/10.1016/j.fct.2012.02.015>; David Ropeik, *How Risky Is It, Really?: Why Our Fears Don't Always Match the Facts* (McGraw Hill Professional 2010).

[60] See, for example, Nicholas Tilmes, 'Disability, Fairness, and Algorithmic Bias in AI Recruitment' (2022) 24 Ethics and Information Technology 21 <https://doi.org/10.1007/s10676-022-09633-2>; Elham Albaroudi, Taha Mansouri and Ali Alameer, 'A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring' (2024) 5 AI 383 <https://doi.org/10.3390/ai5010019>.

AI in recruitment, while others claim the figure is as high as seventy-five percent 75%[61]. These divergent figures often stem from significant methodological flaws. In addition, these results suffer from inadequate scoping, frequently failing to distinguish between low-risk applications, such as using AI-based chatbots for engaging applicants, and high-risk implementations where an algorithm recommends or makes a final hiring decision. In other cases, simple Applicant Tracking Systems are misclassified as genuine AI, a terminological inflation that serves marketing purposes but severely undermines research validity[62].

A similar discrepancy between perceived and real risk emerges when considering the deployment of autonomous systems, such as self-driving vehicles. Public and regulatory reactions to accidents involving autonomous technology are often disproportionately severe compared to the response to the far greater number of incidents caused by human error[63]. The novelty of the technology amplify the sense of risk, even if statistical analyses might eventually demonstrate a superior safety record for the automated system. This illustrates how the perceived risk profile of an AI application can be misaligned with its empirical harm profile, complicating the evaluations for regulation.

This is not to suggest, of course, that AI risks are illusory. There are indeed latent or emergent risks that we are currently ill-equipped to measure or even fully comprehend. For instance, studies are only now beginning to investigate the long-term cognitive effects of increasing reliance on AI tools. The phenomenon of "cognitive debt," wherein the outsourcing of critical thinking, memory, and problem-solving to intelligent systems may lead to an atrophy of human cognitive abilities, represents a subtle, slow-developing risk whose societal impact is presently unknown but potentially profound[64]. In a similar vein, the proliferation of highly convincing synthetic media poses a systemic threat to the shared epistemic commons. The gradual erosion of trust in information sources and the very concept of objective reality is a second-order, societal-level risk that defies simple quantification but could have corrosive effects on democratic processes and social cohesion[65]. The emergence of AI-enabled sexual robots[66] and companion robots[67] will impact human relationships, raising risks not well understood yet.

Finally, beyond the categories of perceived and emerging risk lies a third domain, risks that are existing and well-documented. Concrete evidence demonstrates bias in applications ranging from medical diagnostics[68], where algorithms have shown lower accuracy for certain demographic groups, to automated content moderation, where even well optimized AI systems could exacerbate, rather than relieve, many

[61] See, for example, 'AI Recruitment Statistics: What Is the Future of Hiring?' (*Tidio*, 5 March 2025) <https://www.tidio.com/blog/ai-recruitment/> accessed 5 July 2025; 'How Many Companies Use AI In Hiring & Recruiting? (2025)' (7 February 2025) <https://joingenius.com/statistics/how-many-companies-use-ai-in-hiring/> accessed 5 July 2025.

[62] Andrew Fennell, 'AI in Recruitment Statistics UK 2025 | Latest Reports & Data' (12 June 2024) <https://standout-cv.com/stats/ai-in-recruitment-statistics-uk> accessed 5 July 2025.

[63] Praveena Penmetsa and others, 'Effects of the Autonomous Vehicle Crashes on Public Perception of the Technology' (2021) 45 IATSS Research 485 <https://doi.org/10.1016/j.iatssr.2021.04.003>.

[64] Nataliya Kosmyna and others, 'Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task' (arXiv, 10 June 2025) <https://doi.org/10.48550/arXiv.2506.08872> accessed 4 July 2025.

[65] John Wihbey, 'AI and Epistemic Risk for Democracy: A Coming Crisis of Public Knowledge?' (Social Science Research Network, 20 April 2024) <https://doi.org/10.2139/ssrn.4805026> accessed 5 July 2025.

[66] Piercosma Bisconti, 'Will Sexual Robots Modify Human Relationships? A Psychological Approach to Reframe the Symbolic Argument' (2021) 35 Advanced Robotics 561 <https://doi.org/10.1080/01691864.2021.1886167>.

[67] Piercosma Bisconti Lucidi and Daniele Nardi, 'Companion Robots: The Hallucinatory Danger of Human-Robot Interactions' *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (Association for Computing Machinery 2018) <https://doi.org/10.1145/3278721.3278741> accessed 20 April 2026.

[68] Richard J Chen and others, 'Algorithmic Fairness in Artificial Intelligence for Medicine and Healthcare' (2023) 7 Nature Biomedical Engineering 719 <https://doi.org/10.1038/s41551-023-01056-8>.

existing problems in content moderation due to their opacity and the bias ingrained in the system[69]. The crucial question, however, remains open: whether these documented, evidence-backed harms are, on their own, sufficient to justify the creation of a comprehensive, *sui generis* regulatory framework for AI. It is debatable whether existing legal and professional paradigms, such as established anti-discrimination laws, are not already adequate to address these specific, verifiable failures.

In conclusion, this chapter has not sought to definitively resolve the overarching question of whether the motivations for regulating artificial intelligence are sufficiently robust, nor has it aimed to adjudicate the specific soundness of risk mitigation as a primary justification for regulatory intervention. As the preceding analysis has demonstrated, the complexity of the issue precludes such a definitive conclusion. A principal finding is that the various rationales examined, from the promotion of innovation to the minimisation of risk, appear to lack a firm and comprehensive empirical foundation. This evidentiary deficit stems from a profound uncertainty about how the dynamic between regulation and innovation will unfold in the field of AI, as well as from our current methodological inability to adequately capture and measure the multifaceted nature and impacts of AI-related risks. As will be substantiated through the specific analysis of the EU AI Act, this state of affairs signals a notable departure from the stringent criteria of evidence-based policymaking that has traditionally governed technological risk regulation.

Consequently, while leaving the foundational question of motivation provisionally unresolved, the analysis can pivot to an incontrovertible fact: the regulation of AI is an established and unfolding reality, irrespective of the precise justifications that propel it. The political decision to intervene has been made. For this reason, the focus of the inquiry must now shift from the *motivations for* regulation to the *challenges of* its implementation. The subsequent sections will therefore proceed to dissect the hurdles that emerge once policymakers, acting upon what are held to be compelling reasons, embark upon the task of governing the development and deployment of artificial intelligence systems

3. Challenges

Regulation represents a very complex challenge when directed at emerging technologies whose potential impacts and future developments are highly uncertain, such as AI. This chapter provides a survey of these challenges. The analysis will commence with a discussion of difficulties concerning technology regulation more broadly. Subsequently, the chapter will argue that such general challenges are not only replicated but are substantially exacerbated within the specific context of AI. The analysis will then narrow to address a challenge intrinsic to the technology itself: the singular and often elusive nature of AI-related risks, which complicates conventional approaches to risk identification and mitigation.

3.1. Pacing problems, regulatory disconnect & the Collingridge dilemma

The rapid and relentless pace of innovation in areas like artificial intelligence presents a very well-known challenge for regulation. This phenomenon, often termed the "pacing problem," encapsulates the fundamental challenge arising when the trajectory and velocity of technological development surpass the

[69] R Binns and others, 'Like Trainer, like Bot? Inheritance of Bias in Algorithmic Content Moderation' (2017) 10540 LNCS Social Informatics (SocInfo 2017) <https://ora.ox.ac.uk/objects/uuid:b0723053-f564-4a49-be8f-85e1b0780ea7> accessed 5 July 2025; Robert Gorwa, Reuben Binns and Christian Katzenbach, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945 <https://doi.org/10.1177/2053951719897945>.

capacity of the legal, ethical, and policy frameworks designed to oversee it[70]. This phenomenon has been evocatively captured by the metaphor of "the hare and the tortoise,"[71] where the relentlessly accelerating march of technology assumes the role of the swift hare, while the process of regulation, inherently more deliberate and often constrained by institutional processes, embodies the slower tortoise, perpetually striving to catch up to an ever-receding finish line.

Several key aspects contribute to the genesis and persistence of the pacing problem. Firstly, contemporary innovation cycles for novel technologies are dramatically shorter than those historically associated with industrial or foundational technological shifts, sometimes compressing into a few years rather than spanning decades[72]. This contrasts sharply with the inherently slower, more deliberate, and consultation-heavy cadence characteristic of traditional administrative rulemaking and legislative processes. Secondly, many existing legal frameworks are predicated upon what can be described as a static or comparatively stable view of technology and societal structures, rendering them inherently ill-suited to adapt dynamically and anticipatorily to conditions of continuous and often unpredictable technological flux[73]. Furthermore, regulatory institutions themselves can often be characterised by a degree of stagnation, ossification, and bureaucratic inertia, which inherently attenuates their capacity to adjust promptly and effectively to the challenges posed by rapidly changing technologies[74].

Expanding beyond the temporal lag, "regulatory disconnection" is a more expansive concept that encompasses the pacing problem while also extending to a broader array of mismatches between the prevailing regulatory environment and the dynamically evolving technological landscape. It fundamentally signifies that the current regulatory framework, having been forged in response to past technological realities, is no longer adequate, appropriate, or effectively aligned with the innovations and practices characteristic of the present day[75].

Regulatory disconnection can manifest itself through various modalities[76]. There is, firstly, "descriptive disconnection", which arises when the language, terminology, or conceptual categories embedded within regulations no longer accurately or adequately describe the technology, its functionalities, or the specific practices that the regulation intends to govern or influence[77]. Secondly, "normative disconnection" occurs when the underlying values, fundamental principles, or core policy objectives upon which a regulatory regime is founded are challenged or become questionable in light of

---

[70] For a comprehensive discussion about this problem see, Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, vol 7 (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7> accessed 22 October 2024.

[71] David Rejeski, 'Public Policy on the Technological Frontier' in Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7_4> accessed 5 July 2025.

[72] Mark Fenwick, Wulf Kaal and Erik Vermeulen, 'Regulation Tomorrow: What Happens When Technology Is Faster than the Law?' (2017) 6 American University Business Law Review <https://digitalcommons.wcl.american.edu/aublr/vol6/iss3/1>.

[73] Gary E Marchant, 'The Growing Gap Between Emerging Technologies and the Law' in Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7_2> accessed 5 July 2025.

[74] Gary E Marchant, 'Addressing the Pacing Problem' in Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7_13> accessed 5 July 2025.

[75] Anna Butenko and Pierre Larouche, 'Regulation for Innovativeness or Regulation of Innovation?' (2015) 7 Law, Innovation and Technology 52 <https://doi.org/10.1080/17579961.2015.1052643>.

[76] Carolyn Abbot, 'Bridging the Gap – Non-State Actors and the Challenges of Regulating New Technology' (2012) 39 Journal of Law and Society 329 <https://doi.org/10.1111/j.1467-6478.2012.00588.x>.

[77] Roger Brownsword and Roger Brownsword, *Rights, Regulation, and the Technological Revolution* (Oxford University Press 2008).

new technological developments[78]. This can happen when an emergent technology raises profound ethical dilemmas or poses policy questions that the existing normative framework was neither designed nor equipped to anticipate or resolve. Furthermore, new technologies may emerge into a complete regulatory void, an environment where no specific laws or regulations exist that are directly applicable, or they may encounter significant gaps within existing legislative frameworks, leaving crucial aspects unregulated[79]. This lack of clear regulatory coverage is particularly problematic given the inherent uncertainty surrounding the potential risks and benefits associated with novel innovations. Finally, disconnection can also stem from the inadequacy of existing regulatory frameworks in addressing new social norms, behaviours, or capabilities that are directly enabled or facilitated by technology, such as the widespread digital copying and sharing of music files despite copyright laws designed for physical media[80].

The cumulative impact of the pacing problem and regulatory disconnection gives rise to several critical and interlocking issues. A primary consequence is heightened uncertainty for regulated entities. Innovators may face ambiguity regarding their compliance obligations, unsure of where they stand legally or precisely what rules apply to novel activities or technological deployments, leading to significant difficulties in ensuring compliance. Paradoxically, regulatory rules that are highly specific, often intended to enhance clarity, can quickly become descriptively or normatively disconnected from the technological reality, thereby exacerbating confusion rather than alleviating it[81]. Regulators themselves are frequently confronted with a difficult dilemma, caught between the imperative to act and the lack of comprehensive understanding: they must either risk acting 'recklessly' by regulating without sufficient factual basis or fall into a state of 'paralysis' by doing nothing, allowing risks to proliferate unchecked[82].

These problems, which affect technology regulation in general, appear significantly accentuated within the domain of AI. This amplification is, firstly, attributable to the fact that AI, perhaps more than any other technological field in recent history, has undergone phases of abrupt and rapid acceleration. The demonstrable struggle of regulatory frameworks, including ambitious legislative initiatives aimed at governing AI, to keep pace with these swift shifts is readily apparent. The trajectory of the AI Act serves as a particularly salient example in this regard. When the European Commission initially proposed the regulation in April 2021, sophisticated tools such as LLMs, generative AI systems, AI agents, and other advanced developments characteristic of the subsequent two to three years had not yet achieved their current prominence or widespread adoption. Consequently, the initial proposal for the AI Act contained no specific provisions directly addressing these emergent models or their unique characteristics and risks. It was only in extremis, during the protracted legislative process, that specific measures targeting what were eventually designated as GPAI systems were introduced and incorporated into the text. This reactive incorporation, necessitated by unforeseen technological breakthroughs occurring mid-policy-formation, vividly illustrates the manifestation of the pacing problem within a live regulatory effort. As will be explored in subsequent sections, this phenomenon also points directly to another critical challenge: the inherent information asymmetries that exist between regulatory bodies and the actors operating at the cutting edge of technological innovation.

This rapid and somewhat unpredictable evolution further exacerbates the problem of regulatory disconnection. If, hypothetically, the AI Act had been finalised and adopted just prior to the widespread

[78] ibid.

[79] Lyria Bennett Moses, 'How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target' (2013) 5 Law, Innovation and Technology 1 <https://doi.org/10.5235/17579961.5.1.1>.

[80] Brownsword and Brownsword (n 77).

[81] Roger Brownsword, 'So What Does the World Need Now? Reflections on Regulating Technologies' in Roger Brownsword and Karen Yeung (eds), *Regulating technologies* (Hart 2008).

[82] Fenwick, Kaal and Vermeulen (n 72).

public emergence of powerful generative AI models encompassing text generators (e.g., ChatGPT, Gemini, Claude, LLaMA etc.), image generators (e.g, Midjourney, DALL-E, Stable Diffusion, Imagen etc.), and video generation tools (e.g. Sora, or RunwayML Gen-1/Gen-2, Pika Labs etc.), it is highly probable that the regulatory instrument would have found itself fundamentally disconnected from the technological reality it was designed to govern almost immediately upon entry into force. Given the continuing pace of technological advancement, it cannot be definitively excluded that the AI Act, even in its finally adopted form, may face a similar fate. For instance, the nascent emergence of platform-like distribution models for sophisticated AI Agents already appears to challenge the clear-cut distinction between "providers" and "deployers", a dichotomy that constitutes one of the foundational pillars of the AI Act and upon which the allocation of responsibilities to regulatees is largely predicated. The dynamic nature and potentially emergent behaviours of AI Agents, coupled with complex and multi-layered supply chains, strain this binary classification, potentially leading to ambiguities and gaps in accountability that were not fully anticipated during the legislative drafting process[83].

Consequently, the field of AI regulation appears to squarely confront the challenges encapsulated by the Collingridge dilemma[84]. The Collingridge dilemma describes a fundamental and persistent challenge in the regulation of novel technologies, particularly concerning the optimal timing of regulatory intervention relative to the technology's lifecycle. It presents regulators with two principal hurdles, each fraught with significant risks. Intervening "too early," in the nascent stages of a technology's development, carries the risk of insufficient information. At this juncture, regulators typically lack a comprehensive understanding of the technology's ultimate form, its full spectrum of potential applications, and its eventual societal impacts, both beneficial and detrimental. Consequently, imposing regulations prematurely risks over-regulating the field, potentially stifling valuable innovation before it can fully mature, or alternatively, failing to accurately identify and target the specific risks that will eventually materialise. Conversely, deferring intervention until "too late" means waiting until the technology's impacts, including its problematic consequences, become fully manifest and clearly understood. By this later stage, the technology may have become deeply embedded within societal infrastructures, integrated into existing economic systems, and solidified in its design architecture[85]. At this point, implementing necessary regulatory changes becomes significantly more expensive, politically difficult, and potentially even technically impossible due to phenomena such as "technological lock-in," where the costs of transitioning away from an established technology become prohibitive[86].

The application of the Collingridge dilemma is acutely relevant to the regulation of AI, largely due to AI's inherent characteristics as a rapidly evolving and profoundly disruptive technology. One key facet is the fundamentally unforeseeable nature of AI development. The swift pace of advancements, coupled with the unpredictable and potentially vast number and types of future applications, renders it exceptionally challenging for regulators to anticipate future trajectories and respond effectively in a timely manner. Policymakers frequently operate with incomplete or insufficient knowledge regarding the precise technical nature of emerging AI technologies, the most effective strategies for mitigating unwelcome effects, and precisely which specific applications warrant the most immediate or stringent regulatory attention.

---

[83] Luca Nannini and others, 'AI Agents Under EU Law' (arXiv, 6 April 2026) <https://doi.org/10.48550/arXiv.2604.04604> accessed 20 April 2026.

[84] David Collingridge, *The Social Control of Technology* (London : Frances Pinter 1980) <http://archive.org/details/socialcontroloft0000coll> accessed 5 July 2025.

[85] Moses (n 79).

[86] Mihály Héder, 'AI and the Resurrection of Technological Determinism' (2021) 21 Információs Társadalom: Társadalomtudományi Folyóirat 119.

The dilemma's "too early" horn manifests in the substantial risk of over-regulation. If regulatory bodies intervene prematurely with stringent, potentially prescriptive rules based on incomplete understanding, they risk stifling the very innovation that holds immense potential for societal benefit across diverse fields such as healthcare diagnostics or financial services optimisation. This inherent tension between managing potential AI risks and simultaneously promoting beneficial innovation is a central theme in contemporary policy discussions surrounding AI regulation[87]. The "too late" horn of the dilemma presents an equally daunting prospect. Waiting for the full manifestation of significant AI-related problems, such as the mass manipulation of consumer vulnerabilities, the systemic entrenchment of algorithmic discrimination, or the creation of significant power asymmetries, means that these technologies could become deeply integrated into critical societal infrastructures and economic systems. Implementing meaningful regulation at such a late stage becomes exceedingly costly, politically contentious, and technically complex.

3.2. Uncertainties and asymmetries

Asymmetries between regulatees and regulators constitute an additional, and profoundly multifaceted, challenge in the governance of technology. These disparities manifest primarily across critical dimensions: knowledge, resources, and power. Understanding these imbalances is fundamental to appreciating the complexities inherent in designing and implementing effective regulatory frameworks in rapidly evolving technological landscapes.

A primary challenge is the pervasive information asymmetry that characterises the relationship between regulatory bodies and the entities they are tasked with overseeing[88]. This imbalance signifies that governmental authorities are often structurally incapable of possessing the same depth and specificity of knowledge about a particular industry, its internal operations, technical nuances, and future trajectories as the industry itself holds. This problem transcends a simple lack of accessible data; it points towards a more fundamental fragmentation of knowledge, suggesting that no single actor or institution within the regulatory ecosystem possesses all the necessary information or perspective to fully comprehend and effectively resolve complex, technology-driven problems[89]. This fragmentation includes a critical deficit in expertise and practical experience within state regulatory bodies when compared to the highly specialised technical knowledge resident within the firms developing and deploying cutting-edge technologies[90]. Consequently, regulators may find themselves lacking the sufficient understanding and specific knowledge required to adequately exercise their discretionary powers in regulatory matters, potentially leading to suboptimal or misguided interventions.

This inherent lack of comprehensive information and technical expertise within the state apparatus has profound implications for the operational efficacy and design of regulatory approaches. A purely traditional, state-centric regulatory model, heavily reliant on command-and-control mechanisms predicated on the regulator's presumed superior knowledge, becomes significantly less effective in achieving desired policy outcomes when confronted by industries possessing a substantial informational advantage. Regulators consequently struggle not only to keep pace with the accelerating rate of

[87] Araz Taeihagh, 'Governance of Artificial Intelligence' (2021) 40 Policy and Society 137 <https://doi.org/10.1080/14494035.2021.1928377>.

[88] Araz Taeihagh, M Ramesh and Michael Howlett, 'Assessing the Regulatory Challenges of Emerging Disruptive Technologies' (2021) 15 Regulation & Governance 1009 <https://doi.org/10.1111/rego.12392>.

[89] Julia Black, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (2001) 54 Current Legal Problems 103 <https://doi.org/10.1093/clp/54.1.103>.

[90] Abbot (n 76).

technological development but also to formulate and implement adequate responses to the novel challenges and risks that emerge. The information asymmetry complicates even the most fundamental decisions regulators must make: determining what aspects of a technology or practice warrant intervention, deciding on the optimal timing for such intervention, and selecting the most appropriate regulatory instruments. Establishing reliable, uncontested facts about new technologies, their functionalities, and their impacts can be exceedingly difficult, often involving contested interpretations and confronting the challenge of "unknown unknowns", i.e., risks or consequences that are not only uncertain but entirely unanticipated. Even regulatory approaches designed for flexibility, such as principles-based regulation, require regulators to possess a substantial underlying understanding of the risks and potential choices involved in applying those principles, a prerequisite that is often lacking for genuinely novel technologies.

Beyond the informational deficit, asymmetries in resources and inherent power dynamics further skew the regulatory landscape. Highly organised societal sectors, commanding significant financial and technical resources, can effectively resist political control and regulatory direction through lobbying efforts and regulatory capture. Technology firms, in particular, are adept at strategically framing policy issues, often portraying areas of concern as inherently uncertain or requiring highly specialised technical expertise beyond the grasp of regulators. This framing can serve to delay regulatory action or steer policy design in directions more favourable to industry interests, as documented in the field of AI[91].

The challenges to regulatory legitimacy and public trust are also deeply intertwined with these asymmetries. The perceived conflict of interest that arises when governments are simultaneously tasked with promoting technological innovation and regulating its potential harms can significantly undermine public confidence in the fairness and impartiality of regulatory schemes. Concerns about agency capture influenced by well-organised interest groups further exacerbate public scepticism towards policymakers and scientific advisors involved in regulation. The development and operationalisation of complex AI systems, for instance, have frequently been shaped by a comparatively small and often non-diverse group of actors, creating inherent power imbalances that can leave other jurisdictions, stakeholders, and even national regulators in positions of dependence on the technical decisions made elsewhere. While concepts like "trustworthy AI" are ostensibly aimed at generating public confidence and legitimacy, they can also be strategically instrumentalised by industry as means to legitimise innovative practices and gain market advantage, potentially compromising the purported aim of centring fundamental rights within governance due to the persistent influence of industry priorities[92].

Finally, these asymmetries also impinge directly upon the practical difficulties of enforcing regulations and defining their appropriate scope[93]. Regulators may simply lack the necessary technical wherewithal to fully comprehend, monitor, and effectively enforce compliance with obligations related to highly complex technologies, such as intricate AI model behaviours. This technical deficit can render enforcement efforts impotent or easily circumvented. Furthermore, information and resource asymmetries can lead to instances of asymmetric regulation, where nominally comparable situations or

[91] Kevin Wei and others, 'How Do AI Companies "Fine-Tune" Policy? Examining Regulatory Capture in AI Governance' (Social Science Research Network, 20 August 2024) <https://doi.org/10.2139/ssrn.4931927> accessed 9 July 2025.

[92] Andrea Saltelli and others, 'Science, the Endless Frontier of Regulatory Capture' (2022) 135 Futures 102860 <https://doi.org/10.1016/j.futures.2021.102860>.

[93] Bronwen Morgan and Karen Yeung (eds), 'Regulatory Enforcement and Compliance' *An Introduction to Law and Regulation: Text and Materials* (Cambridge University Press 2007) <https://doi.org/10.1017/CBO9780511801112.005> accessed 9 July 2025.

entities are treated differently by the regulatory system[94]. This can occur, for instance, if regulators possess detailed information and oversight capabilities for traditional market actors but lack them for new digital entrants, or if well-resourced firms can navigate complex rules more effectively than smaller players. This can distort competition, create unfair advantages, and potentially protect established incumbents against disruptive new entrants.

In essence, the inherent differences in knowledge, access to resources, and capacity for influence between those who innovate and those who regulate create a persistent and fundamental tension. This tension significantly impedes the capacity of regulatory systems to adapt rapidly to technological change, make fully informed decisions based on comprehensive understanding, ensure fairness and equitable treatment among regulatees, and ultimately maintain public trust in the legitimacy and effectiveness of governance mechanisms in an era of accelerating technological transformation. Addressing these deeply rooted asymmetries mandates a nuanced approach, often requiring a blend of regulatory tools involving both state and non-state actors, a commitment to continuous learning and capacity building within regulatory institutions, and mechanisms to actively counterbalance the epistemic and power advantages held by regulated industries.

Once again, these challenges are not merely present but are significantly intensified within the domain of AI. Knowledge, resource, and power asymmetries manifest as particularly acute impediments to the effective regulation of AI.

The challenge of knowledge asymmetry is fundamental to AI regulation. From a "decentred" perspective on regulation, this is not merely a temporary deficit of information but an acknowledgement that no single actor, including the state, possesses all the knowledge necessary to fully grasp the complexities of AI and its multifaceted societal implications, let alone devise perfect solutions to the problems it presents. The issue transcends simple information asymmetry. As Black puts it, knowledge itself is understood as socially constructed and fragmented across various distinct, "autopoietically closed sub-systems"—such as the political, administrative, legal, and technological realms—each of which constructs its understanding of other systems through its own inherent logic and, crucially, its own "distorting lens"[95]. This means that even access to raw data does not guarantee shared understanding; the interpretation and significance of technical details about AI models, their capabilities, and their risks are viewed differently through the distinct lenses of, say, a computer scientist designing the model, a lawyer attempting to draft liability rules, and a policymaker weighing economic competitiveness against fundamental rights. This fragmentation is compounded by a demonstrable lack of deep technical expertise and practical experience within many state regulatory bodies when contrasted with the highly specialised knowledge pools and lucrative talent markets commanded by firms operating at the technological frontier. Regulators may thus find themselves inadequately equipped with the specific technical knowledge required to exercise discretion effectively in regulatory matters, understand complex system behaviours, or anticipate novel failure modes.

Compounding these epistemic challenges are profound resource asymmetries. These highlight the vastly unequal distribution of the assets essential not only for developing and deploying cutting-edge AI but also for effectively understanding and regulating it. The global AI market exhibits a highly concentrated, effectively oligopolistic structure dominated by a small cohort of key players, primarily located in the United States and China[96]. These firms control disproportionate shares of the critical

[94] Daniel Gervais, 'The Regulation of Inchoate Technologies' (2010) 47 Houston Law Review 665.

[95] Black (n 58), p. 107.

[96] Lily Ballot Jones, Julia Thornton and Daswin De Silva, 'Limitations of Risk Based Artificial Intelligence Regulation: A Structuration Theory Approach'.

resources required for advanced AI development. This control extends to tangible physical assets, including massive clusters of high-performance GPUs, access to cheap and reliable electricity, and vast data storage infrastructure. It also encompasses vital epistemic resources, such as the ability to attract and retain top AI researchers through competitive salaries and cutting-edge facilities, and crucial informational resources, including access to proprietary, extensive, and often labelled datasets, alongside control over live systems that facilitate continuous data collection and real-world experimentation for model refinement.

Finally, power asymmetry reflects the uneven distribution of influence and control over the trajectory of AI development and its governance. This dispersion of power is particularly evident in the realm of private governance through contracts and standards. Private actors increasingly employ contractual terms and licenses (such as emerging "Responsible AI Licenses") to govern the use and even dictate the acceptable outputs of AI models[97]. These private instruments sometimes directly incorporate, or mirror behavioural restrictions or ethical guidelines proposed in regulatory discussions. Furthermore, as we shall see in greater details in the next chapters, industry-led technical standards, historically developed by self-regulatory bodies and industry associations, often extend beyond mere functional requirements to embed specific values and influence design patterns related to human rights or ethical considerations. This grants standard-setting bodies a significant, albeit often understated, global governance role, as substantive policy decisions can become baked in at a fundamental technical level. Consequently, governments are increasingly adopting hybrid regulatory approaches that seek to incentivise or formally reference private standards to align them with broader legal and policy objectives[98].

Ultimately, the legitimacy and accountability of AI regulation are fundamentally challenged by the involvement of powerful non-state actors and the increasing reliance on governance mechanisms embedded within technology itself. Concerns arise about the potential for the delegation of powers to technical committees or private standard-setting bodies whose decision-making processes may lack the transparency and public accountability characteristic of democratic institutions[99]. The inherent opacity of many complex AI systems and their underlying business models further hinders the ability of individuals or regulators to understand how decisions are made, to challenge those decisions effectively, or to determine responsibility when things go wrong, impacting fundamental principles of due process and fairness. This also elevates the risk of regulatory capture, where regulatory agencies become overly influenced by the powerful industry they are tasked with overseeing. In essence, the interwoven nature of knowledge, resource, and power asymmetries means that those who possess the capacity to develop, control access to, and deploy AI technologies often exert significant influence over the terms under which these technologies are understood, debated, and ultimately regulated, posing a fundamental challenge to the pursuit of effective and equitable regulatory outcomes.

3.3. The nature of AI risks

Whereas the previous sections has delineated regulatory challenges pertinent to technology writ large, challenges that are acutely applicable to the domain of AI, the discussion in this last section pivots to a distinct and intrinsic challenge for AI regulation. This challenge emanates not from characteristics of technologies in general, but from the very nature of the AI risks that regulatory frameworks are designed

[97] Michael Veale, Kira Matus and Robert Gorwa, 'AI and Global Governance: Modalities, Rationales, Tensions' (2023) 19 Annual Review of Law and Social Science 255 <https://doi.org/10.1146/annurev-lawsocsci-020223-040749>.

[98] ibid.

[99] Irene Kamara, *Legal Aspects of Standardisation: Relationship of Standards and Law in the EU* (European Commission 2023).

to minimise and mitigate. As established at the outset of the present Chapter, the profound complexity of risks associated with AI constitutes a primary challenge for effective regulation.

AI risks are indeed profoundly multifaceted, encompassing a broad spectrum of potential harms that arise from a complex and often unpredictable interplay of technical properties, human-computer interaction dynamics, and overarching systemic contexts. In response to this sprawling and fragmented landscape of potential harms, significant scholarly and institutional efforts have been directed towards systematisation. A preeminent example of such an endeavour is the AI Risk Repository developed at the MIT[100]. Through exhaustive meta-reviews, the creation of comprehensive databases, and the development of structured taxonomies, this repository seeks to impose order on the conceptual chaos. The project synthesizes a vast corpus of existing scholarship, meticulously extracting and codifying 1612 distinct risk types from a diverse collection of 65 pre-existing taxonomies. The explicit objective of this large-scale synthesis is to furnish a more unified and systematically structured framework for categorising and, consequently, understanding these multifarious hazards.

The conceptual architecture of the repository hinges upon two principal, complementary taxonomical frameworks that disaggregate the multifaceted nature of AI risk along different analytical axes: the Causal Taxonomy of AI Risks and the Domain Taxonomy of AI Risks.

The Causal Taxonomy provides a framework for classifying risks based on the underlying mechanisms or circumstances that precipitate their emergence – essentially addressing how, when, or why a risk might materialise. It posits three principal causal factors. The first, Entity, identifies the primary agent presented as the source of the risk due to a particular decision or action. Within this category, 'Human-caused risks' are those attributed to deliberate or unintentional actions undertaken by human actors, such as tthe intentional design of AI systems for malicious purposes, or the improper or negligent use of a deployed AI system. Empirical data from the repository demonstrates that certain risk categories are overwhelmingly attributed to human agency; for instance, risks such as fraud, scams, and targeted manipulation are presented as human-caused in a significant majority of cases (79%), while the risk of power centralisation and unfair distribution of benefits is likewise largely seen as originating from human decisions (72%). Conversely, 'AI-caused risks' are those deemed to originate from the decisions, behaviours, or emergent actions of the AI system itself. The data indicates that risks like the generation of false or misleading information are overwhelmingly presented as caused by AI systems (90%), and risks related to AI systems pursuing their own goals in conflict with human goals or values are predominantly seen as AI-caused (72%). A third category, 'Other/Ambiguous', accommodates risks where the primary causal agent is not clearly delineated as either human or AI, or where the attribution is contested or unclear. Overall, the repository's database reveals a nearly equal attribution across the two primary categories, with 41% of risks presented as AI-caused and 39% as human-caused.

The second causal factor, "Intent", distinguishes whether a risk is an expected (intentional) or an unexpected (unintentional) outcome resulting from the pursuit of a goal, whether by a human or the AI system itself. "Intentional risks" occur as a consequence of a deliberate choice or programmatic design, such as an AI system explicitly programmed to exhibit deceptive behaviour or a human actor intentionally misusing an AI tool for harmful ends. Risks falling under the broad domain of "Malicious actors & misuse" are, perhaps unsurprisingly, overwhelmingly classified as intentional. For example, risks related to cyberattacks, weapon development or use, and causing mass harm are attributed to intentional actions in 85% of cases. In contrast, "Unintentional risks" are those that arise as unintended consequences, such

[100] Peter Slattery and others, 'The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence' (arXiv, 10 April 2025) <https://doi.org/10.48550/arXiv.2408.12622> accessed 14 July 2025.

as an AI system inadvertently developing discriminatory biases due to subtle patterns present in incomplete or unrepresentative training data. Empirical analysis within the repository suggests that risks like unfair discrimination and misrepresentation and unequal performance across different demographic groups are largely classified as unintentional outcomes (80% and 88% respectively). Similar to the Entity factor, an "Other/Ambiguous" category is used when intent is not clearly specified or where a risk could plausibly arise from both intentional and unintentional actions. Notably, a significant proportion of risks within the repository's data (31%) were presented with ambiguous or unspecified intent, highlighting a limitation in the clarity of existing risk descriptions.

The third causal factor, "Timing", categorises the risk based on the stage in the AI lifecycle at which it is presented as occurring. "Pre-deployment risks" are those that arise before the AI system is fully developed, validated, and placed into operational use, such as vulnerabilities introduced during the coding phase or biases embedded during data collection and model training. The repository's data indicates that only a relatively small proportion (13%) of the risks identified were presented as occurring specifically pre-deployment. The vast majority, classified as "Post-deployment risks", arise after the AI system has been deployed and is in active use, such as the misuse of the system for harmful purposes, emergent unintended behaviours in real-world conditions, or security breaches exploiting vulnerabilities in the deployed system. A significant majority of risks (62%) are presented as occurring post-deployment. For instance, risks related to fraud, scams, and targeted manipulation are predominantly categorised as post-deployment (93%).

Complementing the "Causal Taxonomy", the "Domain Taxonomy of AI Risks" categorises risks into seven major domains, further subdivided into 24 specific subdomains. This taxonomy focuses on describing the specific hazards and harms associated with AI, providing a structured overview of the broad spectrum of potential impacts. Here is where the multidimensionality of AI risks becomes evident. The seven domains are: (1) "Discrimination & toxicity", which encompasses risks such as unfair discrimination and misrepresentation, exposure to toxic or harmful content, and unequal performance across different user groups; (2) "Privacy & security", covering the compromise of sensitive information through obtaining, leaking, or inferring data and vulnerabilities and attacks targeting AI systems themselves; (3) "Misinformation", pertaining to the generation of false or misleading information and the broader pollution of the information ecosystem leading to a loss of consensus reality; (4) "Malicious actors & misuse", which explicitly focuses on the intentional use of AI for harmful ends, including disinformation, surveillance, and influence at scale, cyberattacks, weapon development or use, and mass harm, and fraud, scams, and targeted manipulation; (5) "Human-computer interaction", addressing issues arising from human engagement and reliance on AI systems, such as overreliance and unsafe use and the potential loss of human agency and autonomy; (6) "Socioeconomic & environmental harms", encompassing broader societal and ecological impacts including power centralisation and unfair distribution of benefits, increased inequality and decline in employment quality, economic and cultural devaluation of human effort, risks stemming from competitive dynamics driving rapid or unsafe deployment, governance failure due to inadequate regulatory frameworks, and the environmental harm associated with the energy consumption and carbon footprint of AI; and finally, (7) "AI system safety", failures & limitations, which addresses technical safety concerns and inherent limitations of AI systems, such as AI pursuing its own goals in conflict with human values, AI possessing dangerous emergent capabilities, lack of capability or robustness under varying conditions, lack of transparency or interpretability hindering understanding and accountability, ethical considerations regarding AI welfare and potential rights, and risks arising from interactions between multiple AI agents.

The analytical strength of the AI Risk Repository is significantly enhanced by its capacity to combine and cross-reference these two taxonomies, thereby revealing nuanced patterns and the differential interplay of causal factors across various risk domains and subdomains. For instance, risks classified under the domain of "Malicious actors & misuse" are consistently attributed to "Human entities" acting with "Intent" and typically occur "Post-deployment", reflecting the deliberate application of AI tools for harmful human agendas. In stark contrast, risks such as unfair discrimination and misrepresentation, unequal performance across groups, and the generation of false or misleading information are generally presented in the literature as unintentionally caused by the AI system itself (even if often rooted in human choices earlier in the lifecycle, the direct cause of the harmful output is the system's behaviour). However, other risk domains, including compromise of privacy, overreliance and unsafe use, loss of human agency and autonomy, and governance failure, exhibit less consistency in their causal attribution. These risks are often presented as being caused by both human and AI entities, and resulting from both intentional and unintentional actions, highlighting the complex, distributed, and frequently contested nature of these specific risk areas within existing research and policy discussions.

The synthesis conducted by the repository underscores a fundamental challenge permeating both research and policy discourse on AI: the absence of a coherent, shared conceptualization of "AI risks." Prior classifications exhibit significant uncoordinated and inconsistent approaches, where identical terminology frequently denotes disparate concepts. This pervasive lack of harmonised vocabulary substantially impedes comprehensive dialogue, collaborative research endeavours, and the formulation of effective, holistic policy responses to the multifaceted challenges posed by AI systems. Quantitative analysis starkly illustrates this fragmentation: none of the 65 documents reviewed encompassed all 24 distinct risk subdomains identified within the repository framework. Coverage averaged a mere 8 subdomains per document, ranging from a minimal focus on only 1 to a maximum of 19. This distribution reveals a pronounced tendency towards narrow analytical scopes within existing literature.

Furthermore, attention is demonstrably unevenly distributed across risk domains. Subdomains such as AI system safety, failures, and limitations (discussed in 75% of documents) and socioeconomic and environmental harms (76%) dominate the discourse. Conversely, critical areas remain profoundly underexplored. AI welfare and rights risks constitute less than 1% of all identified risks and are mentioned in only 3% of documents, while multi-agent risks account for just 3% of risks and feature in only 5% of documents. Even domains of significant societal relevance, such as misinformation (covered in 46% of papers) and human-computer interaction risks (49%), receive comparatively less scholarly and policy attention than their potential impact warrants.

This taxonomy renders evident that the term "AI risks" does not denote a singular, well-defined phenomenon. Rather, it represents a complex conglomerate of disparate concerns, differing fundamentally across multiple dimensions, including the entities affected, underlying causes, application domains, and temporal horizons. While the concept of "AI risk" provides a unifying framework to think about the implications of AI, it masks profoundly diverse underlying realities. The relationship between various AI risks arguably resembles a Wittgensteinian "family resemblance," where similarities are strongest between adjacent risk types and diminish to near non-existence between conceptually distant ones.

The fragmented nature of risks associated with artificial intelligence, as revealed by the demonstrable lack of consistent classification schemes, uneven research focus, and fundamental heterogeneity across multiple dimensions, presents not merely a practical complication but a profound ontological challenge to the very project of AI regulation. This challenge manifests in several critical, interrelated ways that strike at the core of regulatory efficacy. A primary manifestation is the intractable boundary problem

inherent in defining the regulatory subject. Effective regulation presupposes a reasonably definable subject matter, yet the repository's findings demonstrate that "AI risk" lacks ontological unity. It serves as a capacious label for phenomena as disparate as immediate physical system failures, long-term socioeconomic disruptions, insidious psychological manipulation, and speculative existential threats from hypothetical multi-agent systems. This heterogeneity creates an almost insurmountable boundary problem for regulators. Consequently, legislative attempts to delineate categories such as "high-risk AI systems," a central feature of the AI Act, become exercises in drawing arbitrary demarcations through a complex spectrum of interconnected yet fundamentally dissimilar hazards. The absence of shared conceptual foundations means that regulatory frameworks risk either substantial over-inclusion, by capturing low-impact activities under burdensome regimes designed for severe and unrelated risks, or dangerous under-inclusion, by failing to cover novel or poorly understood risk vectors that do not conform to established, narrow categories. The slipperiness of the core concept means the regulatory target constantly eludes precise and stable definition.

This definitional impasse directly gives rise to a second, equally critical challenge: the problem of regulatory targeting and the consequent risk of "one-size-fits-none" interventions. Effective regulation demands instruments that are meticulously tailored to the specific nature of the harm they seek to prevent or mitigate. The profound distinctions between a safety risk, such as an autonomous vehicle sensor failure precipitating a collision, and a fundamental rights risk, like algorithmic profiling leading to discriminatory denial of housing or employment, are not merely of degree but of kind. Regulatory tools effective for one are often irrelevant or even counterproductive for another. Mandating rigorous, standardised pre-deployment testing for physical safety is a conceptually coherent and viable approach for the autonomous vehicle. However, applying an analogous rigid certification process to mitigate subtle, context-dependent biases that infringe upon fundamental rights is an endeavour fraught with conceptual and practical difficulty. Such harms often emerge dynamically post-deployment and are deeply entangled with the complex social structures within which the systems operate. This fragmentation thus forces regulators toward blunt, overly generalized instruments or necessitates a convoluted regulatory patchwork that struggles to address each unique risk profile adequately, inevitably creating significant gaps and systemic inefficiencies.

The corollary of this "one-size-fits-none" dilemma is that any single, comprehensive regulatory framework attempting to govern such diverse phenomena will inevitably be beset by internal tensions and logical contradictions. This represents a coherence problem, where requirements designed to mitigate one category of risk can inadvertently exacerbate another. For instance, mandating high levels of system transparency and explainability to address safety concerns or ensure accountability might directly conflict with legal imperatives to protect proprietary trade secrets or the personal data privacy of individuals, which itself constitutes a fundamental rights risk. Furthermore, a regulatory focus on highly visible, immediate risks such as system failures may divert attention, resources, and institutional capability away from addressing less tangible but potentially more severe long-term risks, including the systemic erosion of collective human agency or the degradation of information ecosystems through pervasive, algorithmically amplified misinformation. The underlying fragmentation of the risk landscape ensures that the regulatory framework itself will lack internal coherence, as the distinct logics, values, and priorities essential for governing different risk subdomains inevitably pull in competing and often opposing directions.

This inherent variability and contextual dependence demonstrates how the very nature of AI risks presents a non-trivial challenge for AI regulation. As will be explored further, this fragmentation constitutes an equally significant challenge for AI technical standards. Standards bodies cannot

realistically provide universal, substantive solutions for every conceivable AI risk manifestation across its diverse dimensions. Their primary function is necessarily constrained to establishing formal, process-oriented frameworks for risk management – delineating methodologies for identification, assessment, mitigation, and documentation. While establishing robust processes is undeniably important, it constitutes significantly less than what is often implicitly expected from standards: namely, the provision of clear, universally applicable technical specifications guaranteeing safety or rights compliance across the entire, heterogeneous landscape of AI applications and associated risks. Standards can provide essential scaffolding for risk governance, but they cannot resolve the underlying ontological complexity or substitute for the nuanced, context-specific judgements required in both development and regulation. The expectation that standards alone can definitively "solve" the AI risk problem is therefore misplaced, reflecting a misunderstanding of both the fragmented nature of the risks and the inherent limitations of standardisation in addressing complex socio-technical challenges. Process frameworks offer necessary structure but not substantive certainty across the diverse spectrum of AI risk, a spectrum whose full contours and interconnections remain partially obscured by the very inconsistencies in discourse and focus identified by the repository synthesis.

4. The tool kit for the regulation of AI

Having analysed the compelling motivations underpinning AI regulation, alongside the significant and multifaceted challenges that complicate such endeavours, we shift focus towards examining the repertoire of regulatory tools available to address these issues.

This exploration of available tools is structured by drawing upon Freiberg's notion of a "Regulatory tool kit"[101]. Freiberg's work posits that regulatory governance is not limited to a single mode of intervention but is effectively discharged through a diverse array of mechanisms and instruments, each possessing distinct characteristics, strengths, and limitations. While Freiberg's comprehensive framework encompasses a wide spectrum of potential tools, our focus here will be directed towards a select subset of these instruments, i.e., those that are particularly pertinent to, or are currently being deployed for, the regulation of AI.

4.1. Overview of the regulatory tool kit

The concept of regulatory tool-kit offers a comprehensive lens through which to understand the diverse array of methods and resources available to governments and other actors seeking to influence and shape the behaviour of individuals and entities for various public purposes. This perspective transcends a narrow focus on purely legislative or rule-making functions, encompassing a wider spectrum of influences that operate concurrently and interactively to produce regulatory outcomes.

At its core, the tool-kit approach rests upon an expansive definition of regulation itself. In line with Black's definition adopted throughout this dissertation, regulation is here understood not merely as a set of state-imposed commands or prohibitions but as sustained and deliberate attempts by any actor to modify the conduct of others with the aim of addressing collective problems or achieving predefined societal goals. This influence is typically mediated through a combination of formal rules, informal norms,

[101] Arie Freiberg, *The Tools of Regulation* (Federation Press 2010) <https://research.monash.edu/en/publications/the-tools-of-regulation> accessed 10 April 2024.

coupled with mechanisms for their implementation and enforcement. Crucially, within this broad conceptualisation, these mechanisms need not be exclusively legal; they can derive from technical standards, economic incentives, social norms, or other forms of structured influence.

A key feature distinguishing this framework is its explicit rejection of traditional, often rigid, dichotomies commonly employed in regulatory analysis. It consciously eschews sharp analytical distinctions between "regulation" and supposed "alternatives to regulation," or between characterisations such as "heavy-handed" versus "light-touch," or "soft" versus "hard" regulation. Instead, the emphasis shifts to analysing the differential regulatory burdens imposed by the various tools and methods within the kit and assessing their relative efficiency and effectiveness in achieving specific policy objectives, recognising that these characteristics may vary significantly depending on the specific context, the nature of the problem being addressed, and the characteristics of the regulatees[102].

Within this conceptualisation, the state's role is not viewed as binary, either actively regulating or entirely absent, but rather as occupying a position along a dynamic continuum of engagement. Its relationship with the private sector and other non-state actors is highly variable and context-dependent, encompassing a broad range of possibilities regarding the locus of rule-making powers, the delegation of norm creation, the responsibility for investigation, the process of adjudication, the application of sanctions, the requirements for accountability, and the precise role afforded to judicial oversight. The state might, for example, delegate significant standard-setting power to industry while retaining overarching oversight and enforcement capabilities, or conversely, maintain tight central control over rule design while outsourcing monitoring or advisory functions to non-state experts. This nuanced view captures the fluidity and adaptability of modern governance structures.

Furthermore, the tool-kit perspective often incorporates the profound insight that human behaviour is, in a fundamental and pervasive sense, perpetually "regulated," irrespective of overt state intervention. Whether intentionally designed through formal rules or emerging as a complex by-product of routine human interactions, behaviour is continuously shaped by a multitude of influences operating within diverse social settings. These include, but are not limited to, the internalised norms and explicit rules governing conduct within families, the disciplinary structures and expectations of educational institutions, the behavioural codes enforced within religious communities, the subtle or overt pressures of peer groups, and the formal or informal strictures of the workplace. State regulation thus operates within and upon an already-regulated social terrain, interacting with, reinforcing, or occasionally conflicting with these non-state regulatory forces.

Consequently, changes observed in the nature or locus of governance are more accurately described not as instances of "de-regulation," which implies a simple withdrawal of all regulatory influence, but rather as complex processes of "regulatory reconfiguration." This term more precisely captures the ongoing, dynamic reorganisation of the various forms of regulation – encompassing state-led, private, and social modalities – and the shifting balances of responsibility and influence between state and non-state actors. It highlights a discernible movement away from an exclusive or primary reliance on purely "command and control" mechanisms towards approaches that actively incorporate and leverage the distinct capacities, specialised knowledge, and influential positions of the private sector, civil society organisations, and other non-state actors in the intricate and challenging task of governing complex phenomena. This dynamic perspective provides a valuable analytical lens through which to understand contemporary shifts in the architecture of governance, particularly relevant in fast-moving technological

[102] Arie Freiberg, 'Re-Stocking the Regulatory Tool-Kit' <http://regulation.huji.ac.il/dp.php>.

fields where state capacity may be limited and non-state actors possess critical expertise and proximity to innovation.

Freiberg's conceptualisation delineates six broad categories of power or tools through which behavioural change., i.e., the objective of regulation, can be achieved.

The first category encompasses "Economic Regulation". This modality involves the manipulation of the production, allocation, or use of material resources such as money or property. It manifests in various forms, including the imposition of taxes, charges, or levies, the provision of bounties or subsidies, and the establishment of sophisticated tradable permit schemes. These tools are particularly well-suited for influencing activities that are demonstrably "price-sensitive" or otherwise responsive to financial incentives and disincentives. Within this category, governments can engage in "Making Markets", creating entirely new marketplaces as regulatory instruments – examples include auctions for spectrum licenses or cap-and-trade schemes for pollution allowances. Alternatively, they can focus on "Influencing Markets", altering the dynamics of existing markets by modifying price signals, adjusting supply or demand conditions, regulating natural or artificial monopolies, or actively removing barriers to entry or exit.

Distinct from purely economic manipulation is "Transactional Regulation". This occurs when the state leverages its substantial economic resources through direct commercial transactions with regulatees. Often referred to as "regulation by contract", this category includes the government's extensive use of procurement contracts and the distribution of grants. These transactions, while ostensibly serving primary purposes such as acquiring goods or funding research, can be strategically designed or conditioned to pursue extraneous regulatory objectives, such as mandating adherence to specific labour standards, environmental controls, or, increasingly, ethical guidelines related to technology development. By making compliance a prerequisite for accessing state resources or contracts, the government can exert considerable influence on behaviour outside of traditional legislative mandates.

A particularly relevant category, and one which will form a central focus of the subsequent analysis, is "Regulation as authorisation". This relies fundamentally on the state's unique sovereign power to authorise, permit, allow, recognise, or legitimate specific activities, statuses, or premises. Mechanisms within this category, such as licensing, permits, registration requirements, certification processes, and accreditation schemes, serve as crucial "tokens of trust". They function to address pervasive information asymmetries within the market by signaling that an individual, entity, product, or service meets certain predefined standards of competence, quality, or safety. By granting or withholding authorization, the state filters market access and confers legitimacy, thereby structuring behaviour and mitigating potential harms. This modality is critical in fields where ensuring minimum standards of safety and reliability is paramount, acting as a gatekeeping function predicated on verified compliance with specified criteria.

A different modality of control is found in "Structural Regulation". These tools operate by altering the physical, environmental, process, or technological environment in a manner that intrinsically structures or constrains behaviour, effectively removing or severely limiting alternative choices such that regulatees are left with no practical option but to act in accordance with the desired regulatory pattern. This can involve "Physical Design", such as the architectural layout of spaces to limit movement (e.g., correctional facilities), implementing target hardening measures like locks and alarms to deter unwanted actions, or engineering physical environments to reduce risks or facilitate desired behaviours (e.g., designing road networks with safety features). It also includes "Environmental Design", consciously altering the built environment or surrounding context to influence social norms and behaviours (e.g., urban planning strategies aimed at crime prevention or promoting public health). Process Design systematically structures sequences of activities or tasks to embed compliance directly into workflows

(e.g., the design of pay-as-you-earn tax systems or the implementation of mandated quality management systems like ISO standards). Critically in the context of technology governance, this category includes regulation embedded directly into technology itself – designing systems that inherently prevent certain behaviours, such as internet filtering technologies engineered to block access to specified harmful content, age verification protocols integrated into platforms distributing restricted material, or the encoding of usage restrictions directly into digital media programs.

Information serves as its own powerful regulatory tool in "Informational Regulation". This approach leverages the strategic use of information to correct existing asymmetries, persuade actors towards desired behaviours, change attitudes, develop necessary capabilities, and form or modify prevailing norms. Key instruments include "Disclosure requirements", mandating that regulatees provide specific information (e.g., prospectus requirements for financial offerings, ingredient labelling on food products, health warnings on hazardous goods) to empower others to make informed choices. "Performance Indicators" involve making public information about the performance of regulatees relative to certain metrics (e.g., publishing "report cards" for healthcare providers or educational institutions) to incentivise improvement and inform consumer choice. Furthermore, this category includes efforts aimed at "Capability, Advice, and Attitude Change" through public information campaigns, provision of expert advice, educational programs, and training initiatives designed to foster desired behaviours and enhance understanding of regulatory requirements and underlying policy goals.

Finally, "Legal Regulation" encompasses the system of rules formally promulgated and backed by the coercive power of the state through sanctions. Law, in its various forms, remains a major and often foundational mode of regulation. This category includes various types of legislation, ranging from primary legislation (Acts of Parliament or equivalent) and delegated legislation (rules created under powers granted by primary law) to "soft law" or quasi-legislative instruments. Soft law instruments, such as regulatory guides, voluntary codes of conduct, technical standards, and guidelines, influence behaviour through their persuasive authority or by clarifying expectations, even without being legally binding per se. However, soft law can gain significant regulatory force through legal sanctions or by being referenced within legally binding instruments – for instance, by being incorporated into primary legislation, mandated as conditions for licenses or permits, or included as terms within binding contracts. Ethics and values, when formally recognised and enforced through legal frameworks, also acquire regulatory force within this category. Legal sanctions represent the forms of punishment or coercive measures imposed following a breach of legal rules, designed primarily to enforce compliance and deter future violations. These range across criminal sanctions (imprisonment, fines, probation, mandated compliance programs, compensation or restoration orders), civil penalties (monetary sanctions imposed in civil proceedings without criminal conviction), and administrative sanctions (warnings, banning orders, injunctions, mandatory disclosure orders, improvement notices, seizure of assets, infringement notices, and legally enforceable undertakings).

It is crucial to append several important caveats and nuances to this classification. Firstly, the regulatory tool-kit is inherently non-hierarchical; no single tool or category is universally superior, and different tools often operate simultaneously and interactively to achieve regulatory outcomes. Secondly, enforcement, while vital, is merely one element within this broader array of tools, operating alongside mechanisms aimed at incentivising, informing, or structuring behaviour. Thirdly, the categories presented are analytical constructs rather than rigid, mutually exclusive bins; real-world regulatory interventions often combine elements from multiple categories. Finally, law itself plays a pervasive and foundational role across almost all categories, providing the ultimate authority for creating rules, establishing regulatory agencies, authorising the use of other tools (like taxes or licenses), and providing the backing of ultimate

sanctions. Notably, concepts such as self-regulation, co-regulation, and meta-regulation are best understood not as pure forms of regulatory power in themselves, but rather as mechanisms or techniques through which the underlying forms of power (economic, authorizational, legal, etc.) are deployed and operationalised. A working knowledge of this wide and varied repertoire of instruments equips regulators with an array of potential interventions, allowing for a more nuanced and adaptive approach to achieving desired policy outcomes in complex and dynamic fields, while acknowledging that there is no single ideal regulatory configuration universally applicable across all contexts.

The categories delineated above do not map directly or orthogonally onto the more commonly employed tripartite categorisation distinguishing between state regulation, self-regulation, and co-regulation[103]. Rather than representing alternative sets of instruments, the concepts of regulation, self-regulation, and co-regulation describe distinct mechanisms or approaches to governing behaviour, whereas the "regulatory toolkit" provides the comprehensive inventory of instruments or tools that can be deployed through any of these mechanisms to achieve desired regulatory outcomes. They are not mutually exclusive alternatives, but rather represent different modalities through which the diverse tools available in the regulatory toolkit are utilised and operationalised.

Within this triad, "Regulation" (often termed "State Regulation") refers to a process inherently linked to the authority of the sovereign State, typically mediated through a dedicated, and often statutorily independent, regulatory body empowered to make decisions. Classic examples include state agencies responsible for issuing licenses for radio and television broadcasting and supervising the activities of the broadcast industry. This traditional paradigm of state regulation is fundamentally characterised by a relationship of "control by a superior" entity and embodies a primarily "directive function", where rules and requirements flow from the authoritative state entity towards the regulated parties. However, the limitations of conventional state law-making processes, including the often slow and cumbersome procedures for enacting legislation or the reactive nature of judicial adjudication regarding technological content or practices, have frequently proven inadequate or even fruitless when confronted with the dynamic and rapidly evolving nature of emerging technologies such as AI.

In contrast, "Self-Regulation" involves actors within a particular sector or domain independently identifying problems and devising solutions among themselves. This is often motivated, at least in part, by a desire to preempt or minimise direct intervention by the state in their affairs. Self-regulation gained prominence with the emergence of industry associations establishing their own codes of conduct to govern the behaviour of their members, such as in the context of standard setting analysed in Chapter 2. This model typically involves representatives from the industry and relevant professional bodies making judgments on complex matters of practice, often through peer review mechanism). This type of self-regulation frequently operates, however, under the shadow of the State, implying that while the state is not directly involved in the day-to-day norm creation or enforcement, all participating parties are acutely aware that the state retains the potential to intervene with binding measures should consensual solutions fail to materialise, or if significant public interests are perceived to be seriously threatened. A newer manifestation of industry self-regulation is observed in the digital realm, based on codes of practice voluntarily adopted by Internet companies to regulate issues such as online privacy, public decency, and the protection of minors. This approach represents a deliberate effort to move away from reliance on traditional state bureaucracy, which was often perceived as too slow, inflexible, and lacking the specific technical understanding necessary to effectively govern the rapidly evolving and intricate specifics of Internet communication.

[103] Hans J Kleinsteuber, 'Self-Regulation, Co-Regulation, State Regulation' <https://cdn.osce.org/sites/default/files/f/documents/7/d/36117.pdf>.

Finally, "Co-regulation" signifies a process where the State and private regulators engage in active cooperation, often within joint or formally recognised institutions. This mechanism is sometimes referred to as "regulated self-regulation," particularly when the framework for self-regulatory activity is structured and sanctioned by the State, even if the State is not directly involved in the operational implementation of the self-regulatory norms[104].

Crucially, self-regulation, co-regulation, and state regulation are not, as previously stated, the tools themselves, but rather the mechanisms, techniques, or governance structures through which these tools are deployed and operationalised. To illustrate, under a model of State regulation, the primary tools employed might be formal legal rules (embodying a command-and-control approach) backed by the coercive power of criminal or administrative sanctions, implemented and enforced through dedicated licensing bodies or state agencies. In contrast, Self-regulation might involve industry associations utilising tools such as codes of conduct (a form of quasi-legislation or soft law) to set norms, enforcing these norms through informational sanctions like public censure, or by leveraging authorisational tools such as exclusion from the association or the revocation of industry-specific certifications. Co-regulation involves a collaborative approach where, as in the case of the New Legislative Framework analysed in Chapter 3, a legislative act might utilise a legal regulation tool to set broad objectives (e.g., essential safety requirements for products), but then delegates the development of specific technical standards (a form of structural/process design or legal/soft law tool) to industry bodies or standardisation organisations to meet those objectives, often with ongoing state oversight or the implicit threat of state intervention if the standards prove inadequate.

In essence, these distinct forms of regulation (state, self, and co) primarily describe who is exercising the regulatory function (the state, an industry or group of non-state actors, or a combination thereof) and the nature of the control being exerted (direct and authoritative, autonomous and peer-based, or shared and collaborative). The regulatory toolkit, conversely, provides the catalogue of specific methods, resources, and instruments (economic incentives, authorisations, structural constraints, information dissemination, legal rules) that any of these actors or mechanisms can potentially utilise to influence behaviour. They are, therefore, interconnected elements within a broader and more complex regulatory landscape, with self-regulation and co-regulation functioning as specific techniques or operational mechanisms that draw upon the diverse array of tools available in the regulatory toolkit, frequently operating effectively only due to the latent or explicit influence exerted by the state, functioning as the omnipresent shadow of the State.

4.2. Co-regulation tools for AI: Codes of practice, regulatory sandboxes and technical standards

Given the distinct characteristics of AI systems and the particular challenges they pose for effective governance, the argument advanced here is that a co-regulatory approach, one that strategically combines the capacities and perspectives of state regulators, industry actors, civil society organisations, and other interested parties, represents one of the most promising avenues for navigating the complex landscape of AI regulation.

This conclusion is reached, in large part, through a process of elimination, drawing heavily upon the insightful arguments articulated by Michèle Finck regarding the inherent limitations of both traditional state regulation and pure self-regulation in the context of digital platforms[105]. By examining Finck's

[104] ibid.

[105] Michèle Finck, 'Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy' (Social Science Research Network, 20 June 2017) <https://doi.org/10.2139/ssrn.2990043> accessed 30 June 2025.

critique of these two poles of the regulatory spectrum, it becomes evident that the same fundamental reasons that render them inadequate for governing the platform economy apply with equal, and arguably greater, force to the challenge of regulating AI.

Turning first to the limitations of traditional State Regulation, particularly its manifestation as "command-and-control" top-down legislation, Finck persuasively argues that this approach is fundamentally ill-suited for governing dynamic digital ecosystems such as the platform economy, and these reasons resonate profoundly when considering AI. A primary impediment lies in the pervasive information asymmetry and the resultant lack of requisite expertise within governmental and regulatory bodies. Regulators and lawmakers simply do not possess sufficient, granular, or up-to-date information about the intricate internal workings of advanced technologies like AI. The algorithms, training data, and operational dynamics of cutting-edge AI systems frequently function as proprietary "black boxes," opaque even to external technical scrutiny. Attempting to craft and enforce detailed legislation without this foundational understanding is fraught with significant risks, potentially leading to unintended consequences or miscalibrated interventions. This epistemic deficit is only exacerbated by the accelerating pace of AI development, which quickly renders any static understanding obsolete.

Furthermore, traditional state-led regulation, if ill-informed or overly rigid, carries a substantial risk of stifling the very innovation it might seek to govern or protect the public from, as already mentioned in this Chapter. Rules devised by state bodies that lack deep technical insight can easily mandate outdated technical approaches, fail to anticipate beneficial future developments, or impose compliance burdens that are disproportionate to the actual risks, thereby harming economic vitality and innovation without effectively achieving the desired public good. This inherent inflexibility is a poor match for the characteristics of AI, a field defined by "acute and novel issues that are subject to rapid change". Static legislative instruments struggle to adapt dynamically to continuous technological evolution and the emergence of unforeseen capabilities or risks.

The "one-size-fits-all" problem, particularly relevant for supranational legislation like that emanating from the European Union, also poses significant difficulties when applied to technologies with diverse manifestations and impacts. While Finck primarily illustrates this using the example of platforms whose external effects (like Airbnb's impact on housing markets) vary dramatically depending on the local context, the analogy holds for AI. Different AI systems, deployed in varied contexts for disparate purposes, present a highly heterogeneous risk landscape. A single set of high-level, prescriptive rules is likely to be both under-inclusive and over-inclusive, failing to capture the specific nuances and risks of certain high-impact applications while potentially imposing unnecessary burdens on lower-risk systems or struggling to accommodate diverse operational environments.

Finally, traditional state regulation often faces considerable difficulty in enforcement when applied to complex digital phenomena. Even when seemingly clear rules are enacted, the task of monitoring compliance, detecting violations, and applying sanctions can be very burdensome to enforce for state authorities. Finck points to the challenges faced by German cities attempting to enforce bans on short-term rentals, contrasting this with how much simpler it would be for the platform itself to implement such restrictions via its underlying code. In the context of AI, this challenge is magnified by the opacity of systems, the global reach of many AI models and services, and the technical capacity required to audit complex AI behaviours or trace responsibility through intricate supply chains.

Conversely, turning to the limitations of a purely Self-Regulation model, often advocated by the tech industry on the basis of on their superior technical knowledge and perceived operational efficiency, Finck's analysis reveals critical and potentially fatal flaws. A fundamental weakness lies in the inherently biased incentives and the pervasive lack of adequate external accountability. When platforms (or, by

extension, AI developers) are left to govern themselves, they will inevitably prioritise their own commercial interests, such as profit maximisation, brand image management, and market dominance, over the broader public good or the legitimate interests of other stakeholders (users, affected individuals, civil society). In the absence of robust "external checks," there is insufficient structural pressure to ensure they act in the public interest.

Furthermore, a purely self-regulatory approach risks exacerbating existing concentrations of power and fostering anti-competitive behaviour. Leaving governance entirely to incumbent platforms (or leading AI firms) risks transforming them into self-regulating oligopolies. These powerful entities could leverage their rule-setting authority to create artificially high barriers to entry, designing technical or operational requirements that effectively prevent new competitors from entering the market or challenging their dominance. This risk is particularly pertinent in the highly concentrated AI market.

While industry claims of superior knowledge are partially valid, especially in highly specialised fields such as AI, self-regulation also suffers from incomplete knowledge, albeit of a different kind than the state's deficit. While platforms or AI developers possess vast amounts of operational data and technical know-how, they typically lack the holistic, societal perspective required for sound regulation. As Finck notes, Airbnb knows about rental patterns, but city authorities possess broader knowledge about housing shortages, urban planning objectives, and residents' overall needs. No single actor within the ecosystem possesses all the knowledge required to make comprehensive and equitable regulatory decisions.

Finally, a purely self-regulatory model leads inevitably to legal uncertainty. The absence of uniformly binding standards developed or endorsed through a public process result in a fragmented legal landscape, where applicable rules are determined on a case-by-case basis, often as a consequence of litigation. This fragmentation and reliance on unpredictable litigation are undesirable for all parties involved – platforms seeking clear operational guidelines, users desiring predictable rights and protections, and regulators aiming for systemic oversight.

Drawing upon Finck's compelling analysis concerning digital platforms, the argument holds that traditional state regulation, with its inherent rigidity, slow pace, and susceptibility to information asymmetry, is ill-equipped to effectively govern dynamic technological fields like AI. Simultaneously, pure self-regulation, while leveraging industry expertise, is fundamentally compromised by biased incentives, a lack of accountability, and the risk of entrenching market power, rendering it unreliable for safeguarding the public good. These limitations of the extremes, often amplified by the specific characteristics of AI discussed earlier, i.e., its rapid, unpredictable development, its opacity, and the high concentration of resources and knowledge within a few frontier labs, point strongly towards the necessity of a middle ground as the best, or the least worst, option for AI regulation.

The operationalisation of a co-regulatory framework for AI necessitates a specific and adaptable set of tools made available in the regulatory toolkit. These tools are the very mechanisms through which the collaborative process between public authorities, industry developers, and civil society can be structured and made effective. They function as the procedural architecture for shared governance, enabling the state to leverage the expertise and agility of private and third-sector actors while retaining its essential role as the ultimate guarantor of the public interest. In the rapidly evolving domain of AI, the instruments that have garnered the most significant attention and practical application are codes of practice, regulatory sandboxes, and technical standards. Each of these tools offers a distinct method for navigating the complex trade-offs between fostering innovation, managing risk, and ensuring accountability in a manner that is more flexible and informed than traditional legislation.

Within the broader taxonomy of regulatory strategies, these instruments are most aptly categorised under the paradigm of "regulation as authorisation". Regulation as authorisation is predicated upon the

unique power of the state to permit, allow, recognise, or legitimate specific activities, entities, or processes. Mechanisms such as formal licensing, permits, registration schemes, and accreditation function as official tokens of trust issued by a public authority. Their principal regulatory purpose is to mitigate the significant information asymmetries that characterise complex markets, especially the chasm of understanding between expert providers lay consumers or the public at large. By establishing and endorsing these mechanisms, the state guarantees a minimum threshold of competence, safety, or quality. Consequently, authorisational regimes are foundational to the management of novel risks and the cultivation of public confidence in a vast range of technological and professional activities.

This categorisation makes profound sense in the context of AI governance. The link is most explicit in the case of regulatory sandboxes, which are, by definition, a form of temporary and conditional authorisation[106]. A regulator grants a firm a formal permit to test a novel AI system in a controlled, live environment, often with waivers from certain existing rules, under strict supervision[107]. Successful navigation of the sandbox can then lead to a broader, more permanent authorisation for market deployment. For codes of practice and technical standards, the mechanism of authorisation is more indirect but equally potent. While these instruments are often developed by industry consortia or multi-stakeholder standards bodies, their regulatory force is activated when a public authority officially recognises or endorses them. By referencing a specific technical standard in procurement policies or by establishing a code of practice as a "safe harbour" for demonstrating compliance, the state effectively authorises that standard or code as a legitimate benchmark for responsible conduct. Adherence thus becomes a de facto condition for market access or a means of mitigating legal liability, functioning as precisely the kind of state-sanctioned "token of trust" that defines authorisational regimes. A more granular examination of each of these instruments is therefore warranted to fully appreciate their respective roles and limitations within the co-regulatory toolkit for artificial intelligence.

4.2.1. Regulatory sandboxes

Regulatory sandboxes represent an excellent example of "regulation as authorisation". This regulatory paradigm shifts the focus from purely prohibitory or prescriptive norms to frameworks that actively enable and structure innovation under controlled conditions. By establishing a predefined, supervised space for experimentation, sandboxes explicitly authorise activities that might otherwise face significant legal uncertainty.

To fully appreciate this, it is necessary to introduce the concept of regulatory sandboxes and contextualise their emergence, particularly in the field of AI. Fundamentally, regulatory sandboxes are delineated legal frameworks engineered to facilitate limited testing of innovative technologies under the aegis of regulatory supervision. They function as concrete structures providing a formal context for experimentation, allowing for the evaluation of innovative technologies, products, services, or methodologies within a real-world environment[108]. This testing is circumscribed by both temporal limitations and a restricted sectoral or geographical scope, all conducted under vigilant regulatory oversight to guarantee the presence of adequate safeguards.

---

[106] Alexey Yefremov, 'Regulatory Sandboxes and Experimental Legislation as the Main Instruments of Regulation in the Digital Transformation' in Daniel A Alexandrov and others (eds), *Digital Transformation and Global Society* (Springer International Publishing 2019) <https://doi.org/10.1007/978-3-030-37858-5_7>.

[107] Filippo Bagni and Seferi Fabio (eds), *White Paper on Regulatory Sandboxes for AI and Cybersecurity* (Cybersecurity National Lab 2025).

[108] Jon Truby and others, 'A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications' (2022) 13 European Journal of Risk Regulation 270 <https://doi.org/10.1017/err.2021.52>.

The genesis of the regulatory sandbox concept is firmly rooted in the financial services sector. The United Kingdom's Financial Conduct Authority (FCA) is widely recognised as the pioneering entity, having inaugurated and deployed the inaugural formal regulatory sandbox in 2015[109]. This initiative was specifically designed to aid in the management and oversight of products emerging from the rapidly evolving field of financial technologies (fintech). Since this initial deployment, the model has experienced rapid global diffusion, leading to the establishment of numerous fintech-focused sandboxes by national and subnational financial regulators worldwide. The European Council has also formally acknowledged the practical utility of regulatory sandboxes and associated experimentation clauses, noting their applicability extends well beyond FinTech, encompassing broader technological growth across sectors[110].

The operational architecture of regulatory sandboxes is fundamentally designed to simultaneously foster innovation and ensure requisite safety through a combination of enhanced legal certainty, targeted regulatory engagement, and built-in flexibility. They furnish a mechanism by which novel technologies can be meticulously observed and rigorously tested on a small scale prior to consideration for wider market deployment. Legislators have increasingly explored these legal instruments as a means to enable controlled real-world experimentation, underscoring the perceived importance of flexibility and a capacity for iterative experimentation in effectively navigating the complex regulatory challenges posed by technological innovation[111].

The common objectives and operational mechanisms underpinning regulatory sandboxes elucidate their function as instruments of regulation as authorisation. Primary among these is the goal of enabling innovation and facilitating market access. Sandboxes aim to streamline the development and deployment pathways for innovative systems, thereby lowering barriers to market entry for providers and consequently stimulating investment and growth in nascent technologies[112]. This mechanism directly contributes to accelerating market access for AI systems while simultaneously endeavouring to improve legal certainty for innovators. A key feature enabling this is the provision of legal certainty and guidance. Regulatory sandboxes are typically endowed with specific legal powers enabling the competent supervisory authorities to render legal guidance. These authorities can assess whether a given product or service aligns with existing legal requirements and furnish either individualised advice tailored to a specific project or general guidance on the pathways towards achieving compliance. This function is twofold: it seeks to enhance legal certainty for participating innovators and to augment the regulatory knowledge base of the authorities themselves, fostering a dynamic understanding of evolving technologies.

Adding another layer to the authorising function, supervisory authorities within sandboxes can issue 'no-enforcement letters'[113]. These constitute formal commitments by the regulator to refrain from initiating enforcement action in specific, defined instances. Such instruments can serve either to clarify compliance requirements in novel contexts or, in certain circumstances, act as a form of temporary legal exemption, mitigating the risk of inadvertently breaching legal provisions during the critical testing phase. However, the legality and appropriateness of no-enforcement letters are generally confined to narrowly

---

[109] Deirdre M Ahern, 'Regulators Nurturing FinTech Innovation: Global Evolution of the Regulatory Sandbox as Opportunity Based Regulation' (Social Science Research Network, 9 March 2020) <https://doi.org/10.2139/ssrn.3552015> accessed 23 December 2025.

[110] 'Regulatory Sandboxes and Experimentation Clauses as Tools for Better Regulation: Council Adopts Conclusions' (*Consilium*) <https://www.consilium.europa.eu/en/press/press-releases/2020/11/16/regulatory-sandboxes-and-experimentation-clauses-as-tools-for-better-regulation-council-adopts-conclusions/> accessed 23 December 2025.

[111] Yefremov (n 106).

[112] Thomas Buocz, Sebastian Pfotenhauer and Iris Eisenberger, 'Regulatory Sandboxes in the AI Act: Reconciling Innovation and Safety?' (2023) 15 Law, Innovation and Technology 357 <https://doi.org/10.1080/17579961.2023.2245678>.

[113] ibid.

defined circumstances, necessitating the implementation of robust safeguards for third parties who might be affected. Under the AI Act, for example, no administrative fines shall be imposed for infringements occurring within the sandbox, provided that participants strictly adhere to the agreed sandbox plan and the guidance provided by the authorities.

Beyond mere non-enforcement, sandboxes can incorporate provisions for direct legal exceptions or exemptions. This mechanism permits innovators to deviate from specific legal rules that might otherwise impede or render impossible the necessary testing procedures. This power is designed to inject a degree of regulatory flexibility, effectively exempting innovators from certain otherwise binding legal constraints on the testing process itself. To counterbalance the potential reduction in protection that such exemptions might entail, the supervising authority typically has the power to impose compensatory conditions. These might include requirements for intensified supervision, stricter reporting obligations, or explicit limitations on the scope and scale of the testing activities. This mechanism is frequently facilitated by "experimental clauses" embedded within broader legislation, which explicitly authorise supervisory authorities to permit deviations from statutory law for specified experimental purposes[114]. This illustrates a direct form of regulatory authorisation, allowing activities that would normally be proscribed under specific, controlled conditions.

Central to the legitimacy and safety of sandboxes, despite their authorising function, is robust regulatory supervision and the implementation of appropriate safeguards[115]. Innovations are not tested in a vacuum but under the direct oversight of the regulator, ensuring that risks are managed proactively. The AI Act specifically underscores that sandbox activities are to be conducted under the "direct supervision and guidance" of competent authorities, explicitly linking this supervision to the objective of ensuring compliance not only with the AI Act itself but also with other relevant Union and Member State legislation. Furthermore, a critical operational aspect is the facilitation of testing "where appropriate in real-world conditions". This moves beyond laboratory simulations, allowing for the gathering of more reliable data and a more accurate assessment of conformity in the environments where the AI system is intended to operate. This real-world testing, though authorised and supervised, provides invaluable empirical evidence for both the innovator and the regulator.

Importantly, the authorisation provided by sandboxes is inherently constrained. Participation is limited in duration, with the time allowed based on the complexity and scale of the project. Similarly, tests are typically conducted within a limited part of a sector or area, preventing uncontrolled deployment across an entire market. Liability remains a significant consideration; participants generally retain liability for any damage caused during their involvement in the regulatory sandbox under applicable Union and Member State legislation. This aspect is subject to ongoing debate, as the potential for significant liability, even during supervised testing, may serve as a limiting factor on the incentives for participation, potentially counteracting the authorising effect[116].

Beyond facilitating individual innovation, a crucial supplementary goal of regulatory sandboxes is mutual information sharing and learning. They are designed to enable regulators themselves to gain a deeper, more practical understanding of emerging products and services. Sandboxes provide a structured environment for regulators to observe how general legal provisions translate into practice, identify potential unintended consequences, and determine when amendments to existing regulations or the

[114] Marion Ho-Dac, Cécile Pellegrini and Bernard Long, 'The Academic Guide to AI Act Compliance' (Social Science Research Network, 30 September 2025) <https://doi.org/10.2139/ssrn.5775382> accessed 23 December 2025.

[115] Walter G Johnson, 'Caught in Quicksand? Compliance and Legitimacy Challenges in Using Regulatory Sandboxes to Manage Emerging Technologies' (2023) 17 Regulation & Governance 709 <https://doi.org/10.1111/rego.12487>.

[116] Giovanni Maria Riccio, 'Tort Liability and Regulatory Sandboxes' (Social Science Research Network, 4 February 2025) <https://papers.ssrn.com/abstract=5165196> accessed 23 December 2025.

creation of new ones might be necessary. The insights and empirical evidence gleaned from sandbox experimentation are explicitly anticipated to inform and improve the development of harmonised technical standards for AI systems, further integrating the experimental insights into the broader regulatory framework. They formalise and structure the dialogue and exchange of information between regulatory bodies and the entities they regulate, building capacity on both sides[117].

Ultimately, sandboxes serve as instruments specifically calibrated to manage the inherent uncertainty and complexity characteristic of emerging technologies like AI. By providing a controlled environment for exploration, they aid in the proactive identification of potential technical, ethical, or legal issues and facilitate the development of informed policy responses while simultaneously mitigating systemic risk to the public. Overall, regulatory sandboxes manifest as a significant evolution in regulatory reform, drawing upon traditions of 'flexible' and 'experimentalist' regulation. They are designed to adapt regulatory approaches to novel scenarios, actively involve third parties in the process of regulatory learning, and promote capacity building for all actors within the regulatory ecosystem.

4.2.2. Codes of Practice

Codes of Practice, emerging as a prominent co-regulatory tool, also manifest aspects of "regulation as authorisation". Their development is inherently collaborative, involving a spectrum of actors spanning regulatory bodies, industry stakeholders, and representatives of civil society. This multi-stakeholder genesis results in the formulation of agreed principles and standards. Crucially, these collaboratively developed frameworks subsequently receive formal endorsement from the state or another legitimate authority, such as the EU. This official imprimatur, while not necessarily conferring the same degree of strict legal enforceability as 'hard law', bestows upon the Code a significant normative weight and regulatory legitimacy. It effectively sanctions certain practices and methodologies as aligned with regulatory expectations, thereby authorising specific approaches to compliance and risk management within a sector. This mechanism thus operates by guiding behaviour and shaping market norms through a process of consensual standard-setting validated by formal regulatory acceptance, rather than relying exclusively on traditional command-and-control directives

A Code of Practice, in essence, constitutes a form of regulatory governance often categorised under the rubric of soft law. It comprises a compendium of agreed principles, guidelines, and standards intended to provide direction on how entities should conduct their operations, particularly in addressing potential market failures or mitigating environmental and societal risks. Unlike statutory law or binding regulations, a code of practice typically operates without a direct, formal enforcement mechanism for ensuring compliance, relying instead on voluntary adherence, collaborative efforts, and industry self-regulation. Such codes are strategically adopted with a view to achieving specific policy objectives, serving either as an interim measure or a more permanent alternative to traditional legislative interventions where flexibility or rapid adaptation is paramount.

In the context of European Union regulation, particularly concerning the governance of large digital platforms, a salient example is the EU Code of Practice on Disinformation. Inaugurated in late 2018, this initiative sought to foster voluntary collaboration among major social media and digital platform

[117] Alessio Tartaro and Enrico Panai, 'Leveraging Technical Standards within AI Regulatory Sandboxes: Challenges and Opportunities' (Social Science Research Network, 4 February 2025) <https://doi.org/10.2139/ssrn.5235132> accessed 21 August 2025.

operators, including entities such as Facebook, Google, Twitter, Microsoft, and Mozilla, for the express purpose of mitigating the proliferation of online disinformation. The Disinformation COP exemplifies a multistakeholder approach to what was intended as transparent self-regulation. Under its terms, signatory platforms committed to submitting both baseline and subsequent annual reports detailing their progress across predefined categories, encompassing areas such as the rigorous scrutiny of advertisement placements, political advertising transparency, ensuring the integrity of services, empowering consumers with information, and fostering greater access for the research community.

However, the practical implementation of the Disinformation COP has yielded results that might be characterised as mixed[118]. While it undoubtedly represented a logical and necessary initial step in addressing a complex societal challenge, it has not fully succeeded in cultivating robust trust across the requisite spectrum of stakeholders, including industry, governmental bodies, academia, and civil society organisations. A pivotal challenge identified lies in the fundamental asymmetry of priorities between the European Union, which foregrounds democratic rights, transparency, and public interest objectives, and the major digital platforms, whose focus tends to gravitate towards optimising customer experience, managing public image, and streamlining internal operational processes. Consequently, the degree of compliance demonstrated by companies often appears contingent upon the extent to which a particular regulatory theme aligns with or is prioritised within their internal corporate strategies and operational mandates. Evidence suggests that platforms generally exhibit a higher level of formal commitment – that is, pledging to implement stated principles and general actions – compared to symbolic commitment, which might entail agreeing in principle without committing to specific, measurable actions. Implementation tends to be more pronounced in areas deemed highly salient for the platforms themselves, such as ensuring the integrity of their services (e.g., through measures like closing fake accounts) and coordinating with third parties on content moderation or fact-checking efforts. Conversely, areas directly implicating fundamental rights or freedom of expression have frequently received minimal concrete action or demonstrated commitment from many platforms. This discrepancy underscores the observation that soft governance mechanisms can be effective where there is a clear convergence between regulator and industry priorities, but conversely, they may lead primarily to superficial changes, often termed window dressing, or cost-effective operational adjustments in domains that are less prioritised by industry, potentially leaving critical policy objectives unmet[119].

A Code of Practice is envisioned as the primary, and effectively default, mechanism by which providers of General Purpose AI (GPAI) models will demonstrate their compliance with the relevant provisions under the AI Act[120]. The fundamental objective of such a Code is to operationalise the legislative requirements, thereby providing the necessary technical and procedural means to achieve compliance with the GPAI obligations outlined in Article 53 and 55 of the AI Act. Once again, a co-

[118] Stephan Mündges and Kirsty Park, 'But Did They Really? Platforms' Compliance with the Code of Practice on Disinformation in Review' (2024) 13 Internet Policy Review 1 <https://doi.org/10.14763/2024.3.1786>.

[119] James Pamment, 'EU Code of Practice on Disinformation: Briefing Note for the New European Commission |' <https://carnegieendowment.org/research/2020/03/eu-code-of-practice-on-disinformation-briefing-note-for-the-new-european-commission?lang=en> accessed 23 December 2025; Gabriela Borz and others, 'The EU Soft Regulation of Digital Campaigning: Regulatory Effectiveness through Platform Compliance to the Code of Practice on Disinformation' (2024) 45 Policy Studies 709 <https://doi.org/10.1080/01442872.2024.2302448>.

[120] 'The General-Purpose AI Code of Practice | Shaping Europe's Digital Future' <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai> accessed 23 December 2025.

regulatory tool is selected as the most appropriate solution to showcase compliance with the requirements of the legislation[121].

Irrespective of their specific scale or associated risk profile, all providers of GPAI models are subject to a foundational set of following compliance pillars. First, providers are mandated to create and maintain comprehensive technical documentation. This requirement entails a thorough description of the model's architecture, its training methodologies, and its functional capabilities, all of which must be kept available for inspection upon request by the AI Office or relevant national competent authorities. Second, there is an obligation to ensure transparency for downstream providers. This necessitates the supply of sufficient information and documentation to third-party companies or developers who intend to integrate the model into their own AI systems. The rationale for this information flow is to empower downstream actors to fully comprehend the modelis capabilities and inherent limitations, thereby enabling them to fulfil their own distinct legal obligations.

Furthermore, the regulatory framework imposes strict adherence to intellectual property standards. Providers must implement a robust policy to ensure compliance with EU copyright law, a requirement that specifically involves identifying and respecting opt-outs exercised by content creators and rights holders. Complementing this protection of intellectual property is the fourth requirement: the publication of a training data summary. Providers are obligated to release a sufficiently detailed summary of the content utilized during the training of the model, ensuring a degree of public accountability regarding the inputs that shape the model's behaviour[122].

While these baseline obligations apply universally, the regulatory regime becomes significantly more rigorous for GPAI models classified as possessing "systemic risk." This classification is typically determined, not without due criticism[123], by high computational thresholds, such as models trained with a compute volume exceeding 10^{25} Floating Point Operations (FLOPS). For providers of these powerful systems, the Act imposes a series of cumulative duties designed to preemptively address large-scale safety concerns. Foremost among these is the requirement for adversarial testing; providers must conduct rigorous model evaluations, including "red teaming," to systematically probe the system for vulnerabilities and potential failure modes, such as those related to jailbreaking, which can be achieved even via crafted poetry-like prompts[124]. Concurrently, these providers are tasked with active risk mitigation, requiring them to assess and address systemic risks at the European Union level. This risk management strategy is reinforced by a mandate to ensure adequate levels of cybersecurity protection, safeguarding both the model and its underlying physical infrastructure against external threats. Finally, to ensure regulatory visibility into critical failures, providers must adhere to strict incident reporting protocols, promptly notifying the newly established AI Office of any serious incidents involving the model.

The Code of Practice functions as a voluntary regulatory instrument designed to facilitate adherence to the above requirements. Its primary utility lies in establishing a provisional yet clear framework for providers of GPAI models, particularly during the interim period prior to the full crystallisation of

---

121 Alexander Peukert, 'A New Kid on the EU's Co-Regulatory Block: Codes of Practice under the AI Act' (Social Science Research Network, 15 December 2025) <https://papers.ssrn.com/abstract=5923162> accessed 23 December 2025.

122 Paul Voigt and Nils Hullen, 'Which Requirements Apply to GPAI Models?' in Paul Voigt and Nils Hullen (eds), *The EU AI Act: Answers to Frequently Asked Questions* (Springer 2024) <https://doi.org/10.1007/978-3-662-70201-7_5> accessed 23 December 2025.

123 Stefan Larsson and Jockum Hildén, 'Flip or Flop? The Pacing Problems of European Systemic Risk GPAI Regulation' (Social Science Research Network, 15 August 2025) <https://doi.org/10.2139/ssrn.5527918> accessed 23 December 2025.

124 Piercosma Bisconti and others, 'Adversarial Poetry as a Universal Single-Turn Jailbreak Mechanism in Large Language Models' (arXiv, 16 January 2026) <https://doi.org/10.48550/arXiv.2511.15304> accessed 20 April 2026.

European harmonised standards. By offering a structured pathway for compliance, the Code serves as a central mechanism for ensuring that GPAI models are developed, deployed, and maintained with the requisite levels of safety and transparency mandated by the burgeoning regulatory landscape.

To operationalise these goals, the Code details specific rules derived from the AI Act, organizing its mandates into a tripartite structure. These distinct chapters, which were drafted by specialized working groups, address transparency, copyright policies, and systemic risk management respectively, which strongly rely on existing industry practices[125]. The first two components apply universally to all GPAI providers, mandating the disclosure of relevant information to both downstream providers and the AI Office to demonstrate compliance with Article 53 of the Act, alongside the establishment of robust policies to ensure alignment with EU copyright laws. Conversely, the third chapter remains exclusive to providers of models possessing systemic risks. This section focuses specifically on the assessment and mitigation of potential widespread disruptions, demonstrating compliance with Article 55.

The legitimacy of this framework is derived in part from its development trajectory, which was characterised by a rigorous, iterative, and inclusive process facilitated by the European AI Office. Commencing with a "Kick-off Plenary" in September 2024 and culminating in the Code's publication on July 10, 2025, the drafting involved nearly one thousand stakeholders. This broad coalition included GPAI providers, downstream developers, industry organisations, civil society actors, rightsholders, and independent experts, all operating under the guidance of independent Chairs and Vice-Chairs appointed by the AI Office. The GPAI CoP was eventually signed by a big number of model providers, including Amazon, Anthropic, Google, Microsoft, Mistral AI and other[126].

However, the efficacy and legal robustness of the resulting document have been the subject of significant debate, with scholars arguing that the final text contains significant omissions and inaccuracies [127]. These deficiencies may create regulatory loopholes or, more precariously, expose providers to compliance risks by suggesting that adherence to the Code equates to full legal satisfaction when it may fall short of statutory requirements.

A focal point of this critique pertains to transparency and the granular handling of data provenance. While Annex XI of the AI Act explicitly requires detailed accounts of data provenance, the Code has been criticised for reducing this obligation to a mere categorical selection, a "checkbox" approach where providers might simply indicate sources like "web crawling" or "synthetic data." This simplification ignores the computer science definition of provenance, which necessitates a record of data lineage and history, thereby rendering the provision insufficient for meeting the legal standard. This issue of opacity is compounded by the Code's allowance for non-comparable metrics regarding the quantity of training data. By permitting providers to choose their own units of measurement, such as tokens, words, or sentences, the Code makes it effectively impossible to benchmark or compare different models. Furthermore, the suggestion that complex data cleaning and filtering methodologies can be adequately explained in descriptions of merely 300 words is viewed by critics as wholly inadequate for models trained on massive, high-dimensional datasets.

The critique extends to the handling of bias detection and copyright compliance. Veale and Quintais identify a potential *contra legem* alteration in the text regarding bias: whereas the AI Act requires documentation of "all other measures" used to detect bias, the Code omits the word "all," granting

[125] Lily Stelling and others, 'Existing Industry Practice for the EU AI Act's General-Purpose AI Code of Practice Safety and Security Measures' (arXiv, 21 April 2025) <https://doi.org/10.48550/arXiv.2504.15181> accessed 30 June 2025.

[126] 'The General-Purpose AI Code of Practice | Shaping Europe's Digital Future' (n 120).

[127] Michael Veale and João Pedro Quintais, 'The Obligations of Providers of General-Purpose AI Models' (Social Science Research Network, 13 November 2025) <https://doi.org/10.2139/ssrn.5744602> accessed 23 December 2025.

providers the discretion to withhold information selectively[128]. Regarding copyright, the Code's limitation of compliance obligations to machine-readable opt-outs (such as robots.txt) is seen as problematic. Critics note that the Copyright Directive does not restrict rights reservations to machine-readable formats for all content types; thus, ignoring reservations expressed in natural language terms and conditions could constitute a violation of the law. Moreover, the removal of references to extraterritoriality in the final draft creates ambiguity regarding whether providers must adhere to EU copyright laws when data scraping occurs outside the jurisdiction, potentially incentivizing regulatory arbitrage. Finally, the critics highlight a misinterpretation of Article 78 regarding confidentiality, arguing that the Code improperly extends authority-specific confidentiality privileges to private providers, thereby obstructing the transparency necessary for the value chain, while simultaneously failing to define what constitutes a "material change" that would trigger mandatory documentation updates.

4.2.3. Technical standards

The final modality within the "regulation as authorisation" paradigm that warrants consideration in this section is that of technical standards, and most notably, harmonised standards. While the development of technical standards in general resides primarily within the purview of stakeholders operating outside the direct regulatory apparatus of the state or Union institutions, typically through recognised standardisation bodies, the system of harmonised standards, particularly as instantiated under the NLF, should qualify unequivocally as a potent form of co-regulation. This status derives from the peculiar institutional architecture and operational mechanisms of the NLF. Within this framework, EU institutions require the development of standards by the ESOs and these developed standards, in turn, acquire regulatory significance through a formal process of recognition and referencing, most notably via their publication in the Official Journal of the European Union, as we will explore in greater details in Chapter 3.

This formal endorsement effectively transforms voluntary technical specifications into de facto compliance pathways. Adherence to harmonised standards referenced under the NLF typically provides a presumption of conformity with the essential requirements laid down in the relevant Union harmonisation legislation. Thus, entities electing to follow these standards are, in essence, being authorised by the regulatory system to do so, with their compliance effort thereby validated by the formal regulatory endorsement of the standard itself. This intricate nexus between stakeholder-led technical development and formal regulatory recognition positions harmonised standards squarely within the ambit of 'regulation as authorisation', enabling compliance through formally accepted, though not directly state-mandated, technical means. The profound implications of this system, particularly concerning its role and inherent limitations in addressing the unique challenges posed by the regulation of Artificial Intelligence technologies, will be subject to a more extensive and granular analysis in subsequent chapters of this dissertation, delving into the specific mechanisms of standardisation applicable to AI and evaluating their efficacy and.

[128] ibid.

5. Conclusion of the Chapter

This chapter has served to lay essential groundwork for the remainder of the research, which will pivot towards a more specific and detailed examination of the role and inherent limitations of technical standards in the context of Artificial Intelligence regulation.

The chapter commenced by identifying the principal rationales advanced within scholarly discourse and policymaking spheres for intervening to regulate AI systems. As demonstrated, among the various justifications, the minimisation or mitigation of risk emerges as arguably the most recurrent and dominant motivation underpinning contemporary AI governance initiatives, particularly evident in prominent legislative efforts like the AI Act. Concurrently, however, we observed that this seemingly straightforward motivation is attended by significant ambiguity. The notion of "risk" in the AI context is itself multifaceted and contested, encompassing disparate phenomena ranging from immediate safety failures and discriminatory outcomes to long-term societal impacts and potential existential threats. The dominant focus on "risk minimisation" often sidesteps crucial questions about whose risks are being prioritised, how risks are to be identified and measured (especially given the opacity and emergent properties of some AI systems), and whether the goal is to eliminate risk entirely (an impossibility), reduce specific types of risk, or manage risk within acceptable societal trade-offs. This ambiguity in the foundational regulatory motivation inevitably translates into complexity and potential incoherence in the design and application of regulatory tools intended to achieve it.

At the current juncture, however, the political and policy commitment to regulating AI appears largely indisputable, regardless of the precise or fully resolved underlying motivations. Therefore, having established the 'why', the discussion transitioned to an analysis of the profound and intricate challenges that inevitably emerge when moving from the decision to regulate AI to the practical task of how to regulate effectively. These challenges were shown to derive both from more general difficulties inherent in the regulation of technology, such as the pervasive "pacing problem" and various forms of structural and informational asymmetries, difficulties that are significantly exacerbated in the particularly dynamic and complex field of AI, as well as from more specific problems directly attributable to the unique nature and multifaceted characteristics of AI risks themselves, as mapped, for instance, by detailed risk taxonomies. The confluence of rapid technological change, inherent uncertainties, informational imbalances between regulators and regulatees, and the diverse, often unpredictable, manifestations of AI risks presents a formidable obstacle course for traditional regulatory approaches.

Finally, in light of these identified motivations and challenges, the chapter explored the potential utility of various instruments available within the conceptual framework of a 'regulatory tool kit'. Drawing distinctions between different modes of governance – traditional state regulation, self-regulation, and co-regulation – the analysis suggested that approaches leaning towards co-regulation, particularly those rooted in the concept of 'regulation as authorisation', appear to be among the most promising for addressing the specific complexities of AI. This preference stems from the noted limitations of purely state-centric command-and-control mechanisms, which often struggle with information asymmetry and rapid adaptation, and the inherent drawbacks of unfettered self-regulation, which is prone to conflicts of interest and insufficient public accountability. Co-regulatory models, by contrast, offer the potential to leverage the technical expertise and proximity to innovation held by industry and non-state actors, while grounding these efforts within a framework structured and ultimately overseen by legitimate state authority, thereby offering a potential pathway to navigate the aforementioned challenges.

The remainder of this research will delve specifically and in granular detail into one particular tool situated prominently within this co-regulatory, 'regulation as authorisation' paradigm: technical standards.

The subsequent chapters will critically analyse the multifaceted role that technical standards, and especially harmonised standards within regulatory frameworks like the New Legislative Framework, are intended to play in the regulation of AI systems.

Chapter 2.
Standards and standardisation: what they are and why they matter

1. Aim of the Chapter

In Chapter 1, we explored the multifaceted motivations for regulating emerging technologies, particularly AI, alongside the challenges inherent to such efforts and the diverse regulatory tools available. Building on this foundation, the present Chapter 2 narrows its focus to one such tool: technical standards.

This chapter prioritises technical standards because they are instrumental to the implementation of the AI Act, serving as the operational backbone for its risk-based regulatory framework. Although distinct from formal legal rules, they possess significant socio-political, economic, and regulatory weight. Unlike traditional legal frameworks, these standards are primarily developed by private actors, the ESOs, but their influence extends to shaping markets, technologies, and behaviour. This importance, however, is intertwined with complex issues like legitimacy, enforcement, and power dynamics, which will be further examined in Chapter 3.

This chapter is structured into four principal sections. The first section provides a concise historical overview of the evolution of formal standardisation. Following this, the second section offers a multidimensional perspective on standards, identifying their various forms, typologies, and functional roles. The third section undertakes an analysis of the standardisation process itself, delineating its key stakeholders, foundational principles, and the internal dynamics of the working groups. The fourth and final section scrutinises the mechanisms through which technical standards are enforced, with a particular emphasis on conformity assessment procedures (CAPs) and the challenges inherent in their practical application.

2. A short history of standards and standardisation

2.1. Three waves of standardisation

Standardisation has continuously reshaped human interactions. From its earliest trickles in shared concepts to the mighty currents of global digital protocols, it illustrates humanity's persistent drive to bridge distances, simplify complexities, and ensure a common ground for progress, whether through agreed-upon units or the intricate rules governing a globalized world. This process, which can be defined as the establishment and implementation of common and repeated use of rules, conditions, guidelines, or characteristics for activities or their results, is a fundamental precondition for any form of cultural and economic integration. At its core, standardisation is an exercise in creating shared conceptual frameworks and semantic congruity; without such a foundation, meaningful communication, let alone complex transactional exchanges, cannot be established. Every stratum of human interaction, from the elemental

structure of a common language to the intricate legal and social conventions that permit communal living, relies upon a degree of implicit or explicit standardisation[129].

While the impulse to standardise is an age-old and transcultural phenomenon, not one specifically indigenous to Europe, attempts to impose norms and standards on a grand, systematic scale gained unprecedented momentum on the continent toward the end of the eighteenth century[130]. This period introduced an entirely new rationale to the process, shifting it from a matter of local custom to a project of scientific rationalism and universalist ambition. In their modern incarnation, standards represent a form of codified technical knowledge that enables the predictable and efficient development of products and processes. They perform a tripartite function: regulative, by setting parameters for safety and quality; cognitive, by simplifying information and reducing complexity; and normative, by establishing benchmarks for best practice[131]. The synergistic effect of these functions is an increase in cooperation, a reduction in transaction costs, and the generation of powerful network effects. Once established, a standard creates an infrastructure—both physical and conceptual—that incentivizes its continued use and expansion, thereby reinforcing its own dominance. This process, however, is rarely seamless or uncontested. The history of standardisation is replete with tensions between local and regional interests, which often possess deeply entrenched and particularistic traditions, and the centralizing, homogenizing forces that seek to impose a universal norm. The negotiation between these poles has been a central dynamic in the evolution of modern standardisation.

The earliest roots of this process, such as the emergence of basic units for weights and measures, are relatively straightforward to account for, driven by the rudimentary necessities of barter, taxation, and construction[132]. Similarly, the development of money as a medium of exchange represents a highly sophisticated and early form of standardisation, abstracting value into a universally accepted token. Historical evidence suggests that the utility of such standards is so profound that they can be imposed by non-coercive, consensus-based mechanisms, even within what might be termed an anarchic society, simply because their advantages in facilitating cooperation and exchange are self-evident[133]. A notable precursor to industrial standardisation can be identified in the late seventeenth century, when Dutch shipbuilders developed the "herring buss," a type of fishing vessel constructed from pre-fabricated, modular parts. This innovation, which followed a much older tradition of modular galley construction in the Venetian Arsenal during the medieval period, significantly limited the variety of required replacement parts and enabled rapid, almost assembly-line production, thereby conferring a substantial economic advantage[134].

However, the watershed moment that marks the transition from ad-hoc or commercially driven standardisation to a systematic, science-based endeavor was the comprehensive revision of weights and measures initiated during the French Revolution[135]. This was the first instance of standardisation conceived and executed through scientific deliberation and international conference. The metric system

---

[129] Roland Wenzlhuemer, 'The History of Standardisation in Europe' [2010] Ego. European History Online <https://ieg-ego.eu/en/threads/transnational-movements-and-organisations/internationalism/roland-wenzlhuemer-the-history-of-standardisation-in-europe>.

[130] JoAnne Yates and Craig N Murphy, *Engineering Rules* (Johns Hopkins University Press 2019) <https://doi.org/10.1353/book.66187> accessed 16 November 2023.

[131] Panagiotis Delimatsis, 'Global Standard-Setting 2.0: How the WTO Spotlights ISO and Impacts The Transnational Standard-Setting Process' [2018] SSRN Electronic Journal <https://doi.org/10.2139/ssrn.3184824> accessed 4 April 2024.

[132] V Krislov, *How Nations Choose Product Standards and Standards Change Nations* <https://upittpress.org/books/9780822956228/> accessed 16 November 2023.

[133] ibid.

[134] ibid.

[135] Wenzlhuemer (n 88).

was revolutionary not only in its political context but also in its philosophical design. Its proponents sought to create a system of "natural" standards, deriving the meter from a quadrant of the Earth's meridian, thus grounding it in a universal physical reality rather than the whim of a monarch or the historical accident of a nation. The system was characterised by a rational interconnectedness between units of length, volume, and mass, and a consistent application of decimal divisions, features designed to ensure its universal intelligibility and acceptability, free from any overt national symbolism[136]. The political upheaval of the Revolution provided a rare tabula rasa, a moment in which long-standing, chaotic systems of measurement could be swept away. This ambitious project was propelled by a confluence of imperatives: the pressing need for common national and international standards to rationalize commerce and trade, the administrative necessity of managing large, newly consolidated territories, and the desire among scientists to facilitate collaboration through a shared, precise metrological language. Indeed, the scientific community, populated by figures such as Lavoisier, Condorcet and Laplace, served as the primary and most effective lobbyists for this new standardised regime[137].

The nineteenth century subsequently witnessed the first tentative forays into applying this new model of standardisation on an international scale, building upon the long-standing networks of contact and exchange that characterised the European continent. Throughout the century, this impulse toward international agreement manifested in several key domains, each following a slightly different trajectory but all driven by the logic of interoperability in an increasingly interconnected world. The exigencies of international telecommunication led to harmonised telegraphic codes; the expansion of railway networks necessitated agreements on standardised track gauges to allow trains to cross borders seamlessly; the global postal system required consensus on letter sizes and weights; and the coordination of transport and communication schedules compelled the establishment of internationally agreed-upon time zones[138].

A powerful and often brutal impetus for industrial standardisation came from the realm of military and national defense[139]. The demands for mass-produced, interchangeable weapons and equipment, particularly during the escalating arms races of the late nineteenth century and the cataclysmic world wars of the twentieth, created an undeniable case for rigorous standards. The ability to manufacture ammunition that would fit any rifle of a given model, or to replace a damaged part on a piece of artillery with one made in a different factory, became a matter of strategic necessity.

One of the most persistent and vexing problems driving the early development of technical standards was the issue of screw threads[140]. With the more and more important role of industrial machines in production processes, the inability to ensure that a bolt from one manufacturer would fit a nut from another created enormous inefficiencies and maintenance nightmares. The first systematic attempt to resolve this chaos came in Great Britain, where the engineer Joseph Whitworth developed a standard for thread angle and pitch in 1841, which soon became a de facto national standard. A similar effort was undertaken in the United States by William Sellers, whose system was adopted as the US Standard in 1868. The critical, and often life-threatening, importance of such interoperability was starkly illustrated by the great Baltimore fire of 1904. Firefighting crews arriving from neighboring cities were rendered

[136] David M Fahey and Charles A Gliozzo, 'The Metric System and the French Revolution' (1969) 68 South Atlantic Quarterly 74 <https://doi.org/10.1215/00382876-68-1-74>.

[137] Jonathan H Grossman, 'Standardization (Standardisation)' (2018) 44 Critical Inquiry 447 <https://doi.org/10.1086/696912>.

[138] Krislov (n 132).

[139] Wang Ping, 'A Brief History of Standards and Standardization Organizations: A Chinese Perspective' [2011] East-West Center Working Papers.

[140] Stephen Jay Mihm, 'Visible Hands and Invisible Standards: The Nuts and Bolts of the Second Industrial Revolution' <https://www.hbs.edu/businesshistory/Documents/Mihm%20HBS%202014.pdf>.

helpless as their hoses could not connect to Baltimore's hydrants, which used a different, non-standardised coupling thread. The catastrophic failure of inter-city assistance provided a dramatic and public lesson on the vital need for standardisation in public safety infrastructure[141].

Between the end of XIX century and the beginning of XX century, governments in the leading industrializing nations began to establish formal institutions to support and promote industrial and scientific standardisation. Germany founded its prestigious Physikalisch-Technische Reichsanstalt in 1887 to link precision measurement with industrial application, which was followed by the creation of the Deutsches Institut für Normung (DIN) in 1917. The British Engineering Standards Committee (ESC) was established in 1901—the world's first national standards body and the forerunner of the British Standards Institution (BSI). A pivotal moment in international organisation was the establishment of the International Electrotechnical Commission (IEC) in 1906, an body dedicated to standardisation in the vast and rapidly growing field of electrical and electronic engineering. Its creation was largely driven by the practical needs of telegraph engineers and the electrical power industry, who required universally accepted definitions for units like the volt, ampere, and ohm to ensure the compatibility and safety of their globalizing systems[142].

This period from the late nineteenth century through the interwar years can be characterised as the first great wave of modern standardisation, defined by the professionalization of engineering and the rise of private, voluntary standard-setting bodies[143]. As engineering coalesced into a distinct profession with its own specialized knowledge, associations, and educational systems, its practitioners became the chief architects of the new standards regime. Within these national professional and industry associations, technical committees were established to address pressing industrial problems. Early efforts were concentrated in three primary areas: safety, particularly concerning the catastrophic potential of steam boiler explosions; interoperability, epitomized by the aforementioned struggle over screw threads; and performance, focused on materials like steel rails, where standardised specifications were essential for ensuring the durability and safety of the expanding railway networks[144].

From these early endeavors, a set of core principles emerged that would govern the practice of voluntary standardisation for the next century until today. Central among these was the principle of voluntary adoption, distinguishing these standards from government-mandated regulations. Success depended on a process of consensus-building among a balance of interested parties, a tripartite structure that typically included producers, who were concerned with manufacturability and cost; users and consumers, who prioritised performance and safety; and unaffiliated general-interest experts, often from academia or government, who could provide objective technical input. The process itself was designed to be deliberative and respectful, ensuring that all viewpoints were heard and considered before a standard was approved. In the international arena, this principle of balanced representation was translated into a rule of one vote per country, a cornerstone of bodies like the IEC and ISO still today.

The momentum of the standardisation movement waned during the economic turmoil of the Great Depression, only to be revived with renewed urgency by the logistical and industrial demands of the Second World War. This set the stage for the second major wave of standardisation, which was characterised by a more explicitly global focus and a close alignment with the project of international market integration. The International Organisation for Standardization (ISO) was officially launched in

[141] Andrew L Russell, "'Industrial Legislatures": Consensus Standardization in the Second and Third Industrial Revolutions' (2009) 10 Enterprise & Society 661 <https://doi.org/10.1093/es/khp038>.

[142] ibid.

[143] Yates and Murphy (n 130).

[144] ibid.

1947. The initiative was spearheaded by a committee of delegates from the Allied and neutral countries. The stated purpose of ISO was to promote the development of international standards with a view to facilitating the international exchange of goods and services and to fostering cooperation in the spheres of intellectual, scientific, technological, and economic activity. Initially, ISO's primary function was to act as a clearinghouse, recommending the adoption of well-regarded existing national standards as international norms[145].

A major success story from this period, and one that profoundly reshaped the global economy, was the work of ISO technical committees on freight containers, established in the 1950s and 60s[146]. The standardisation of shipping container dimensions, along with related standards for pallets and packaging, created a seamless, intermodal system of transport that drastically reduced the costs and time associated with moving goods across the globe. This seemingly mundane set of technical specifications was a key enabler of the post-war boom in international trade. However, the path to international agreement was not always smooth. The failure to establish a single global standard for colour television broadcasting serves as a salient counterexample, highlighting how political and economic rivalries could trump technical merit. The proliferation of three competing and incompatible systems, the American NTSC, the German PAL, and the French SECAM, was a direct result of national industrial policy and geopolitical maneuvering, fragmenting the global market for decades[147].

This era also saw the formal integration of international standards into the framework of global trade governance. The General Agreement on Tariffs and Trade (GATT) negotiated its Agreement on Technical Barriers to Trade, commonly known as the Standards Code, between 1973 and 1979[148]. This landmark agreement required signatory nations to use relevant international standards as the basis for their domestic technical regulations wherever they existed, unless such standards were deemed inappropriate for reasons of national security, public health, or environmental protection. This provision effectively elevated ISO and IEC standards to a position of presumptive legitimacy in international law, creating a powerful incentive for both nations and industries to participate in their development[149]. Concurrently, the process of economic integration in Europe prompted the formation of regional standardisation bodies, the European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (CENELEC), and later the European Telecommunications Standards Institute (ETSI)[150]. Created from the 1960s through the 1980s, their primary mission was to unify standards activities within the European Economic Community. These bodies pioneered the "New Approach" to technical harmonisation and standards," whereby legislation would define broad, mandatory "essential requirements" for health and safety, leaving industry to meet these objectives through the development of detailed, voluntary technical specifications, or "harmonised standards", which will be further analysed in Chapter 4.

The third wave of standardisation commenced in the 1980s, propelled by the transformative power of computer networking and digital technologies[151]. This era witnessed the rise of a new breed of private standard-setting organisations, such as the Internet Engineering Task Force (IETF) and the World Wide

[145] JoAnne Yates and Craig N Murphy, 'Coordinating International Standards: The Formation of the ISO' (2007) Working Paper <https://dspace.mit.edu/handle/1721.1/37156> accessed 20 July 2025.

[146] TM Egyedi, 'The Standardized Container: Gateway Technologies in Cargo Transportation' in MJ Holler and E Niskanen (eds), *EURAS Yearbook of Standardization* (EURAS 2000).

[147] Rhonda J Crane, *The Politics of International Standards: France and the Color TV War* (Bloomsbury Academic 1979).

[148] Yates and Murphy (n 130).

[149] Delimatsis (n 131).

[150] Wenzlhuemer (n 129).

[151] Yates and Murphy (n 130).

Web Consortium (W3C), whose work would have a profound and ubiquitous influence on the global economy. These "third-wave" organisations were distinct in their aspirations; they were conceived as inherently global entities from their inception, rather than as international federations of national bodies as in the case of ISO and IEC. Their processes, while echoing traditional models in some respects, also introduced significant innovations. The IETF, for example, relies on a global community of volunteer technical experts and develops standards through a consensus-based process, much like its predecessors. However, its definition of consensus is famously and pragmatically summarized as "rough consensus and running code," prioritising technical feasibility and demonstrated implementation over formal voting procedures established in SDOs such as ISO.

This period also saw standardisation expand its scope beyond purely technical specifications into the realm of management systems and social governance. The ISO 9000 series of standards on quality management, first published in 1987, and the ISO 14000 series on environmental management, introduced in 1996, became globally recognised benchmarks for business operations. Certification to these standards became a de facto requirement for participating in many international supply chains, demonstrating the power of standards to shape corporate behaviour on a massive scale[152]. More recently, the principles of standardisation have been applied to complex social issues with the emergence of social responsibility standards. Standards such as SA8000, which focuses on labour rights, and ISO 26000, which provides guidance on social responsibility, address a wide range of issues including human rights, labour practices, environmental stewardship, and anti-corruption measures. The development process for ISO 26000 was particularly noteworthy for its explicit effort to broaden the traditional stakeholder model, involving national delegations composed of representatives from six distinct groups: industry, government, labour, consumers, non-governmental organisations, and a final category for service, support, research, and others, including academics[153].

2.2. Towards a fourth wave of standardisation

While the "three-wave model" of the history of standardisation by Yates and Murphy offers a comprehensive historical framework, recent transformations in the field suggest that we are now entering a fourth wave of standardisation[154]. This new era is characterised by significant changes in the nature of standardisation itself. In particular:

- the nature of standardisation undergoes a significant transformation. In the context of the fourth wave, standardisation is not only about interoperability, efficiency, and the like, but also about responsibility.
- new actors have entered the standardisation arena. Unlike previous waves where technical committees were predominantly composed of representatives from specific industrial sectors and technical fields, the current wave welcomes a wider range of stakeholders from various segments of society.

---

[152] Charles J Corbett, 'Global Diffusion of ISO 9000 Certification Through Supply Chains' in Christopher S Tang, Chung-Piaw Teo and Kwok-Kee Wei (eds), *Supply Chain Analysis: A Handbook on the Interaction of Information, System and Optimization* (Springer US 2008) <https://doi.org/10.1007/978-0-387-75240-2_7> accessed 11 August 2025.

[153] Michaela A Balzarova and Pavel Castka, 'Stakeholders' Influence and Contribution to Social Standards Development: The Case of Multiple Stakeholder Approach to ISO 26000 Development' (2012) 111 Journal of Business Ethics 265 <https://doi.org/10.1007/s10551-012-1206-9>.

[154] Alessio Tartaro, 'The Fourth Wave of Standardisation: A New Kind of Standard-Setting Is Emerging and Standards for Responsible and Trustworthy AI Are Part of It.' *Proceedings of the 28th EURAS Annual Standardisation Conference - Comprehensive Standardisation for Societal Challenges, 19-21 June 2024, Delft, The Netherlands* (Forthcoming).

- the scope of standards, i.e., what is being standardised, is broadening. While traditionally centered on technical specifications for products or services, there is now a move towards standards that cover complex governance processes.
- standards are becoming more and more vague. As they move from purely technical matters into normative issues and value judgments, standards become difficult to follow unambiguously, and claims of compliance become debatable.

These characteristics of the fourth wave of standardisation are not isolated; rather, they are deeply interconnected, each influencing and reinforcing the other[155]. They all represent a significant transformation for standards and standardisation, but it does come with its challenges. In the rest of this section, we will explore each of the four characteristics in greater detail.

2.2.1. Responsible standardisation

As society becomes increasingly dependent on technologies, the need to design, develop, and use them responsibly has become more pressing. Discussions on Responsible Research and Innovation (RRI) have existed for decades; however, the concept of "responsible standardisation" has only recently gained attention[156]. This topic is increasingly important because standards facilitate the deployment of those technologies that significantly impact society.

Responsible standardisation can be defined as "a process that establishes socially desirable standards" [157]. For a standardisation process to be deemed responsible, it must adhere to certain criteria, such as openness and inclusivity. The assumption here is that aligning the process with specific norms enhances the development of standards beneficial to society. However, other authors argues that this is not always the case[158]. Under some circumstances (e.g., technical domain such as USB standards), adherence to norms such as "openness" and "inclusivity" may not always be necessary or beneficial. On the contrary, such norms might hinder progress rather than foster it.

Further exploration into the norms governing responsible standardisation reveals additional dimensions other than "openness" and "inclusivity." Key attributes of responsible standardisation, such as being "anticipatory," "reflective," and "responsive," align with the principles of Responsible Research and Innovation[159]. An extended list of principles includes "open meeting," "consensus," "due process," "open intellectual property," "transparent membership," "open roles," and "open history."[160] However, current practices within main standard development organisations often fall short of embodying these principles comprehensively[161]. This discrepancy prompts a call for reform, particularly for technologies with significant societal implications, advocating for a standardisation process that prioritises the early evaluation of societal impacts alongside maintaining rigorous technical standards. This revised approach

---

155 Rohan Light and Enrico Panai, 'The Self-Synchronisation of AI Ethical Principles' (2022) 1 Digital Society 24 <https://doi.org/10.1007/s44206-022-00023-1>.

156 Almar Meijer and others, 'Towards Responsible Standardisation: Investigating the Importance of Responsible Innovation for Standards Development' (2023) 0 Technology Analysis & Strategic Management 1 <https://doi.org/10.1080/09537325.2023.2225108>.

157 Martijn Wiarda and others, 'Responsible Innovation and De Jure Standardisation: An In-Depth Exploration of Moral Motives, Barriers, and Facilitators' (2022) 28 Science and Engineering Ethics 65 <https://doi.org/10.1007/s11948-022-00415-z>.

158 Kai Jakobs, 'Responsibility by Design?! – On the Standardisation of "Smart" Systems' *Smart Technologies and Fundamental Rights* (Brill 2020) <https://doi.org/10.1163/9789004437876_014> accessed 22 January 2024.

159 Meijer and others (n 156).

160 Kai Jakobs, 'Today's Situation and Some Future Problems' (2023) 2 Standardization - Journal of Research and Innovation 16 <https://doi.org/10.58831/2831-7920-2.4.16>.

161 ibid.

encourages more inclusive, anticipatory, communicative, and responsive engagement across various disciplines and stakeholders.[162].

Efforts to align with some of these principles are underway. The ISO, for example, includes the promotion of inclusivity and diversity within its operations as one of its priorities[163]. Similarly, the IEC Council Board established a Task Force on Diversity in 2019[164]. At the European level, amendments to Regulation 1025/2012 on European standardisation seek to amplify the voice of social stakeholders (i.e., Annex III organisations representing consumers, workers, SMEs and environmental stakeholders) in CEN, CENELEC, and ETSI[165]. Despite these advancements, however, the implementation of certain principles remains incomplete. For instance, the opaque composition of WGs and the anonymity of contributors to standards development in some SDOs challenge the principles of transparent membership and open history, respectively. A related problem concerns the copyright protection of standards themselves. As recently ruled by European Court of Justice in the Public.Resource.Org case (ECJ 05.03.2024, C-588/21 P), harmonised standards, i.e., technical standards that can be used to achieve a presumption of conformity with some EU legislations, shall be publicly and freely accessible to the public. At the same time, however, most of these standards are protected by copyright and are accessible only under the payment of a fee. The ruling of the EUCJ will likely change this situation[166].

2.2.2. New actors

The concept of responsible standardisation, emblematic of the fourth wave, underscores the pivotal roles of openness and inclusiveness. These principles seek to address the long-standing issue of underrepresentation in the standardisation process. Historically, standardisation predominantly involved engineers, specifically those representing industry stakeholders in the respective fields[167], despite the increasing social impact of standards and the technologies they enable. The pursuit of responsible standardisation endeavours to rectify this through diversified participation, which can also enhance the legitimacy of standards[168].

Recent research in standardisation has evolved to recognise a broader array of stakeholders involved in the standardisation process. This includes users, manufacturers, suppliers of complementary goods, standards-setting organisations, and government bodies [169]. The dynamics among these actors, and their influence on the standardisation process, are significantly shaped by the mode of standardisation—be it committee-based, market-based, or government-based [170]. Despite the theoretical inclusivity of committee-based standardisation, practical engagement reveals disparities in the influence exerted by different groups. Users, consumer advocates, labour organisations, and civil society groups often find

---

[162] ibid.

[163] ISO, 'ISO Strategy 2030' <https://www.iso.org/publication/PUB100364.html>.

[164] IEC, 'Inclusion and Diversity' <https://iec.ch/basecamp/inclusion-and-diversity> accessed 8 February 2024.

[165] Regulation (EU) 2022/2480 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 1025/2012 as regards decisions of European standardisation organisations concerning European standards and European standardisation deliverables (Text with EEA relevance) 2022 (OJ L).

[166] Olia Kanevskaia, 'Is It Really All about the Money? The Future of European Standardization after PublicResourceOrg' (2025) 16 European Journal of Risk Regulation 344 <https://doi.org/10.1017/err.2024.64>.

[167] Yates and Murphy (n 130).

[168] Raymund Werle and Eric J Iversen, 'Promoting Legitimacy in Technical Standardization' (2006) 2 Science, Technology & Innovation Studies 19.

[169] Seungyeon Moon and Heesang Lee, 'The Primary Actors of Technology Standardization in the Manufacturing Industry' (2021) 9 IEEE Access 101886 <https://doi.org/10.1109/ACCESS.2021.3097800>.

[170] Paul Moritz Wiegmann, HJ de Vries and K Blind, 'Multi-Mode Standardisation: A Critical Review and a Research Agenda' (2017) 46 Research Policy 1370 <https://doi.org/10.1016/j.respol.2017.06.002>.

themselves marginalised due to a lack of necessary resources and expertise, thereby perpetuating the dominance of major market actors[171].

Nonetheless, the fourth wave of standardisation is characterised by the increasing involvement of stakeholders other than the traditional ones. The development of ISO 26000 serves as a notable example, as it involved consumer representatives, labour unions, NGOs, and an additional group of service, support, research, and other entities[172]. ISO 26000, however, represented an exceptional case, as it was the result of a conscious decision by the ISO to broaden participation. Yet, in scenarios reliant on volunteer experts, obstacles related to expertise and resource persist, potentially limiting the scope for meaningful engagement.

2.2.3. Broadening scopes

The fourth wave of standardisation marks an expansion in the scope of standards. Unlike earlier waves, where standards primarily provided technical specifications for products, the fourth wave encompasses more complex processes. These include, for example, risk management, anti-corruption measures, innovation, workers' health and safety, and environmental responsibility, for each of which an international standard exist. In the standardisation lexicon, these are referred to as "management system standards"[173]. From another perspective, scholars in public policy and political science—who are interested in standards as de facto substitutes for government regulation—describe these as "standards of good governance" [174].

These particular kinds of standards embody the principles of optimisation and rationalisation of traditional standards within the context of complex societal challenges that are highly influenced by value judgments[175]. Although such standards aim commendably to offer solutions for the public good, they have not been immune to criticism, as highlighted by critical perspectives in the literature[176]. Most of these critiques centre on the legitimacy of private standards serving as governance tools for processes of public significance[177]. However, a more fundamental issue lies in the very nature of these standards. They presuppose the existence of "correct" approaches to managing processes whose correctness criteria are not only contestable but actively contested.

The case of risk management is illustrative in this context. Standards such as ISO 31000 provide guidelines for managing risk across various domains and by different actors, yet they overlook some assumptions underlying the process itself. This oversight is evident in two aspects: the definition of risk and the acceptability of risk levels.

---

171 Jean-Christophe Graz, 'Global Corporations and the Governance of Standards' in Andreas Nölke and Christian May (eds), *Handbook of the International Political Economy of the Corporation* (Edward Elgar Publishing 2018) <https://doi.org/10.4337/9781785362538.00037> accessed 8 February 2024.

172 Balzarova and Castka (n 153).

173 Pavel Castka and Charles J Corbett, 'Management Systems Standards: Diffusion, Impact and Governance of ISO 9000, ISO 14000, and Other Management Standards' (2015) 7 Foundations and Trends® in Technology, Information and Operations Management 161 <https://doi.org/10.1561/0200000042>.

174 Jens Steffek and Philip Wegmann, 'The Standardization of "Good Governance" in the Age of Reflexive Modernity' (2021) 1 Global Studies Quarterly ksab029 <https://doi.org/10.1093/isagsq/ksab029>.

175 Stefan Timmermans and Steven Epstein, 'A World of Standards but Not a Standard World: Toward a Sociology of Standards and Standardization' (2010) 36 Annual Review of Sociology 69 <https://doi.org/10.1146/annurev.soc.012809.102629>.

176 Lawrence Busch, 'Standards and Their Problems: From Technical Specifications to World-Making' in Mara Miele and others (eds), *Research in Rural Sociology and Development*, vol 24 (Emerald Publishing Limited 2017) <https://doi.org/10.1108/S1057-192220170000024005> accessed 8 February 2024.

177 Lawrence Busch, *Standards: Recipes for Reality* (The MIT Press 2011) <https://doi.org/10.7551/mitpress/8962.001.0001> accessed 16 November 2023.

Firstly, definitions of risks adopted in international standards such as ISO 31000 (e.g., "effect of uncertainty on objectives") conceal the normative dimension of the concept, allowing different actors to interpret these definitions based on their positions, interests, and objectives [178]. Consequently, what is considered a risk in a specific context remains undefined within the standard itself. This leads to the almost paradoxical situation where, as the object of standardisation broadens, it eventually transcends the scope of the standard.

Secondly, the goal of the risk management process is often described as reducing risk to an acceptable level. However, "acceptable" is a normative concept as well [179], making the notion of "acceptable risk" a thick concept. The determination of what is "acceptable" in each context is influenced by the positionality of the actors involved, i.e., the varied ways in which individuals perceive a situation.

This variability renders the outputs of the standard contestable. An organisation may claim conformity with a risk management standard, yet still face criticism from those with a different interpretation of what constitutes risk or when a level of risk is acceptable. While this may seem self-evident, the novelty lies in the fact that such criticisms can be leveraged by appealing to the very standard that the organisation claims to conform to. The root cause of this issue is the inherent vagueness of these standards, as detailed further.

This is not to say that standards in risk management and other areas such as environmental, social responsibility, anti-bribery, etc., are inherently flawed. On the contrary, they often represent the best tools we have at our disposal. This discussion merely aims to acknowledge the reality that, as the scope of standards expands, they evolve and encounter new challenges.

2.2.4. Vagueness

Standards are fundamentally texts and documents, and so they are imbued with meaning and interpreted by people using them. Although, to my knowledge, no existing studies have proposed textual analysis of the content of standards, this chapter posits that such kind of analysis is useful for understanding the nuances of standards in the fourth wave. It particularly illuminates the growing vagueness of these standards, marking a significant departure from the historical precision characterising technical specifications.

The previous section's analysis of risk management standards provides an illustrative example of this vagueness. Terms like "risks" and "acceptable levels of risk" are interpreted variably across different groups and contexts. This aligns with Hempel's theory that the vagueness of a term is a function of its application by a group to a set of objects, leading to discrepancies in the application of the term [180]. This results in the fact that "different observers will not always give the same answer to the question as to whether the symbol does or does not apply to a given case; in other words, the vagueness of a symbol has been characterised by certain variations occurring in its use"[181].

This paper extends this argument to standards in the fourth wave. In order to demonstrate their increasing vagueness, a comparative analysis of three requirements and recommendations from three distinct standards is proposed here:

1) "AUI cables for 10BASE-FP and 10BASE-FL shall not exceed 25 m. (IEEE 802.3-2015, 13.3).

[178] Kirsten Juhl, 'Standardization of Risk versus the Risk of Standardization: A Conceptual Analysis' *Standardization and Risk Governance* (Routledge 2019).

[179] Odd Einar Olsen, 'The Standardization of Risk Governance' *Standardization and Risk Governance* (Routledge 2019).

[180] Carl G Hempel, 'Vagueness and Logic' (1939) 6 Philosophy of Science 163.

[181] ibid.

2) "The quality objectives shall be relevant to conformity of products and services and to enhancement of customer satisfaction "(ISO 9001:2015, 6.2.1).
3) "An organisation should actively promote ethical behaviour by developing and using governance structures that help to promote ethical behaviour within the organisation, in its decision making and in its interactions with others" (ISO 26000:2010, 4.4).

The first requirement, from an older version of IEEE 802.3 for Ethernet, mandates a precise measurement for AUI cables. This specification is very precise and leaves little room for interpretation. It is easy for different groups of people to agree that a particular cable does not, for example, exceed 25 meters. Such precision is crucial for ensuring interoperability and network functionality. In contrast, the second requirement, from ISO 9001, introduces more ambiguity with terms like "relevance" and "customer satisfaction," which are open to interpretation based on organisational context and customer expectations. Although methods exist to assess these aspects, the standard demands organisations define their own criteria, leading to variability in standard application. The third requirement, from ISO 26000, plunges further into vagueness by dealing with "ethical behaviour" and its promotion through appropriate governance structures—concepts highly subjective and context-dependent. The lack of measurable criteria for evaluating these requirements introduces a significant degree of vagueness, necessitating qualitative assessments and consensus on ethical principles, which can vary widely.

It should be observed, however, that vagueness in standards is not necessarily a bug, but can be a feature. This intentional flexibility allows standards to be adaptable across diverse industries, cultural contexts, and organisational structures and so it encourages entities to tailor the standards to their specific needs. The open-ended nature of standards like ISO 9001 and ISO 26000 facilitate broader applicability and encourage organisations to engage in deeper reflection and dialogue about their practices and values.

These are just examples, but they should be enough to illustrate the point effectively. Vagueness is not an attribute exclusive to ISO 26000 but extends to all standards with broadly defined subject matter and boundaries, such as social responsibility, risk management, innovation, environmental management etc. In fact, this vagueness results exactly from broadening the scope of standards. In turn, the latter is a consequence of the development of standards that address complex societal issues, which is linked to responsible standardisation. One effect of broadening scopes is the increasing inclusion of new actors in standard-making, as more actors have a stake in the standards. Finally, it can be speculated that the inclusion of these new stakeholders, each bringing unique perspectives, expertise, and experiences, further contributes to the vagueness of the standards due to the intensified negotiation process they necessitate. This dynamic underscores that the four characteristics of the fourth wave of standardisation are not isolated; rather, they are deeply interconnected, each influencing and reinforcing the other. Ultimately, the interconnectedness of these trends underscores a fundamental shift in how standards function, moving from rigid rules to adaptable frameworks for navigating complex societal and business challenges, including those related to AI.

3. Standards: a multidimensional perspective

3.1. "A document…": defining standards

Having delineated the broad historical process, including its more recent developments, that culminated in the advent of standardisation as it is currently understood, the inquiry must now shift its

focus to the standards themselves. This analysis will commence with an examination of the very definition of a standard and its inherent functions.

The most authoritative and widely cited definition is provided by the ISO, which defines a standard as a "document, established by consensus and approved by a recognised body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context."[182] In a note, the ISO/IEC Guide 2:2004 further specifies that "Standards should be based on the consolidated results of science, technology and experience, and aimed at the promotion of optimum community benefits." A meticulous exegesis of this comprehensive definition not only allows for the identification of the essential constituents of a standard but also illuminates its multifaceted functions, which will be explored in greater depth subsequently in this chapter.

The first part of the ISO definition, "a document established by consensus and approved by a recognised body," serves to delineate the procedural origins and institutional legitimacy of a standard. The designation of a standard as a "document" constitutes its most fundamental, yet critically important, attribute As a document, a standard is not merely inert; rather, it possesses the capacity to exert agency, a concept extensively discussed within the literature on the philosophical concepts of documentality[183]. Precisely because they are documents, standards can have a profound effect on the world, as has been previously mentioned in preceding chapters and will be further explored in subsequent ones.

The principle of establishment by "consensus" is the cornerstone of a standard's development and adoption. ISO/IEC Guide 2:2004, consensus in defined as "general agreement, characterised by the absence of sustained opposition to substantial issues by any important part of the concerned interests and by a process that involves seeking to take into account the views of all parties concerned and to reconcile any conflicting arguments". In this way, "consensus need not imply unanimity". However, the consensus-based standardisation process is sealed by the approval of a "recognised body," which confers formal legitimacy and authority upon the document. This endorsement by an entity such as ISO, IEC, or a national standards body like ANSI or BSI, signifies that the standard is not a mere private agreement but has been subjected to a rigorous, transparent, and impartial vetting procedure. The recognised body thus acts as a guarantor of procedural integrity and a stamp of quality, instilling confidence in the validity and reliability of the standard among its potential users.

The subsequent portion of the definition, which states that a standard "provides, for common and repeated use, rules, guidelines or characteristics for activities or their results," elaborates upon the content, nature, and intended application of a standard. The stipulation "for common and repeated use" is central to a standard's economic and practical utility. It underscores that standards are engineered not for singular, bespoke applications but as scalable solutions to recurring problems. The immense value derived from standards for A4 paper dimensions, USB connectivity, or intermodal shipping containers is predicated entirely on their consistent and repetitive application by a multitude of actors. This focus on repetition is the very mechanism that generates interoperability, market predictability, and profound efficiencies in production and global supply chains. Furthermore, the definition's flexible specification of content as "rules, guidelines or characteristics" allows standards to serve a wide array of purposes. Rules are typically prescriptive and mandatory, often pertaining to critical areas such as safety or direct

[182] 'ISO/IEC Guide 2:2004 - Standardization and Related Activities — General Vocabulary' 2 <https://www.iso.org/standard/39976.html> accessed 12 August 2025.

[183] Barry Smith, 'How to Do Things with Documents' [2012] Rivista di estetica 179 <https://doi.org/10.4000/estetica.1480>; Barry Smith, 'Document Acts' in Anita Konzelmann Ziv and Hans Bernhard Schmid (eds) (Springer Netherlands 2014) <https://doi.org/10.1007/978-94-007-6934-2_2> accessed 6 December 2023; Maurizio Ferraris, *Documentality: Why It Is Necessary to Leave Traces* (Fordham University Press 2013).

interoperability, for instance, defining a specific electrical voltage range. Guidelines, in contrast, are recommendatory, offering best-practice frameworks for achieving desired outcomes without being obligatory, as exemplified by the ISO 26000 standard on Social Responsibility[184]. Characteristics refer to quantifiable attributes of a product, material, or service, such as the minimum tensile strength of steel, thereby defining objective parameters for quality and performance. This content structure is complemented by the critical duality of a standard's scope, applicable to "activities or their results." This allows standardisation to govern either processes (the *how*) or outputs (the *what*). The renowned ISO 9001 standard, for example, specifies requirements for a Quality Management System, an activity, rather than for a particular product. Conversely, myriad standards define the explicit properties of a final product, from the thread pitch of a screw to the data format of a digital image. This dual focus is profoundly significant, as it enables standards to shape everything from managerial methodologies and production processes to the final characteristics of goods and services.

The ultimate purpose of a standard is articulated in the definition in the phrase "aimed at the achievement of the optimum degree of order in a given context." This clause reveals the core teleological drive of standardisation. Here, "order" is to be understood as the establishment of predictability, consistency, and efficiency through the strategic reduction of unnecessary variety and complexity. The most sophisticated and crucial term in this passage is "optimum." The definition deliberately eschews "maximum" order, which would invariably lead to market rigidity, stifle innovation, and foreclose competition. Instead, "optimum" order signifies a carefully calibrated equilibrium, a strategic balance that imposes sufficient structure to guarantee safety, interoperability, and efficiency, while preserving the necessary latitude for technological advancement and market differentiation. It is a search for a 'sweet spot' between chaos and over-regulation. This objective is further grounded in reality by the qualifier "in a given context," which acknowledges that a standard is not a universal, abstract law but a tool designed for a specific technological, industrial, or social domain. The requirements for achieving optimum order for sterile medical equipment operate in a vastly different context from those for managing enterprise IT services. This contextual specificity ensures that standards remain relevant, fit-for-purpose, and effective, preventing the misapplication of rules and recognising that the ideal state of order in one field may represent dysfunction in another.

Finally, the addendum to the definition provides the epistemological and ethical foundations upon which standards must be constructed, stating they "should be based on the consolidated results of science, technology and experience, and aimed at the promotion of optimum community benefits." This first principle establishes the imperative for an evidence-based foundation. Standards are not to be derived from arbitrary opinion or narrow commercial interest; they must be a synthesis of the collective wisdom of a field. This includes the empirical data and theoretical frameworks of "science," the state-of-the-art and practical achievability of "technology," and the invaluable real-world knowledge and lessons learned from expert "experience." Taken together, these inputs ensures that standards are technically sound, robust, credible, and reflect a consolidated understanding at a given point in time. The concluding phrase, "aimed at the promotion of optimum community benefits," elevates the purpose of standardisation from a purely technical objective to a societal one. It provides the ultimate rationale for the endeavor, asserting that the creation of order is not an end in itself but a means to a greater good, linking the technical specifications of standards to tangible, positive outcomes for society as a whole. These community benefits are vast and varied, encompassing consumer safety and confidence, enhanced

[184] Stéphanie Bijlmakers, 'No ISO Fix for Human Rights: A Critical Perspective on ISO 26000 Guidance on Social Responsibility' *Research Handbook on Global Governance, Business and Human Rights* (Edward Elgar Publishing 2022) <https://www.elgaronline.com/edcollchap/edcoll/9781788979825/9781788979825.00017.xml> accessed 23 May 2024.

economic efficiency through facilitated trade and reduced costs, environmental protection via limits on pollution or mandates for energy efficiency, and the promotion of social equity through, for instance, standards for accessibility.

3.2. Anatomy of a standard

Technical standards are distinguished not only by their objectives and the way they are developed, but also by their peculiar structure as documents. In a sense, the structure of a standard is itself standardised. The recurring elements of a standard, scope, references, definitions, and normative-informational divisions, operate as mechanisms of epistemic governance, constructing domains of authority while embedding contested social and technical assumptions into institutional frameworks.

The "scope" section of a standard serves as its foundational epistemic boundary, delineating the terrain of applicability through declarative assertions. Far from being a passive introduction, the scope actively defines what constitutes a legitimate object of standardisation, often materializing conceptual categories that previously lacked institutional recognition. The definition of a scope transforms abstract problems into bounded, manageable entities subject to technical prescription. By establishing what is included and, just as critically, what is excluded, the scope performs a powerful act of epistemic closure. The ISO/IEC Directives, Part 2, in Clause 14, mandates that the scope must "clearly define the subject of the document and the aspects covered"[185], thereby making this very practice of boundary-work as a prerequisite for standardisation itself.

Following the scope, the "normative references" section anchors the standard within a broader ecosystem of interrelated documents. These citations are not merely bibliographic formalities but functional linkages that bind standards into a recursive network of mutual reinforcement. Such intertextuality ensures technical coherence across domains, yet it also entrenches path dependencies. When a new standard normatively references an existing one, for example for a specific test method or a set of material specifications, it inherits and propagates the technical assumptions and compromises embedded in the referenced document. This creates a dense web of interdependencies that can stifle radical innovation by increasing the costs and complexity of deviation from established norms. The authority of any single standard is thereby amplified by the perceived solidity of the entire system to which it is linked, a system whose internal consistency is a primary objective. This self-referential network solidifies the hegemony of the standardizing body, making it difficult for alternative systems or specifications to gain traction.

Another crucial part of a standard is its "definitions" section. As stipulated in ISO/IEC Directives, Part 2, Clause 16, definitions must adhere to stringent syntactic and semantic rules, such as grammatical alignment between term and explanation, taxonomic classification via genus-differentia structures, and the exclusion of unit-based descriptors, all aimed at eliminating ambiguity. However, this veneer of semantic neutrality belies the constructive power of definitions. To define a term like "sustainability" or "state of the art" within a standard is not a passive act of description but an active and often politically charged intervention that forecloses alternative meanings. The chosen definition privileges certain values, technical approaches, or economic interests while marginalizing others. This act of terminological closure constructs the very reality it purports to describe, establishing a shared conceptual universe that is a

[185] 'ISO/IEC Directives Part 2' <https://www.iec.ch/standards-development/isoiec-directives-part-2> accessed 12 August 2025.

prerequisite for coordinated action but may also serve to normalize a particular view of the problem addressed in the standard.

After the preliminary sections of scope, references, and definitions, the subsequent clauses of a standard articulate its substantive content through a carefully managed distinctions between requirements, recommendations and permissions. Some clauses, marked by the performative modal verb "shall," impose requirements that are enforceable through conformity assessment. This specific linguistic choice, rigorously policed by standard development organisations like ISO, transforms the text from a description into a requirement, creating binding criteria for compliance. In stark contrast, provisions articulated with "should" signal recommendations, while those using "may" indicate permissions. These latter forms often represent negotiated compromises among technical committee members with divergent interests or reflect aspirational benchmarks for which consensus on mandatory enforcement could not be achieved. This epistemological tension between the coercive force of "shall" and the advisory nature of "should" underscores the paradoxical character of standards: they are simultaneously instruments of coercion and sites of contestation, where technical consensus is visibly demarcated from ongoing debate and future development[186].

The anatomy of a standard, therefore, reveals it as both a textual artifact and a sociopolitical institution. Its structural and linguistic conventions, from the boundary-setting scope to the performative force of its verbal forms, are far from incidental; they are the very mechanisms that enact epistemic closures, construct technical and social realities, and mediate power relations. While these features are essential for enhancing the predictability and coordination that standards promise, they also carry the inherent risk of institutionalizing a form of technocratic hegemony. This necessitates critical scrutiny of whose interests and assumptions are inscribed into or, conversely, excluded from the standard's formal architecture. Such an analysis situates the practice of standardisation within broader scholarly debates concerning governance, knowledge production, and the material-discursive practices that shape contemporary sociotechnical systems.

3.3. Different types and functions of standards

Standards, while superficially constituting elementary documents containing precise technical specifications, criteria, and definitions to ensure products, processes, and services fulfill their intended purpose, possess a far more profound significance. They function as fundamental institutional frameworks that structure complex social and economic interactions, providing the essential scaffolding for coordination, exchange, and technological integration within and across societies[187]. Despite this shared foundational characteristic, standards exhibit considerable heterogeneity, having been established to serve a remarkably diverse array of specific objectives. Although multiple taxonomies have been proposed to categorise this variety[188], this work proposes an alternative based on the function of the standard within the broader ecosystem of standardisation.

Foremost among these foundational categories are "Vocabulary standards", encompassing comprehensive glossaries and terminological definitions. These instruments provide the critical uniformity and linguistic cohesion indispensable for the unambiguous interpretation of technical terms

[186] Busch (n 177).

[187] Joseph Farrell and Garth Saloner, 'Coordination Through Committees and Markets' (1988) 19 The RAND Journal of Economics 235 <https://doi.org/10.2307/2555702>.

[188] Knut Blind, 'A Taxonomy of Standards in the Service Sector: Theoretical Discussion and Empirical Test' [2006] The Service Industries Journal <https://doi.org/10.1080/02642060600621597> accessed 12 August 2025.

deployed across the entire spectrum of standardisation efforts and within specialized discourse; without such shared semantic conventions, the potential for costly misinterpretation, operational inefficiency, and systemic failure escalates dramatically, fundamentally undermining the reliability and trust that standards seek to engender.

Equally critical to the edifice of modern technological and scientific endeavour are "Measurement standards", which address the precise definitions and practical realizations of units of measure. With the notable exception of a few fundamental physical constants, units of measurement are essentially arbitrary constructs, deriving their immense practical utility solely from widespread collective agreement and meticulous international codification. This category encompasses the primary standards that define and underpin the seven base units of the International System of Units (SI), such as the metre for length and the kilogram for mass. These base units constitute the immutable bedrock upon which all derived units rest, enabling precise, reproducible, and universally comparable quantification essential for scientific research, industrial manufacturing, global trade, and regulatory enforcement, thereby facilitating objective comparison and collaboration across disciplinary and national boundaries.

Distinct yet vital categories further illustrate the expansive remit of standardisation. "Safety standards", for instance, are explicitly formulated to mitigate hazards and ensure the security of products, processes, workplaces, and environments. They typically prescribe rigorous design parameters, mandatory protective features, specific testing methodologies, and detailed operational protocols aimed at preventing harm to individuals, property, or ecosystems, thereby serving as crucial instruments for risk management, regulatory compliance, and the maintenance of societal trust in technological systems.

Similarly, "Management standards" encompass a diverse array of frameworks and codified techniques designed to govern organisational processes and systemic operations. These range from comprehensive quality management systems (exemplified by the ISO 9000 series) and environmental management systems (ISO 14000 series) to standards for information security management (ISO/IEC 27000 series) and specific documentation protocols governing sectors such as banking transactions or medical records. They provide structured, often auditable, methodologies for achieving consistency, enhancing operational efficiency, ensuring traceability, and fostering continuous improvement in administrative and core business functions, effectively translating strategic organisational objectives into standardised, controlled practice.

However, it is arguably within the domain of "Product standards" that the profound contemporary economic and technological significance of standardisation becomes most immediately tangible and impactful. Product standards meticulously delineate the exhaustive specifications requisite for the consistent production, reliable performance, safe use, and effective marketability of a given material or manufactured item. These specifications frequently encompass exacting requirements regarding physical dimensions, material composition, performance thresholds under defined operational conditions, durability expectations, packaging specifications, labelling information, and the specific, validated testing methods employed to verify conformance. By establishing these common technical baselines, product standards facilitate international trade by reducing technical barriers, ensure consumer protection through predictable quality and safety levels, and enable meaningful technological comparability and substitutability.

The seamless and reliable operation of any complex system composed of interdependent, communicating components is especially contingent upon rigorous adherence to common technical specifications, commonly designated as compatibility or "Interoperability standards". These pivotal instruments explicitly define the interfaces, communication protocols, data formats, signalling conventions, and operational procedures through which disparate technologies or subsystems interact,

exchange information, and function cohesively as a unified whole. Ubiquitous examples include the standards enabling a mobile telephone to authenticate and connect to a cellular network, a software application to execute correctly on an operating system platform, or a peripheral device to interface seamlessly with a host computer. While historically concentrated within the ICT and consumer electronics sectors, the strategic importance of interoperability standards is rapidly proliferating into diverse and critical industrial domains, including advanced manufacturing (Industry 4.0/IoT), smart energy grids, intelligent transportation systems, and integrated digital healthcare technologies. The pervasive density and criticality of such standards within contemporary technological artefacts is strikingly underscored by empirical evidence; a study identified 251 distinct technical interoperability standards implemented within the architecture of a single modern laptop computer, while extrapolating that the total number of relevant standards governing its full functionality, component interactions, and external connectivity would be substantially higher[189].

3.4. Advantages and disadvantages of standards

Technical standards constitute a fundamental infrastructure underpinning modern economies, serving multifaceted public interests by facilitating production, trade, and technological advancement[190]. Quantifying their precise economic impact, however, presents significant methodological challenges due to the complex and pervasive nature of their effects across diverse sectors and temporal scales. Crucially, alongside their widely acknowledged benefits, standards inherently possess the potential to generate unintended adverse consequences, a duality demanding careful consideration. These potential advantages and disadvantages of standards can be systematically examined across several key dimensions.

Standards frequently function as catalysts for innovation and competition[191]. By providing a common technological foundation, they effectively lower barriers to market entry. Standards act as foundational building blocks, encapsulating innovation and best practices, thereby expanding market reach for companies by connecting them to larger networks of buyers, sellers, and partners. This expanded scope directly fosters heightened competition, which typically intensifies later in the product lifecycle, yielding tangible consumer benefits through lower prices, greater choice among sellers, products, and services, and significantly reduced switching costs. Standards act as a bulwark against restrictive practices like tying, enhancing consumer sovereignty and mitigating the risk of technological lock-in, empowering buyers to change suppliers with greater ease and shielding them from the risk of investing in obsolete or unsupported technologies and the associated lack of complementary products. Furthermore, the inherent comparability and common feature sets mandated by standardised offerings simplify evaluation for buyers, enhance market transparency, and foster clearer communication among all market actors. Conformity to an international standard serves as a trusted proxy for quality, safety, or compatibility, enabling informed purchasing decisions without requiring deep expertise in the underlying technology or the specific supplier.

Perhaps most significantly in interconnected sectors and emerging industries, standards are indispensable for achieving interoperability, enabling the seamless combination and interchangeability of

[189] Brad Biddle, Andrew White and Sean Woods, 'How Many Standards in a Laptop? (And Other Empirical Questions)' (Social Science Research Network, 10 September 2010) <https://doi.org/10.2139/ssrn.1619440> accessed 12 August 2025.

[190] Gregory Tassey, 'The Roles and Impacts of Technical Standards on Economic Growth and Implications for Innovation Policy' (2017) 1 Annals of Science and Technology Policy 215 <https://doi.org/10.1561/110.00000003>.

[191] Knut Blind, 'Standardisation as a Catalyst for Innovation' [2009] The Review of Economic Studies 1.

products or services from different vendors[192]. This is acutely evident in ICT, where standards form the bedrock of the global communications infrastructure, ensuring that disparate networks and devices can exchange voice, video, and data seamlessly, effectively speaking the same language. This prevents costly and fragmenting market battles over proprietary technologies and curbs the ability of dominant players to establish exclusionary walled gardens.

From a production perspective, standards drive cost efficiency by enabling economies of scale through interchangeable parts and processes, fostering specialization, and crucially, reducing wasteful duplication of R&D efforts[193]. Historically, as noted in section 2, standardised components and processes were instrumental in enabling cost-efficient mass production, leveraging interchangeability, division of labour, and specialized skills to achieve significant economies of scale and lower unit costs. This efficiency remains indispensable in contemporary economies, where the increasing complexity of industrialized and modernized systems would lead to rampant duplication of R&D efforts were it not for standards codifying benchmarks and best practices. This pooling of R&D resources accelerates technological diffusion and progress, as standards act as critical vectors for technology transfer, enabling the global replication and adaptation of technological breakthroughs achieved in specific locales, thereby accelerating collective innovation and economic growth.

On a broader socio-economic level, international standards promote national development, particularly in emerging economies, by providing vital blueprints for infrastructure development and economic advancement[194]. They facilitate market liberalization and unlock access to broader international markets, enriching domestic players by connecting them to global buyers and sellers.

However, these potential benefits are counterbalanced by significant risks[195]. The standardisation process itself inherently concentrates considerable influence among its participants. The development of a standard by a select group of companies, which may not fully represent all current or future implementers, risks creating outcomes disproportionately favourable to certain interests, thereby detrimentally affecting consumers and rival producers by restricting product choice and impeding competition. Decision-making procedures within SDOs can exhibit biases, often favouring large producers or, occasionally, large institutional buyers like governments or major network operators. Furthermore, participants contributing technology to a standardisation effort typically do so strategically, seeking endorsement to drive widespread deployment and secure competitive advantage. In specific instances, the collective agreement on a common solution can also dampen competitive pressures early in a technology lifecycle, reducing the diversity of available technical approaches compared to a scenario where competing proprietary platforms vie for dominance.

Standards can also be strategically employed or even weaponised for protectionism[196]. When nations unilaterally establish domestic standards diverging from international norms, they can erect significant technical barriers to trade, shielding domestic industries from foreign competition. Compliance with unique national standards can impose prohibitive costs on foreign firms, deterring market entry and

---

[192] Thomas Burns, John Cosgrove and Frank Doyle, 'A Review of Interoperability Standards for Industry 4.0.' (2019) 38 Procedia Manufacturing 646, 4 <https://doi.org/10.1016/j.promfg.2020.01.083>.

[193] Joseph Farrell and Garth Saloner, 'Economic Issues in Standardization' (Massachusetts Institute of Technology (MIT), Department of Economics 1985) Working paper 393 <https://EconPapers.repec.org/RePEc:mit:worpap:393> accessed 21 April 2026.

[194] Dieter Ernst, Heejin Lee and Jooyoung Kwak, 'Standards, Innovation, and Latecomer Economic Development: Conceptual Issues and Policy Challenges' (2014) 38 Telecommunications Policy 853 <https://doi.org/10.1016/j.telpol.2014.09.009>.

[195] Joseph Farrell, 'Economic Issues in Standardization'.

[196] Christopher S Gibson, 'Globalization and the Technology Standards Game: Balancing Concerns of Protectionism and Intellectual Property in International Standards' (2007) 22 Berkeley Technology Law Journal 1403.

fragmenting global markets, as historically exemplified by the proliferation of incompatible television broadcast standards which impeded the development of a unified global TV industry, as discussed in section 2.1.

Market dynamics can also foster inertia in adopting newer or superior standards, particularly when entrenched standards have achieved widespread deployment[197]. A reluctance to migrate from established standards to innovative solutions can hamper competition and technological progress, even when superior alternatives exist. Finally, the widespread adoption of a dominant standard inherently risks a loss of technological and product variety. Standards are inevitably optimized for the broadest market segment, potentially neglecting the specialized requirements of niche user groups.

The strategic significance of standardisation is increasingly acknowledged from both economic and policy perspectives, as it is recognised as a critical mechanism for expanding common markets and promoting sustained socioeconomic development. This recognition is substantiated by the growing attention that policymakers are dedicating to the field. Prior to analysing the role of standards in European Union regulation in the subsequent chapter, the remainder of the current chapter will explore the standardisation process and the mechanisms through which standards are enforced, specifically through conformity assessment procedures (CAPs).

4. The standardisation process

4.1. Standard setting entities

The landscape of standardisation is populated by a diverse array of entities, which can be broadly delineated into three principal categories: single companies, formal SDOs, and industry forums or consortia. The provenance of a standard, that is, the type of entity responsible for its creation, profoundly influences its impact and adoption.

Single companies are the source of what are known as "proprietary specifications," which are often adopted either within the organisation that develops or uses them (e.g., a standard procedure within a McDonald's restaurant), or by other organisations as *de facto* standards. In contrast, formal standards-developing organisations, such as the ISO or CEN-CENELEC, typically produce standards that may be voluntarily adopted by any organisation and are recognised as de jure standards by virtue of their development processes within the SDOs. In the European Union, such standards can also acquire legal value when adopted to support legislations under the NLF. Occupying an intermediate position are forums and consortia. These entities, which include organisations like the IETF and the W3C, most often produce specifications that can be voluntarily adopted like de jure standards but are developed according to the consortia's internal procedures, which often differ from those of formal SDOs.

4.1.1. Proprietary Specifications and De Facto Standards

Standards originating from a single commercial entity, formally termed "proprietary specifications", are characterised by the complete control the developing firm retains over their technical content and future evolution. The development process for such specifications is typically closed, with the parent company dictating the terms of engagement for any external parties, if participation is permitted at all.

[197] Alexander Brem and Petra Nylund, 'The Inertia of Dominant Designs in Technological Innovation: An Ecosystem View of Standardization' (2024) 71 IEEE Transactions on Engineering Management 2640 <https://doi.org/10.1109/TEM.2022.3192094>.

This centralised control confers distinct advantages; proprietary specifications can be brought to market with greater celerity than their consensus-based counterparts, which must accommodate a multitude of competing interests. Moreover, they can be strategically optimised to serve the specific commercial objectives of the originating firm. Should a proprietary specification achieve widespread market penetration, facilitated by the owner's decision to license the requisite intellectual property, it can become a significant source of revenue or confer other strategic benefits. It is crucial to underscore that such specifications are fundamentally different from collaboratively developed standards, and the owners of any associated patents are not bound by the specific licensing commitments, such as RAND terms, that are a hallmark of many formal standardisation processes.

Given that standardisation is primarily a voluntary activity, any company or collective of companies has the latitude to develop and promulgate its own technical specification. It should be noted that if a group of companies collaborates on this effort, the resulting body may be more accurately classified as a forum or consortium, a category to be discussed subsequently. When a publicly available proprietary specification achieves a high degree of market adoption, it is often referred to as a "de facto standard". The emergence of a de facto standard is a potential outcome of informal standardisation, where market success, rather than formal approval by a designated entity, serves as the arbiter of a technology's canonical status.

Upon developing a specification, a company faces a strategic decision: either to actively promote its adoption by others or to maintain it as a guarded asset[198]. The former path is chosen when the firm anticipates greater benefits from fostering a broad, interoperable ecosystem. A larger market platform can attract complementary innovations in devices, software, or content, thereby increasing the value of the original technology for all participants, including the innovator. A classic illustration is the VHS developed by Japan Victor[199]. The company astutely recognised that its interests were best served by promoting the VHS format not only among competing hardware manufacturers but also to tape producers and film studios, whose content was essential for the platform's success. Conversely, a firm may opt to restrict its specification if it believes it is positioned to serve the entire market exclusively[200]. This strategy is observable in the market for computer printers, where manufacturers often attempt to discourage adoption of their proprietary command languages. The legal capacity to prevent such adoption, however, is contingent upon the ownership of essential intellectual property rights and the decision not to license them to third parties.

When a proprietary specification achieves significant market traction, especially in a domain conducive to international standardisation, it is not uncommon for the original developer to transfer stewardship to a formal SDO. In doing so, the specification gains the elevated status of a formal standard, and the SDO assumes responsibility for its maintenance and further development. This transition provides a robust platform for greater global dissemination and adoption, though it necessitates that the original firm relinquish its exclusive control. A prominent case is the PDF, developed by Adobe Systems. Though Adobe made the specification publicly available in 1993, PDF remained a proprietary technology until its submission to the ISO, which formalized it as ISO 32000-1 in 2008. This move made the format more palatable to government and archival institutions seeking a universal open document standard.

---

198 Peter Grindley and Peter Grindley, *Standards, Strategy, and Policy: Cases and Stories* (Oxford University Press 1995).

199 Michael A Cusumano, Yiorgos Mylonadis and Richard S Rosenbloom, 'Strategic Maneuvering and Mass-Market Dynamics: The Triumph of VHS over Beta' (1992) 66 Business History Review 51.

200 Stanley M Besen and Joseph Farrell, 'Choosing How to Compete: Strategies and Tactics in Standardization' (1994) 8 Journal of Economic Perspectives 117 <https://doi.org/10.1257/jep.8.2.117>.

Concurrently, Adobe issued a public patent license to ISO 32000-1, granting royalty-free rights to all of its patents necessary for implementing the standard[201].

4.1.2. Standards Development in Formal SDOs

Recognising the strategic importance of an open and accessible standards system, numerous national authorities have formally established or officially recognised specific standards bodies. These entities, generally known as formal SDOs, provide the institutional framework for cooperative standards development. These organisations are typically membership-driven, convening experts from diverse, and often competing, stakeholder groups including private industry, government, academia, and civil society. Participation models vary; some global SDOs, such as ETSI, permit direct membership from private-sector entities, whereas others, such as ISO, operate through a system of indirect participation. In the latter model, national SDOs, defined by ISO as "the national body most representative of standardisation in its country"[202], are delegated the responsibility of representing the full spectrum of national stakeholder interests.

Formal SDOs are distinguished by their adherence to well-defined procedural rules governing participation rights, consensus-based decision-making, public availability of specifications, and patent disclosure and licensing policies. It should be noted that the "open availability" of a standard does not necessarily imply it is free of charge, as many SDOs rely on the sale of standards documents as a primary funding mechanism. The finalisation of a standard typically involves a rigorous approval process administered by the membership or the organisation's secretariat, guided by the principle of consensus.

The preeminent international SDOs are the ISO, whose purview covers nearly all technical fields; the IEC, specializing in electrical and electronic technologies; and the ITU, which focuses on information and communication technologies. These three bodies collaborate through the World Standards Cooperation (WSC) to ensure the efficient coordination of their activities, with the explicit goal of providing cohesion to a myriad of national and regional standards, thereby harmonizing global best practices, eliminating technical barriers to trade and fostering shared socio-economic advance[203]. ISO and IEC also collaborate directly on ICT standardisation within the ISO/IEC Joint Technical Committee 1 (JTC1)[204].

At the regional level, the European Union officially recognises three pivotal SDOs: CEN, CENELEC, and ETSI[205]. CEN and CENELEC maintain close collaboration with international SDOs, such as ISO and IEC, through the frameworks of the Vienna Agreement[206] and the Frankfurt Agreement[207], respectively. At the national level, most countries possess a government-recognised SDO,

---

[201] 'History of PDF', *Wikipedia* (2025) <https://en.wikipedia.org/w/index.php?title=History_of_PDF&oldid=1316915536> accessed 28 December 2025.

[202] 'International Organization for Standardization (ISO)' (*Oxford Public International Law*) <https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e1887?d=%2F10.1093%2Flaw%3Aepil%2F9780199231690%2Flaw-9780199231690-e1887&p=emailA%2Fb.zRQKUG9aU> accessed 28 December 2025.

[203] 'Who We Are – WSC' <https://www.worldstandardscooperation.org/who-we-are/> accessed 28 December 2025.

[204] 'Home' (*JTC 1*) <https://jtc1info.org/> accessed 28 December 2025.

[205] 'Key Players in European Standardisation - Internal Market, Industry, Entrepreneurship and SMEs' <https://single-market-economy.ec.europa.eu/single-market/goods/european-standards/key-players-european-standardisation_en> accessed 28 December 2025.

[206] 'Agreement on Technical Co-Operation between ISO and CEN (Vienna Agreement)' <https://boss.cen.eu/media/CEN/ref/vienna_agreement.pdf>.

[207] 'CENELEC Guide 13. IEC-CENELEC Agreement on Common Planning of New Work and Parallel Voting (the Frankfurt Agreement)' <https://boss.cenelec.eu/media/Guides/CLC/13_cenelecguide13.pdf>.

with prominent examples including the BSI, the DIN in Germany, the AFNOR in France, and the UNI and UNINFO in Italy.

4.1.3. The Role of Forums and Consortia

A vast and dynamic segment of the standardisation ecosystem, particularly within the fast-paced ICT industry, is composed of forums, consortia, and other informal industry associations[208]. These entities are frequently established by a group of like-minded organisations under the premise that focused, informal cooperation can yield a technical consensus more rapidly than the formal SDO process. These organisations exist on a continuum between single-company proprietary development and the multi-stakeholder processes of formal SDOs. The scope of consortia varies widely; some are formed to develop a single, specific standard, while others are chartered for a longer lifespan to address a broader technology area. Their reach can be national, regional, or global. The sheer number of such entities is substantial; a well-known inventory of standards bodies maintained by Gesmer Updegrove LLP, currently lists over 1100 organisations active in ICT standardisation, the vast majority of which can be classified as consortia[209].

Consortia are not monolithic in their governance or accessibility. They differ significantly in their degree of exclusivity. Some operate with a high degree of openness, welcoming broad participation and satisfying many of the criteria typically associated with open standards. Other consortia are considerably more exclusive, restricting participation to invited members, holding closed proceedings, or establishing stringent membership criteria. Among the most influential and successful of these bodies are organisations that, while lacking the formal recognition of national authorities, otherwise operate with the scale and rigor of formal SDOs. These 'quasi-formal SDOs' include the IETF which is responsible for the foundational Internet Protocol suite (TCP/IP), and the World Wide Web Consortium (W3C), which stewards the core standards of the Web.

The output of these consortia includes some of the most ubiquitous technologies in modern society. The USB standard, for instance, was developed by a consortium of companies including Intel, Microsoft, and IBM to create a universal hardware interface, and its implementations now number in the billions[210]. The DVD standard was similarly developed and is maintained by a membership-driven association known as the DVD Forum[211]. Bluetooth technology, initially conceived by Ericsson, was advanced by the Bluetooth Special Interest Group consortium before its eventual standardisation by the IEEE[212].

4.2. Principles of formal standardisation

In contrast to proprietary specifications and the standards developed by industry consortia or other informal forums, formal SDOs, such as the ISO or CEN-CENELEC, produce de jure standards based on a set of well-defined principles. The provenance of these principles can be traced to the legal

[208] T Schoechle, 'Digital Enclosure: The Privatization of Standards and Standardization' *ESSDERC 2003. Proceedings of the 33rd European Solid-State Device Research - ESSDERC '03 (IEEE Cat. No. 03EX704)* (IEEE 2003) <https://doi.org/10.1109/SIIT.2003.1251210> accessed 28 December 2025.

[209] 'SSO List | ConsortiumInfo.Org' <https://consortiuminfo.org/sso-list/> accessed 28 December 2025.

[210] 'Front Page | USB-IF' <https://www.usb.org/> accessed 28 December 2025.

[211] 'DVD Forum's Mission' <https://web.archive.org/web/20241128004023/http://www.dvdforum.org/about-mission.htm> accessed 28 December 2025.

[212] 'Bluetooth Technology Website | The Official Website of Bluetooth Technology.' (*Bluetooth® Technology Website*, 19 December 2025) <https://www.bluetooth.com/> accessed 28 December 2025.

framework of the World Trade Organisation (WTO), specifically the Agreement on Technical Barriers to Trade (TBT)[213]. While the TBT Agreement itself laid the groundwork, a pivotal moment in the articulation of these tenets occurred in the year 2000, when the TBT Technical Committee issued a formal decision-document that delineated some principle from internation standardisation. This document established six mandatory principles for their development: transparency, openness, impartiality and consensus, effectiveness and relevance, coherence and development dimension[214]. The principles under consideration have since become the essential criteria for defining which standards can be recognized as relevant within the international context.

4.2.1. The Six Foundational Principles of the WTO TBT Committee

The six principles promulgated by the WTO TBT Committee collectively form a framework intended to ensure that international standardisation facilitates, rather than impedes, global trade. These tenets are transparency, openness, impartiality and consensus, effectiveness and relevance, coherence, and a consideration of the development dimension. Each principle addresses a distinct facet of the standardisation process, yet they are deeply interconnected, creating a robust system of checks and balances. The major international standards bodies, including ISO, have thoroughly integrated these principles into their procedural frameworks, and documents such as ISO/IEC Guide 59 provide detailed recommendations for their implementation at the national level[215].

The principle of transparency mandates that sufficient and regularly updated information regarding standardisation activities be made easily accessible to all stakeholders in a timely manner. This requirement is operationalised through several concrete practices. It necessitates the publication of a notice at an early appropriate stage, informing interested parties of a proposal to develop a particular standard. The work programmes of standards bodies must be made publicly available, containing information on the standards currently being prepared and those recently adopted. A crucial aspect of transparency is the open consultation phase, where a draft standard is made available for public comment for an adequate period, typically a minimum of sixty days, allowing for meaningful feedback to be collected and considered. Furthermore, upon adoption, the final standard must be published promptly.

Complementing transparency is the principle of openness, which concerns the rights of participation in the standards development process. Openness requires that participation be available on a non-discriminatory basis to all relevant stakeholders at every stage of development. This encompasses the proposal and acceptance of new work items, technical discussions on their content, the submission and resolution of comments on drafts, the review of existing standards, and the voting and formal adoption of standards. At the national level, this implies that an NSB should not impose undue membership barriers and must actively facilitate the participation of a diverse range of stakeholders, including potentially under-represented groups such as SMEs and representatives of societal interests.

The third principle is a dual concept: impartiality and consensus. Impartiality demands that the standardisation process afford no privilege or unfair advantage to the interests of any particular supplier, country, or region participating in the standardisation process. It must be manifest throughout the entire process, from access to participation and consideration of comments to decision-making and the

[213] World Trade Organization, *Technical Barriers to Trade: The WTO Agreement Series* (WTO 2014) <https://doi.org/10.30875/7c1b8b12-en> accessed 3 August 2023.

[214] 'Principles for the Development of International Standards, Guides and Recommendations' <https://www.wto.org/english/news_e/news19_e/ddgaw_05dec19_e.htm> accessed 28 December 2025.

[215] 'ISO/IEC Guide 59:2019 - ISO and IEC Recommended Practices for Standardization by National Bodies' <https://www.iso.org/standard/71917.html> accessed 28 December 2025.

dissemination of documents. Impartiality is the procedural foundation upon which consensus, the substantive goal, is built. Consensus, as defined in ISO/IEC Guide 2, is a general agreement characterised by the absence of sustained opposition to substantial issues by any important part of the concerned interests[216]. As we shall analyse in section 4.3 of this Chapter, it is a process that necessitates reconciling conflicting arguments and taking all views into account; it does not, however, imply unanimity. To this end, leadership within a technical committee is tasked with impartially assessing whether expressed objections constitute "sustained opposition" and managing the process to either resolve such objections or, if a resolution is not possible without compromising the existing level of agreement, to register the opposition and proceed.

The principles of effectiveness and relevance ensure that standards meet real-world needs. International standards must effectively respond to regulatory and market demands, as well as scientific and technological developments, without distorting global markets or stifling innovation. Standards should be market-driven, addressing needs identified by industry, regulators, and consumers. To maintain their relevance, they must be developed and published in a timely manner, as a standard that appears long after the market has moved on is of little value. This challenge is particularly acute in rapidly evolving technology sectors, where time-to-market is a critical parameter. Moreover, relevance requires that standards, whenever possible, express requirements in terms of performance rather than prescriptive design characteristics, thereby allowing maximum freedom for technical innovation. This is supported by systematic review procedures, typically on a five-year cycle, aimed at identifying and revising or withdrawing standards that have become obsolete.

Coherence addresses the need for a logical and consistent standards landscape. It is essential that international standardizing bodies avoid duplication of, or overlap with, the work of other such bodies to prevent the development of conflicting standards. This requirement has become increasingly critical as technologies converge, demanding closer cooperation between previously distinct domains. To achieve coherence, standards bodies engage in active coordination, establishing memoranda of understanding and pursuing joint work, as seen in the ICT sector. At the national level, an NSB must conduct a thorough study before initiating any new project to determine if complementary or duplicative work already exists at the national, regional, or international level.

Finally, the development dimension acknowledges the constraints that developing countries face in participating effectively in international standardisation. This principle advocates that their concerns be explicitly taken into consideration. Tangible ways of facilitating their participation, such as through technical assistance and capacity-building programmes, should be sought to ensure they are not de facto excluded from the process. Organisations like ISO have institutionalized this principle through policy committees such as DEVCO, which partners a developing country's NSB with a more experienced counterpart for mentorship and knowledge sharing[217].

4.3. The road to consensus: the standardisation process in formal SDOs

The principles established in the preceding discussion constitute the normative foundation upon which formal standardization operates during the development of any given technical specification. This developmental trajectory is not a monolithic event but is instead structured through a series of rigorously defined phases that ensure both procedural integrity and technical consensus. To analyse the mechanics

[216] 'ISO/IEC Guide 2:2004 - Standardization and Related Activities — General Vocabulary' (n 182) 2.

[217] 'ISO - A Committee to Support Developing Countries: DEVCO' (*ISO*) <https://www.iso.org/devco.html> accessed 28 December 2025.

of this progression, we examine the process governing the creation of an International Standard in ISO, as a procedural archetype, effectively encapsulating the essential operational workflows and regulatory requirements that are mirrored across the broader ecosystem of international, regional, and national SDOs. Because ISO's procedures are designed to harmonise diverse technical and geopolitical interests, they provide a blueprint for the consensus-building mechanisms that define the legitimacy of formal standards globally. Consequently, an analysis of the ISO lifecycle offers more than a singular institutional perspective; it clarifies the stages of standard-setting, from the initial proposal and preparatory phases to the committee, inquiry, and final approval stages, that are fundamental to the maintenance of technical consistency and market relevance in a multi-layered governance structure.

The development trajectory for international standards within the ISO constitutes a meticulously structured, multi-phase process expressly engineered to ensure the resultant documents attain the requisite levels of clarity, precision, and unambiguousness, while simultaneously fulfilling stringent criteria for global relevance. This procedural architecture is rigorously governed by the ISO/IEC Directives Part 1[218] and Part 2[219] which collectively delineate the operational and linguistic parameters for standardisation activities. Each standardisation project is assigned to one of several predefined development tracks—18, 24, 36, or 48 months—dictating the timeframe for progression through the six principal stages.

The foundational Proposal Stage (Stage 10) is obligatory, serving the critical function of validating the necessity for a new International Standard within a specific domain, guided by the imperative of demonstrable global relevance. This validation is initiated through the submission of a New Work Item Proposal (NWIP), formally designated as an NP, utilizing Form 4 via the designated electronic balloting portal to the pertinent Technical Committee (TC) or Subcommittee (SC) for formal voting Crucially, the nominated Project Leader is identified within this submission, assuming responsibility for steering the nascent project. Concurrently, potential complications pertaining to intellectual property rights, including copyright and patents, as well as conformity assessment implications, necessitate early and thorough scrutiny to pre-empt future impediments.

Progressing to the Preparatory Stage (Stage 20), the parent committee typically mandates the formation of a dedicated Working Group (WG), composed of subject-matter experts and convened by an individual usually serving concurrently as the Project Leader. The primary objective within this phase is the meticulous preparation of the Working Draft (WD). The finalised WD is subsequently elevated to the parent committee, which exercises discretion in determining whether advancement proceeds directly to the Enquiry Stage or necessitates intermediate scrutiny within the Committee Stage. The ISO/TC platform provides a dedicated infrastructure for collaborative document sharing and refinement during this preparatory phase.

The Committee Stage (Stage 30) presents as an optional phase, the applicability of which is guided by specific criteria outlined in Annex SS of the ISO/IEC Directives Part 1. When invoked, its purpose is to disseminate the Working Group's draft, now formalized as a Committee Draft (CD), to the broader membership of the parent TC or SC for comprehensive review. Utilizing the Electronic Balloting Portal, committee members submit technical comments on the CD. This stage may entail multiple iterations of CD circulation, each followed by comment integration, until a robust consensus on the technical substance is secured. Within the ISO framework, consensus is explicitly defined as general agreement characterised by the absence of sustained opposition to substantial issues from significant concerned interests, achieved through a deliberative process that seeks to reconcile divergent viewpoints and

[218] 'Iso.Org/Sites/Directives/Current/Consolidated/Index.Html' <https://www.iso.org/sites/directives/current/consolidated/index.html> accessed 28 December 2025.

[219] 'ISO/IEC Directives Part 2' (n 185) 2.

conflicting arguments; it is emphasised that this does not necessitate unanimity but rather reflects a resolution that accommodates the core concerns of all relevant parties.

The Enquiry Stage (Stage 40) constitutes an obligatory milestone. At this juncture, the Draft International Standard (DIS) is formally submitted to the ISO Central Secretariat (ISO/CS), which orchestrates its circulation to all ISO member bodies for a mandated 12-week voting and comment period. Ratification of the DIS demands that two-thirds of the P-members (participating members) of the originating TC/SC vote affirmatively, coupled with the stipulation that not more than one-quarter of the total votes cast are negative. Should the DIS secure approval without necessitating substantive technical modifications, the project transitions directly to publication. However, the introduction of technical changes in response to comments received during the Enquiry Stage renders progression to the subsequent Approval Stage mandatory. Drafts at this stage are typically accessible for public commentary through national member bodies.

The Approval Stage (Stage 50) is conditionally obligatory. It is automatically obviated if the DIS garnered approval in the preceding stage without requiring technical amendments. Conversely, the incorporation of technical changes subsequent to the DIS vote necessitates this stage, making it mandatory under those circumstances. The Final Draft International Standard (FDIS) is submitted to ISO/CS and circulated to all ISO members for a condensed 8-week voting period. Approval mirrors the criteria of the Enquiry Stage: a two-thirds majority of P-members in favour and no more than one-quarter negative votes from the total cast.

Conclusively, the Publication Stage (Stage 60) is obligatory. The project secretary submits the finalised document for publication via the designated Submission Interface. If the project traversed the Approval Stage, the submission may include the Project Leader's formal responses to member body comments on the FDIS. Crucially, only editorial refinements are permissible at this terminal phase; no technical alterations are allowed. The ISO Central Secretariat undertakes the formal publication of the document as an International Standard. A final two-week sign-off period is allocated to Committee Managers and Project Leaders prior to official release.

Underpinning the entire developmental continuum are rigorous General Drafting Guidelines, demanding unwavering attention to clarity, precision, and unambiguous expression, as also mentioned in section 3.2 of the present Chapter. This necessitates the employment of Plain English, characterised by concise sentences, frequent subclause headings, and structured presentation to enhance comprehension and minimise translational errors. A direct, active, and impersonal tone is mandated, eschewing personal pronouns and extraneous detail. Oxford English spelling conventions, specifically the use of the -ize suffix, must be applied consistently. Terminological consistency is paramount, and the inclusion of legal or regulatory stipulations within the standard's content is proscribed. Submitted drafts must represent complete, finalised versions devoid of tracked changes. Colour must not be the sole mechanism for conveying information in any element, including figures and tables, to ensure accessibility and legibility across diverse contexts. Tables require specific formatting: they must not be submitted as images, should avoid nesting or vertical text orientation, and generally eschew shading for content delineation. Mathematical expressions, particularly complex formulations, should be displayed on separate lines and preferably numbered. Graphics must be submitted as revisable vector files. Stringently, images or text generated by artificial intelligence tools are explicitly prohibited from use in any ISO content, internal or external[220].

[220] ISO, 'Guidance on Use of Artificial Intelligence (AI) for ISO Committees' <https://www.iso.org/files/live/sites/isoorg/files/developing_standards/who_develops_standards/docs/use%20of%20AI.pdf> accessed 28 December 2025.

Beyond International Standards (IS), ISO produces supplementary deliverables, including Technical Specifications (TS), Publicly Available Specifications (PAS), Technical Reports (TR), and Guides. TS, PAS, and TR bypass the Enquiry Stage vote, requiring solely committee approval for publication, whereas Guides are subjected to the Enquiry Stage voting process. Furthermore, a systematic review is mandated every five years post-publication to evaluate the standard's ongoing utility, adoption, and relevance, culminating in a formal decision to confirm the standard without change, initiate its revision, or proceed with withdrawal.

5. Enforcing standards via conformity assessment procedures

5.1. An introduction to conformity assessment

While the adoption of technical standards is fundamentally characterized by its voluntary nature, they are nonetheless subject to rigorous enforcement processes designed to verify the correct implementation of their requirements. This is the conformity assessment, which is formally defined as the demonstration that specified requirements relating to a product, process, system, person, or body are fulfilled[221]. Fundamentally, it encompasses the totality of activities undertaken to determine whether a given object of assessment, such as a product or service, meets the specific requirements typically codified in a standard. The principal objective of this rigorous process is the cultivation of trust and confidence within economic and regulatory ecosystems. By furnishing a systematic means for verifying conformity against a spectrum of expectations—including standards, regulations, and other technical specifications—it ensures that products, processes, and services deliver on their declared promises. It is crucial to note that this demonstration is thought as a neutral undertaking; the process is equally capable of concluding that specified requirements have not been fulfilled, thereby providing critical information for risk management and quality control.

The significance of conformity assessment is pervasive, underpinning the functional integrity of all modern economies by providing a structured basis for assurance to the users of products, services, and commodities[222]. This mechanism of trust-building is particularly instrumental in facilitating trade at both national and international levels. By obviating the need for every buyer to conduct direct verification of product specifications, it allows transactional decisions to be made on the basis of credible test reports and certificates issued by competent bodies. The global harmonisation of these procedures is a critical factor in overcoming what are known as "technical barriers to trade" thereby promoting greater market integration[223]. Beyond its commercial utility, conformity assessment serves vital public interests by establishing and monitoring adherence to requirements for public health, safety, and environmental protection. This, in turn, contributes to enhanced economic efficiency and provides crucial access to global markets for producers. For consumers, the presence of formal declarations, marks, or certificates of conformity generates confidence in their purchasing decisions. Concurrently, it offers manufacturers

[221] 'ISO - Conformity Assessment' <https://www.iso.org/conformity-assessment.html> accessed 30 December 2025.

[222] Arthur E Appleton, 'Conformity Assessment Procedures' *Research Handbook on the WTO and Technical Barriers to Trade* (Edward Elgar Publishing 2013) <https://www.elgaronline.com/edcollchap/edcoll/9780857936714/9780857936714.00008.xml> accessed 30 December 2025.

[223] However, standards and conformity assessment procedures can become non-technical barrier to trade themselves. See, for example, Sherry M Stephenson, 'Standards and Conformity Assessment as Nontariff Barriers to Trade' (Social Science Research Network, 1 May 1997) <https://papers.ssrn.com/abstract=597242> accessed 21 February 2025.

and service providers a robust framework for ensuring their offerings meet declared specifications, thereby mitigating the substantial costs associated with product failures and recalls[224].

The scope of conformity assessment is remarkably broad, extending to any "object of conformity assessment," a term that can denote a product, process, system, person, or body[225]. The application is highly versatile, encompassing tangible items such as products, materials, and installations, as well as more abstract entities like services, projects, data, designs, claims, and even entire organisations or any combination thereof. The operational context of conformity assessment is often characterised by the relationship between the entity performing the assessment and the object being assessed. This relationship delineates three primary categories of activity. First-party conformity assessment is conducted by the person or organisation that provides the object itself, such as a manufacturer or service provider. A successful first-party assessment typically culminates in a formal supplier declaration of conformity, which represents a self-attestation of compliance by the supplier of the object. Second-party conformity assessment is performed by a person or organisation that possesses a direct user interest in the object, for instance, a purchaser or a corporate client conducting an audit of a supplier. In contrast, third-party conformity assessment is carried out by a person or body that is demonstrably independent of both the provider of the object and any party with a user interest. This category includes specialized entities such as certification bodies, testing laboratories, and inspection bodies. The inherent impartiality of the third-party approach inspires a significantly higher degree of market trust and is frequently stipulated as a mandatory requirement for demonstrating compliance with legal and regulatory obligations[226].

Conceptually, the practice of conformity assessment can be deconstructed into a series of discrete yet interconnected functions, collectively designed to demonstrate the fulfilment of specified requirements[227]. This "functional approach" provides a standardised and logical framework for all assessment activities. The process begins with the selection function, a preparatory stage that involves the planning and gathering of all necessary information and inputs for the subsequent determination. This includes precisely defining the applicable standards, selecting representative samples for evaluation, choosing the most appropriate assessment procedures, such as specific testing or inspection methods, and identifying personnel with the requisite competence to execute the tasks. Following selection is the determination function, which comprises the core technical activities undertaken to develop complete information regarding the fulfilment or non-fulfilment of the specified requirements by the object of assessment. Common determination activities include testing, inspection, and auditing, each a specialized technique for generating objective evidence.

The outputs of the determination phase are then subjected to a tripartite concluding stage, beginning with review. The review function involves a critical consideration of the suitability, adequacy, and effectiveness of the preceding selection and determination activities and their results. Based on this comprehensive review, a formal decision is made, which constitutes a definitive conclusion as to whether the fulfillment of specified requirements has been demonstrated. The final step is attestation, which is the issuance of a formal statement, predicated on the positive decision, that the specified requirements have been fulfilled. This output is commonly referred to as a 'statement of conformity' and serves as the

[224] ISO, IEC and UNIDO, 'Building Trust. The Conformity Assessment Toolbox' <https://www.iso.org/publication/PUB100230.html>.

[225] 'ISO/IEC 17000:2020(En), Conformity Assessment — Vocabulary and General Principles' <https://www.iso.org/obp/ui/#iso:std:iso-iec:17000:ed-2:v2:en> accessed 30 December 2025.

[226] 'Types of Conformity Assessment' <https://www.iec.ch/conformity-assessment/types-conformity-assessment> accessed 30 December 2025.

[227] Wilfried Hesser, *Standardisation in Companies and Markets* (Helmut-Schmidt-Univ 2010).

tangible evidence of a successful assessment. These forms of attestation vary based on the party performing the assessment. A declaration represents a first-party attestation, whereas certification constitutes a third-party attestation related to an object of conformity assessment. A specialised form of third-party attestation is accreditation, which conveys a formal demonstration of a conformity assessment body's competence, impartiality, and consistent operation. For objects subject to ongoing production or continuous service delivery, a fourth function, surveillance, may be required. Surveillance involves the systematic and periodic iteration of conformity assessment activities to ensure the continued validity of the original statement of conformity over time. These subsequent activities may be of a reduced scope or differ from the initial assessment, tailored to monitor ongoing compliance efficiently.

The various technical activities that constitute the determination function are themselves highly specialised. Testing, for instance, is the determination of one or more characteristics of an object according to a defined procedure, a process governed by stringent standards such as ISO/IEC 17025 to ensure the competence of laboratories, the validation of methods, and the traceability of measurements[228]. Inspection, by contrast, is the examination of an object to determine its conformity with detailed or general requirements, a task that heavily relies on the professional judgment of the inspector and is structured by standards like ISO/IEC 17020[229]. Auditing is a systematic and independent process for obtaining evidence and evaluating it objectively against specified criteria, most commonly applied to management systems like those defined by ISO 9001 or ISO 14001. Newer, yet increasingly critical, determination activities include validation and verification, which confirm the plausibility of claims about future use and the truthfulness of claims about past events, respectively, with their governing principles outlined in ISO/IEC 17029[230].

These individual activities are operationalised within the framework of conformity assessment schemes and systems. A conformity assessment scheme provides the specific rules and procedures for a given assessment, identifying the object, the specified requirements, and the methodology to be used. A conformity assessment system, in turn, provides the overarching rules for managing related schemes, often at a national, regional, or international level[231]. The design of such schemes is typically risk-based; the greater the likelihood and consequences of non-conformity, the more rigorous the assessment required, often necessitating the involvement of independent third parties. The choice of scheme is critical, with different models offering varying levels of assurance. Product certification schemes, for example, range from simple type testing (Type 1a) to comprehensive models involving initial testing, factory quality system assessment, and ongoing surveillance of products from both the factory and the market (Type 5), as detailed in ISO/IEC 17067[232]. Similar structured schemes exist for the certification of management systems, personnel competence, and other objects of assessment.

At the heart of this entire structure are the conformity assessment bodies (CABs) that perform the determination activities, for example, the testing laboratories, inspection bodies, and certification bodies. The international community has developed a comprehensive suite of standards, known as the ISO/CASCO Toolbox, to ensure these bodies operate with technical competence, consistency, and

[228] 'ISO - ISO/IEC 17025 — Testing and Calibration Laboratories' <https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html> accessed 30 December 2025.

[229] 'ISO/IEC 17020:2012' (*ISO*) <https://www.iso.org/standard/52994.html> accessed 30 December 2025.

[230] 'ISO/IEC 17029:2019' (*ISO*) <https://www.iso.org/standard/29352.html> accessed 30 December 2025.

[231] ISO, 'Conformity Assessment - Basic Concepts' (*ISO*) <https://casco.iso.org/cms/render/live/en/sites/cascoregulators/home/conformity-assessment/basic-concepts.html> accessed 5 August 2024.

[232] 'ISO/IEC 17067:2013' (*ISO*) <https://www.iso.org/standard/55087.html> accessed 30 December 2025.

impartiality[233]. Standards as those mentioned before, such as ISO/IEC 17025 for laboratories, ISO/IEC 17065 for product certifiers, and ISO/IEC 17021-1 for management system certifiers provide the benchmarks against which a CAB's competence is judged. To provide the ultimate layer of confidence in this system, accreditation exists as a form of meta-assessment[234]. Accreditation is the third-party attestation related to a CAB, conveying a formal demonstration of its competence to perform specific conformity assessment activities. Accreditation bodies, which themselves must comply with the rigorous requirements of ISO/IEC 17011, sit at the apex of the trust pyramid[235]. They are frequently appointed by national governments and form the backbone of national quality infrastructure.

This integrated system of standards, activities and bodies forms a complex yet coherent infrastructure, also called "quality infrastructure", that allows for a robust and internationally recognised mechanism for managing risk and facilitating commerce[236]. The role of this infrastructure in the context of international trade is particularly salient[237]. The WTO TBT was established precisely to ensure that technical regulations, standards, and conformity assessment procedures do not create unnecessary obstacles to the free flow of goods between nations . The TBT Agreement actively encourages WTO members to develop and participate in international conformity assessment systems and to base their national requirements on international standards where they exist. It further promotes the principle of mutual acceptance, urging members to accept the results of conformity assessment procedures from other member countries, provided they offer an equivalent assurance of conformity. Within this paradigm, verified compliance with international guides and recommendations, particularly through the mechanism of accreditation, serves as a primary indication of a conformity assessment body's technical competence, thereby providing a credible foundation for such recognition. It is for this reason that international frameworks, such as the IEC CA Systems, are recommended by both the WTO and the United Nations Economic Commission for Europe (UNECE) as highly effective instruments for averting technical trade barriers and instilling confidence in the quality and safety of imported goods[238].

Beyond its function as a lubricant for global trade, conformity assessment is a cornerstone of modern regulatory practice. Governments and public authorities extensively utilize its tools to enforce national legislation pertaining to public health, safety, and environmental protection. In many jurisdictions, this involves the formal governmental appointment or designation of specific conformity assessment bodies (CABs) authorized to perform assessments against mandatory regulatory requirements. The European Union provides a highly developed and influential model of this regulatory use under the New Legislative Framework, which is analysed in details in Chapter 3[239].

---

233 'ISO/CASCO - Committee on Conformity Assessment' <https://www.iso.org/committee/54998/x/catalogue/> accessed 30 December 2025.

234 APD Dwivedi, 'Laboratory Accreditation: Evolution and Development' in Anuj Bhatnagar and others (eds), *Handbook of Quality System, Accreditation and Conformity Assessment* (Springer Nature 2025) <https://doi.org/10.1007/978-981-99-4637-2_26-1> accessed 20 November 2024.

235 'ISO/IEC 17011:2017 - Conformity Assessment — Requirements for Accreditation Bodies Accrediting Conformity Assessment Bodies' <https://www.iso.org/standard/67198.html> accessed 30 December 2025.

236 Qi Li and others, 'Industry Quality Infrastructure: A Review' *2019 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)* (2019) <https://doi.org/10.1109/QR2MSE46217.2019.9021243> accessed 30 December 2025.

237 National Research Council, Board on Science Policy Technology, and Economic and International Standards Committee Conformity Assessment, and US Trade Policy Project, *Standards, Conformity Assessment, and Trade: Into the 21st Century* (National Academies Press 1995).

238 UNECE, 'Standards for Sustainable Development' (2017) <https://unece.org/fileadmin/DAM/trade/wp6/documents/2017/Conference_Report.pdf>.

239 Helen Delaney, Helen Delaney and Rene Van De Zande, 'A Guide to EU Standards and Conformity Assessment' (0 edn, National Institute of Standards and Technology 2000) NIST SP 951 <https://doi.org/10.6028/NIST.SP.951> accessed 21 February 2025.

5.2. Epistemological challenges in conformity assessment

While the technical and regulatory frameworks governing conformity assessment are well-established, the epistemological dimensions of these processes remain significantly under-researched. This oversight is noteworthy, as conformity assessment provides a fertile ground for investigating the nature of knowledge derived from empirical measurement and the subsequent challenges of formulating definitive statements of truth within a framework of inherent uncertainty.

A foundational epistemological challenge permeating conformity assessment arises from the ontological and epistemic distinction between the "true value" of a characteristic and its "measured value."[240] This distinction acknowledges the inherent impossibility of obtaining a perfect representation of the measured parameter; every analytical result is inextricably bound by associated uncertainty. This fundamental limitation is formally captured by the relationship: *Measured Value = True Value + Measurement Error*. Crucially, both the absolute magnitude of the measurement error and the true value itself remain epistemologically inaccessible, constituting an irreducible epistemic gap. Consequently, the "space of measured values" exists as a distinct conceptual and practical domain, explicitly non-interchangeable with the "space of true values." Measured values, therefore, provide only probabilistic estimates of the true value, each intrinsically accompanied by an area of uncertainty. This axiomatic reality dictates that any knowledge claim regarding conformity – the judgment that a product, process, or system meets specified requirements – is inherently inferential and probabilistic, fundamentally precluding absolute certainty[241]. Modern frameworks, such as that articulated in JCGM 106:2012[242], explicitly embrace this probabilistic nature by encoding knowledge about a measurand's possible values through a Probability Density Function. This Function serves as a formal mathematical expression of the state of incomplete information and the associated uncertainty, marking a decisive departure from binary, deterministic interpretations of conformity.

The centrality of measurement uncertainty to this probabilistic paradigm inevitably introduces the risk of incorrect conformity decisions. A finite, quantifiable uncertainty means that any decision rule based on measured values carries an inherent probability of error. Scholarship rigorously delineates two primary categories of such decision risks[243]. The consumer's risk, defined as the probability of falsely accepting a non-conforming item or batch (false conformity), poses tangible threats, potentially compromising product efficacy, safety standards, or public health outcomes. Conversely, the producer's risk, representing the probability of falsely rejecting a conforming item or batch (false non-conformity), imposes direct economic burdens through unnecessary reprocessing, retesting, or disposal. The ISO/IEC Guide 98-4:2012 - Uncertainty of measurement - Part 4: Role of measurement uncertainty[244] further distinguishes between "specific risk" and "global risk." Specific risk quantifies the probability of an erroneous decision for a singular, specific entity based solely on its measured value, associated measurement uncertainty, and the applicable specification limits. Estimation techniques for specific risk often employ computational methods such as Monte Carlo simulation to propagate uncertainties through

---

[240] Stéphane Puydarrieux and others, 'Role of Measurement Uncertainty in Conformity Assessment' in Sandrine Gazal (ed), *19th International Congress of Metrology (CIM2019)* (EDP Sciences 2019) <https://doi.org/10.1051/metrology/201916003> accessed 23 February 2025.

[241] LR Pendrill, 'Using Measurement Uncertainty in Decision-Making and Conformity Assessment' (2014) 51 Metrologia S206 <https://doi.org/10.1088/0026-1394/51/4/S206>.

[242] 'JCGM106-2012' (*BIPM*) <https://www.bipm.org/en/doi/10.59161/jcgm106-2012> accessed 30 December 2025.

[243] Guilherme Lucarelli Orsay, Khrissy Aracélly Reis Medeiros and Elcio Cruz De Oliveira, 'Use of Uncertainty Information in Conformity Assessment in thePharmaceutical Industry' (2023) 19 Current Pharmaceutical Analysis 673 <https://doi.org/10.2174/0115734129262343231020105935>.

[244] 'ISO/IEC Guide 98-4:2012' (*ISO*) <https://www.iso.org/standard/50465.html> accessed 4 August 2025.

the decision rule. Global risk, however, addresses the expected long-term proportion of entities within a defined population subject to decision errors, either non-conforming items erroneously accepted or conforming items erroneously rejected. This concept is paramount in operational contexts like clinical laboratories or high-volume calibration services, where predicting the average rate of incorrect decisions over time is essential for quality management. Estimating global risk necessitates the integration of prior knowledge concerning the population's historical distribution (e.g., mean and standard deviation) with the measurement uncertainty model, typically utilizing Bayesian inference to derive a posterior distribution predictive of future error rates.

Methodologically, managing these risks often involves the strategic implementation of guard bands[245]. Guard bands constitute a safeguard mechanism whereby specification limits are systematically adjusted inwards (for acceptance) or outwards (for rejection) to create modified acceptance and rejection zones. The size of the guard band is deliberately chosen to control specific risks, for instance, ensuring that the consumer's risk (probability of false acceptance) remains below a predefined maximum admissible threshold, commonly set at 5% in regulated industries. The complexity of risk management escalates significantly in multiparameter conformity assessment scenarios[246]. Here, the total risk of a false conformity decision, where the item is accepted despite failing to meet requirements on one or more parameters, can substantially exceed the maximum admissible risk level, even if each individual parameter measurement suggests compliance when considered in isolation. This amplification stems from combinatorial probabilities and potential parameter interactions. Addressing this challenge requires the development of multivariate acceptance limits. These limits are typically derived using sophisticated computational approaches, such as Monte Carlo simulation incorporating appropriate multivariate coverage factors, which account for the joint probability distribution of the parameters. Critically, the intrinsic correlation between the measured values of different parameters, whether arising from the nature of the item itself (intrinsic correlation) or from shared elements of the measurement process (metrological correlation), must be explicitly considered, as such dependencies exert a profound influence on the calculated total risk. Beyond the probabilistic quantification of risk percentages, contemporary scholarship underscores the imperative to quantify the tangible economic costs and broader impacts associated with incorrect decisions[247]. Truly rational decision-making frameworks must integrate these consequentialist considerations. An outcome that appears statistically equitable in terms of risk percentages may prove economically unfair when the potentially asymmetric costs of a false acceptance versus a false rejection are factored in. Consequently, optimising the measurement process itself, including its inherent uncertainty, becomes an economic exercise aimed at minimising the sum of the costs associated with performing the test and the expected costs arising from the consequences of incorrect decisions.

Modern probabilistic conformity assessment frameworks explicitly acknowledge the critical role of "measurand prior knowledge" as a fundamental epistemic component alongside the immediate measurement result and its associated uncertainty[248]. This prior knowledge encompasses *any* existing, relevant information pertaining to the true value of the measurand available *before* the current

[245] Luciana Separovic and others, 'Measurement Uncertainty and Conformity Assessment Applied to Drug and Medicine Analyses – A Review' (2023) 53 Critical Reviews in Analytical Chemistry 123 <https://doi.org/10.1080/10408347.2021.1940086>.

[246] Elcio Cruz De Oliveira and Felipe Rebello Lourenço, 'Risk of False Conformity Assessment Applied to Automotive Fuel Analysis: A Multiparameter Approach' (2021) 263 Chemosphere 128265 <https://doi.org/10.1016/j.chemosphere.2020.128265>.

[247] Pendrill (n 241).

[248] Puydarrieux and others (n 240).

measurement is undertaken. Within the Bayesian statistical paradigm, this prior knowledge is formally quantified as a prior probability distribution and subsequently combined with the likelihood function derived from the new measurement data. This synthesis yields a posterior probability distribution, representing the updated state of knowledge regarding the measurand's true value. The epistemological gain inherent in this process is significant: the posterior distribution typically exhibits reduced uncertainty compared to either the prior distribution or the distribution implied by the measurement result alone. This reduction stems from the synergistic integration of historical information with fresh empirical evidence, resulting in a more robust and informed probabilistic basis for conformity decisions. However, a substantial epistemological and practical challenge resides precisely in the quantification and formalization of this prior knowledge into statistically valid prior distributions. Transforming often heterogeneous, qualitative, or incomplete historical data, such as analyses of control charts from production processes, expected values derived from validated computational models, supplier-provided specifications, or historical calibration records for instruments, into rigorous probabilistic priors demands sophisticated judgment and introduces its own layer of methodological complexity. Crucially, the incorporation of prior knowledge is indispensable for the robust estimation of global risk, as it enables the fusion of historical data characterising the population's variability with the current measurement uncertainty model. This integration allows for predictive insights into the probability of erroneous conformity judgments (false acceptance or false rejection) for items yet to be tested, thereby informing broader quality management strategies.

Concurrent with the emphasis on prior knowledge, a significant epistemological evolution within conformity assessment scholarship concerns the conceptualization of sampling[249]. This evolution is marked by a deliberate and critical shift away from the historically prevalent term "representative sample" towards the advocated construct of the "appropriate sample." The critique levied against "representative sample" is fundamentally epistemological: it is deemed "not appropriate for use in the modern context of data quality and decision-making"[250] precisely because it primarily signifies the *intention* of the sampler rather than constituting an objectively verifiable and quantifiable *property* of the sample itself. This terminology obscures a profound difficulty in establishing whether any given sample genuinely embodies the properties of the whole population or merely reflects the sampler's aspiration that it should. The ambiguity surrounding "representation" hinders rigorous uncertainty quantification essential for probabilistic conformity assessment. In contrast, the concept of an "appropriate sample" is defined pragmatically and functionally as sampling demonstrably "fit for its intended purpose" within the specific decision-making context of the assessment. This definition carries the critical epistemological requirement that a quantifiable estimate of the uncertainty component attributable *solely* to the sampling process must be assignable. This reframing explicitly acknowledges sampling not as a preliminary or separate activity, but as an intrinsic and non-separable stage of the overall measurement procedure. Consequently, the uncertainty introduced by sampling must be rigorously evaluated and incorporated into the total measurement uncertainty budget; neglecting this contribution fundamentally jeopardizes the reliability of any conformity decision predicated on that total uncertainty. From an epistemological standpoint, only sampling protocols adhering to strict randomness, where every potential increment of the sampling target has an equal probability of selection, can yield unbiased estimates to which meaningful sampling uncertainty can be attached. The observed variance among measured values obtained from successive random samples drawn from a heterogeneous target provides an unbiased empirical estimate

[249] Fernando Cordeiro and Michael H Ramsey, 'Sampling in the Context of Conformity Assessment' (2023) 28 Accreditation and Quality Assurance 181 <https://doi.org/10.1007/s00769-023-01542-1>.

[250] ibid.

of the squared standard uncertainty inherent in the sampling process itself. However, this variance-based estimate inherently excludes systematic effects, such as consistent biases introduced by a particular sampler's technique. Quantifying such sampler bias, potentially achievable through inter-sampler comparison studies, becomes necessary to refine the uncertainty estimate further. It is imperative to recognise that these uncertainty estimates, while crucial for decision-making, are themselves estimates derived from data, allowing for the calculation of confidence intervals bounding the likely location of the true sampling uncertainty. Finally, the epistemological framework dictates that any subsequent statement of conformity must explicitly reference the defined sampling target (e.g., the specific lot, batch, or defined population), not merely the individual laboratory sample derived from it, as the inference pertains to the entire target entity.

5.3. Conformity assessment between trade barriers and trade facilitation

The intricate and heterogeneous nature of conformity assessment (CA) procedures across diverse national jurisdictions frequently constitutes a substantial impediment to international trade by imposing significant costs upon exporters[251]. These costs can be considerable, sometimes representing between two and ten percent of a firm's overall production costs[252]. The mechanisms through which these procedures translate into trade barriers are multifaceted, with several recurring themes emerging from the specific trade concerns raised by WTO Members, as analysed by scholars[253].

A primary contention pertains to the principle of proportionality, specifically whether the stringency of a given CAP is commensurate with the level of risk associated with the product in question[254]. Members frequently challenge regulations that appear excessively strict for low-risk goods. A case in point is India's compulsory registration order for electronics and information technology goods, which the European Union characterised as "excessively burdensome" and "more strict than necessary" for products such as laptops and printers. The EU posited that a less onerous procedure, such as a supplier's declaration of conformity, would be more appropriate for the identified risk level. Similarly, the clarity of risk categorisation itself can become a point of contention. Concerns were raised regarding Colombia's conformity assessment framework, where ambiguity in the definition of "moderate" risk, and the criteria distinguishing it from other risk classifications, created uncertainty and potential for misapplication of unnecessarily stringent procedures[255].

Another significant source of trade friction stems from the duplication of procedures, a practice that arises from the non-acceptance of foreign test results or certifications and compels importers to reiterate costly and time-consuming evaluations in the destination market[256]. Such redundancies not only inflate costs for exporters but can also function as a de facto protectionist measure, insulating domestic enterprises from international competition. For example, Korea voiced concern over the United States Energy Star Program, which imposed supplementary accreditation requirements beyond those of established international frameworks like the ILAC Mutual Recognition Arrangement. This issue also

---

[251] 'Standards, Conformity Assessment, and Trade into the 21st Century'.

[252] Stephenson (n 223).

[253] Devin McDaniels and Marianna Karttunen, 'Trade, Testing and Toasters: Bringing Conformity Assessment Procedures into the Spotlight' (2016) 50 Journal of World Trade <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\TRAD\TRAD2016031.pdf> accessed 4 September 2024.

[254] ibid.

[255] ibid.

[256] ibid.

manifests at the intersection of national and regional regulatory frameworks. The potential for duplicative requirements between Saudi Arabia's national certificate of conformity for toys and the Gulf Cooperation Council Conformity Marking applied at the regional level prompted calls for a streamlined system wherein compliance with one set of requirements would suffice[257].

Beyond the broader issues of proportionality and duplication, the operational mechanics of CA procedures, specifically their required frequency, the volume of testing mandated, and the limited duration of certification validity, are frequently identified as burdensome[258]. The European Union, for instance, challenged Ecuador's stipulation that ceramic tiles undergo certification based on testing "every single lot or shipment," advocating for a more streamlined approach based on product type. Analogous concerns were voiced by Indonesia regarding a United States Environmental Protection Agency proposal for quarterly testing of wood products, which was perceived as an unnecessary imposition on producers. The sheer quantity of mandated tests has also been a point of contention, as exemplified by Korea's critique of Indonesia's steel product CAPs for their "large number of sampling tests." Furthermore, the temporal validity of certifications can present a significant hurdle; Australia's one-year renewal period for washing machine certifications under its WELS scheme was flagged as imposing a considerable and recurring administrative burden[259].

In addition to the implicit costs embedded within procedural design, WTO Members have also explicitly cited excessive delays and direct financial outlays as direct consequences of certain CAPs[260]. India, for example, characterised Brazil's Good Manufacturing Practice (GMP) certification for medical devices as both time consuming and expensive, with reported costs exceeding $20,000 and approval timelines extending up to four years[261]. This concern was echoed by Mexico, which highlighted prohibitive costs and delays in obtaining mandatory certifications for electronic appliances in Brazil, attributing the problem in part to a constrained domestic quality infrastructure. Thailand's certification regime for ceramic tiles drew multifaceted criticism for being "unnecessarily burdensome, redundant and costly," a perception exacerbated by requirements for in-country completion, certification validity limited per importer and product line, disclosure of confidential business information, and mandatory on-site factory inspections by Thai officials[262].

A pervasive and foundational challenge that compounds many of the aforementioned issues is the lack of transparency, a concern cited in over half of all CAP-related specific trade concerns[263]. This manifests in several forms, including the failure to notify new or amended CAPs in draft form, thereby precluding other members from exercising their right to comment, as the United States noted with respect to India's compulsory registration for toys. The use of urgent notification procedures, which bypass standard comment periods, as seen in the case of an Ecuadorian cosmetics regulation, also undermines transparency. Consequently, Members often use the multilateral forum of the TBT Committee to formally request clarification on technical requirements and to seek extended transition periods, in accordance with Article 5.9 of the TBT Agreement, to afford manufacturers sufficient time to adapt. The disruption caused by Turkey's imposition of new pharmaceutical import procedures with

[257] ibid.
[258] ibid.
[259] ibid.
[260] ibid.
[261] ibid.
[262] ibid.
[263] ibid.

a mere three-month compliance window underscores the commercial significance of such temporal considerations[264].

Perhaps the most contentious category of CA-related barriers involves allegations of discrimination, wherein procedures are perceived to afford less favourable treatment to imported goods compared to like domestic products[265]. Such discrimination can be explicit, applying only to foreign producers, or de facto, creating an uneven playing field through seemingly neutral rules. For example, security requirements for telecommunications equipment in India were questioned by the United States and Japan on the grounds that analogous testing and certification did not apply to domestic producers[266]. A similar discriminatory effect was identified in Ecuador's certification requirements for metal cable trays, where a national quality seal that exempted products from conformity certificates was available only to Ecuadorian manufacturers. Preferential treatment for domestic innovation, such as China's expedited approval process for medical devices featuring "Chinese indigenous innovation," has been criticized by the European Union as giving an unfair advantage to Chinese-manufactured devices. Korea's KCGMP requirements for cosmetics were similarly flagged for providing "certain facilities," such as exemptions from batch tests, exclusively to domestic firms[267]. Notably, such concerns are not unidirectional; China argued that an EU directive on active pharmaceutical ingredients imposed a disproportionately higher burden on importers compared to EU-based manufacturers. Even procedurally neutral rules can have discriminatory effects, as illustrated by a US requirement for testing lithium batteries transported by air, which was argued to disproportionately affect importers who rely heavily on air freight, while domestic producers using land transport were less impacted[268].

Addressing such complex challenges necessitates international cooperation, for which the WTO Agreement on Technical Barriers to Trade (TBT Agreement) provides the primary legal framework to balance national regulatory autonomy with the imperative to prevent unnecessary obstacles to trade. Among the most significant instruments for this purpose are Mutual Recognition Agreements (MRAs), which the TBT Agreement explicitly encourages Members to negotiate[269]. These agreements operationalise the principle of "tested once, accepted everywhere," a paradigm that allows manufacturers to conduct product testing in one jurisdiction and have the results and certification accepted across all signatory markets. The practical implementation of this principle, however, is contingent upon a high degree of mutual confidence in the respective parties' laboratory testing facilities, national accreditation programs, and overall procedural integrity. As we shall see in Chapter 5, the European Union's "global approach" to conformity assessment exemplifies a robust application of this concept; for products subject to mandatory technical regulations, a product originating outside the Union that demonstrates conformity with essential requirements would be subject to a single conformity assessment procedure for the entire EU market, supplanting what would have previously been fifteen distinct national tests. In contrast, while the North American Free Trade Agreement (NAFTA) similarly encouraged mutual recognition and the "equivalence" of procedures, its framework was less direct, ultimately leaving the final determination of compatibility to the discretion of the importing country. Further advancing this cooperative paradigm, regional bodies such as the Asia-Pacific Economic Cooperation (APEC) have

[264] ibid.
[265] ibid.
[266] ibid.
[267] ibid.
[268] ibid.
[269] Stephenson (n 223).

made significant strides by developing elements of a model MRA and prioritising specific product sectors, including toys and food safety, for its application[270].

The efficacy of these cooperative frameworks is contingent upon the institutional capacity of all Members, a consideration addressed through technical assistance programs. Initiatives such as those developed by the United States' National Institute of Standards and Technology (NIST) to provide standards-related assistance to key emerging markets are vital for building the institutional mechanisms necessary to comply with international agreements and facilitate trade[271]. Such assistance, often more forthcoming at a regional level, is particularly critical for improving standards systems in developing economies that may lack adequate infrastructure and human capital for laboratory testing and accreditation. Collectively, these cooperative endeavours represent a concerted effort to bridge the chasm between divergent national regulatory philosophies and the inherent uncertainties of metrology, which, as previously discussed, underpin the risks of incorrect conformity decisions.

6. Conclusion

This chapter has transitioned the inquiry from the broad rationales and challenges of technology regulation, established as the "why" and "how" of AI governance in Chapter 1, to a granular examination of technical standards, the specific instrument identified as the operational backbone of the AI Act's risk-based framework. By treating standards not merely as neutral technical specifications but as profound institutional frameworks, the discussion has mapped their evolution, their internal logic, and the intricate mechanisms through which they are developed and enforced.

The chapter opened with a historical contextualisation, tracing the trajectory of standardisation from early modular construction and the scientific rationalism of the metric system to the digital protocols of the internet era. Central to this analysis was the identification of an emerging "fourth wave" of standardisation. As argued, this current phase represents a paradigm shift wherein the remit of standards expands beyond narrow technical interoperability to encompass complex, value-laden governance processes such as risk management, ethical behaviour, and social responsibility. This evolution, however, introduces a critical tension: as standards move into these normative domains, they inevitably sacrifice the mathematical precision of the "first wave" for a degree of linguistic vagueness. This ambiguity, while allowing for contextual flexibility and the inclusion of a broader array of societal stakeholders, renders compliance more contestable and underscores the standard's role as a site of ongoing socio-political negotiation rather than a simple technical benchmark.

Following this historical mapping, the analysis turned to the "anatomy" of the standard itself, deconstructing the ISO definition and the structural components that constitute a standard as a "documentary" agent. It was shown that the formal structure of a standard, ranging from its scope and normative references to the performative distinction between mandatory "shall" requirements and advisory "should" recommendations, functions as a mechanism of epistemic governance. By establishing shared semantic frameworks and "optimum order," standards do more than facilitate trade; they construct the technical and social realities within which innovation occurs. This section underscored that the power of a standard lies in its ability to embed contested assumptions into a seemingly neutral, consensus-based document, thereby exerting influence over global supply chains and regulatory outcomes alike.

[270] ibid.

[271] 'Standards, Conformity Assessment, and Trade into the 21st Century' (n 251).

The discussion then delineated the diverse ecosystem of standard-setting entities and the rigorous principles that govern formal standardisation. By distinguishing between the proprietary specifications of single firms, the agile but often exclusive outputs of industry consortia, and the de jure standards produced by formal SDOs like ISO, CEN, and CENELEC, the chapter highlighted the varying degrees of legitimacy and authority inherent in different standardisation pathways. The analysis emphasised that the legitimacy of formal standards is predicated on the six foundational principles of the WTO TBT Agreement, i.e., transparency, openness, impartiality, effectiveness, coherence, and the development dimension. These principles ensure that the standardisation process remains a robust, consensus-driven endeavour, capable of balancing the competing interests of producers, users, and regulators in a globalized market.

Finally, the chapter examined the critical final link in the regulatory chain: conformity assessment. This process of verifying that specified requirements are fulfilled was revealed to be a fundamental exercise in trust-building, yet one fraught with epistemological and practical complexities. The analysis moved beyond the technicalities of testing and certification to explore the inherent uncertainty of measurement, highlighting the "epistemic gap" between true and measured values. In doing so, it was demonstrated that conformity assessment is never a binary determination of truth, but rather a probabilistic management of risk—balancing the consumer's risk of false acceptance against the producer's risk of false rejection. Furthermore, the discussion addressed the dual nature of these procedures in international trade, where they serve both as essential facilitators of market integration and as potential technical barriers to trade when proportionality and transparency are lacking.

In sum, this chapter has provided a comprehensive theoretical and practical mapping of the standardisation landscape. It has shown that standards are far more than manuals for machines; they are complex instruments of governance that manage uncertainty, facilitate global commerce, and bridge the gap between abstract legal requirements and practical technical implementation. However, having explored the mechanics and history of standards, the research must now confront their integration into the European regulatory framework for products and technologies, as discussed in the next Chapter 3.

Chapter 3.
Integrating technical standards for regulatory implementation in the EU

1. Aim of the Chapter

In Chapter 2, we examined the historical evolution, functional typologies, and procedural foundations of technical standards at a global level. Building on this broad conceptual framework, the present Chapter 3 pivots toward the specific European dimension, where standardisation occupies a role of unique salience within the regulatory landscape governing products and technologies.

This chapter prioritises the European standardisation system because it provides the institutional and legal blueprint upon which the AI Act is constructed. Within the European Union, technical standards are not merely voluntary benchmarks; they are integrated into a sophisticated regulatory architecture that emerged with the "New Approach" in the 1980s and evolved into the contemporary New Legislative Framework. As will be demonstrated, these standards serve as the essential bridge between high-level legislative mandates and the technical realities of the internal market, acting as a foundational pillar for the Union's digital and green transitions. Although primarily developed by private actors, the ESOs, these standards acquire a quasi-legal character when cited in the OJEU, making them central to issues of market access, industrial competitiveness, and public safety. Understanding this complex interplay between statutory law and technical specification is a necessary prerequisite for appreciating the regulatory choices underpinning AI governance in Europe.

The chapter is structured into four main sections. The first section explores the genesis and logic of the New Approach and the NLF, detailing the strategic division of labour between European legislators and technical experts. Following this, the second session provides an analysis of the formal legal framework established by the Standardisation Regulation (1025/2012), delineating the relationship between the European Commission and the ESOs. Next, the third section scrutinises the operational modalities of the system, with a particular focus on the mechanism of standardisation requests and the role of notified bodies in the conformity assessment process. Finally, the fourth section examines the contemporary challenges and unresolved questions facing the system, ranging from the legal status and accessibility of harmonised standards to persistent concerns regarding legitimacy, stakeholder participation, and the strategic objectives of the recent European Standardisation Strategy.

2. Evolution of the legal framework

2.1 The Old Approach

Prior to the introduction of technical standards as a tool for harmonizing technical legislation within the European single market, technical legislation in the countries of the Union was delegated to individual member states under what is commonly termed the "Old Approach." Driven by the ambition to complete a common internal market, this method relied primarily on detailed legislative directives to align national

technical regulations[272]. While aiming for free movement of goods, this approach proved cumbersome, fragmented, and ultimately fell short of its ambitious goals, setting the stage for subsequent reforms.

At its core, the old approach involved a product-by-product, or "vertical," harmonisation strategy. The 1969 General Program, for instance, envisioned adopting over a hundred directives for industrial products by the end of 1970, targeting sectors such as motor vehicles, agricultural tractors, machinery, and measuring instruments[273]. Each directive would define precise technical specifications that products had to meet. Two main methods underpinned this harmonisation: "total harmonisation" and "optional harmonisation"[274]. Total harmonisation aimed to completely replace divergent national regulations with uniform Community-wide specifications, requiring all products to conform to the directive to be marketed within the EC. This represented the most far-reaching form of harmonisation, demanding significant consensus at the Community level. Conversely, "optional harmonisation" allowed manufacturers the choice between adhering to national law or Community-level requirements. Products meeting the Community standards could circulate freely, but Member States retained the right to maintain their own divergent national specifications.

A key characteristic of the old approach was the inclusion of technically detailed regulations within the directives themselves[275]. This meant that EC legislation directly prescribed intricate technical specifications, a task that frequently proved unwieldy. For example, directives in the automotive sector often spanned hundreds of pages of technical details and testing instructions. This highly prescriptive nature had several significant drawbacks. It led to long preparatory periods and substantial delays in adoption, with directives sometimes taking an average of nine and a half years to be approved[276]. The detailed nature also meant that the directives were susceptible to becoming outdated quickly due to rapid technical progress, necessitating continuous adaptation. Moreover, the process overloaded the high-level political decision-making procedure with technical details making effective decision-making difficult[277].

The reliance on Article 100 EEC for harmonisation further exacerbated these issues due to its requirement for unanimity among Member States. This often resulted in directive content not being a well-balanced compromise but just a sum of all national ideas, reflecting the lowest common denominator and leading to less ambitious outcomes[278]. The problem was compounded by the fact that the scope of harmonisation efforts remained limited compared to the vast number of national technical standards. For instance, in 1984 alone, Germany and Britain issued thousands of national standards, while the Council adopted only 16 directives[279]. This disparity meant the Community struggled to keep pace with the proliferation of national regulations, leading to continued market fragmentation. Member States also retained the ability to impose "safeguard clauses," allowing them to temporarily prohibit or restrict the marketing of products, even if compliant with EC directives, if a hazard arose[280]. This mechanism, while intended to protect public health and safety, underscored the limitations of harmonisation and the ongoing assertion of national regulatory autonomy.

---

[272] Josef Falke and Christian Joerges, 'Traditional Law Approximation Policy Approaches to Removing Technical Barriers to Trade and Efforts at a Horizontal European Product Safety Policy' (2010) 6 Hanse Law Review 237.

[273] Publications Office of the European Union, 'General Programme of 28 May 1969 for the Elimination of Technical Barriers to Trade Which Result from Disparities between the Provisions Laid down by Law, Regulation or Administrative Action in Member States, CELEX1' (*Publications Office of the EU*, 28 May 1969) <https://op.europa.eu/en/publication-detail/-/publication/74d47cdf-6604-417a-83fc-9a0cb323a617/language-en> accessed 14 August 2025.

[274] Falke and Joerges (n 272).

[275] ibid.

[276] ibid.

[277] ibid.

[278] ibid.

[279] ibid.

[280] ibid.

By the mid-1980s, the criticisms of the old approach were undeniable. It had achieved only "fragmentary" results, with many sectors remaining outside its reach. The process of converting directives into national law was often delayed, and monitoring compliance proved challenging. In this way, the old approach to eliminating technical barriers, while ambitious in its intent, suffered from fundamental structural flaws. Its reliance on detailed, product-specific directives, the unanimity voting rule, and the persistent assertion of national safeguard measures resulted in a slow, inefficient, and often incomplete harmonisation process. This piecemeal progress, coupled with the overwhelming proliferation of national standards, demonstrated that a more flexible and efficient strategy was needed to truly achieve a barrier-free internal market, paving the way for the "new approach" that would address these inherent limitations.

2.2 The New Approach

In response to the manifest deficiencies of the Old approach, the European Community initiated a fundamental shift in strategy in 1985, commonly referred to as the "New approach" to technical harmonisation and standardisation. This revised strategy, predicated on a more flexible and pragmatic framework, sought to overcome the limitations of its predecessor by incorporating three key elements. Firstly, it aimed to proactively prevent the emergence of new technical barriers to trade through a mutual information directive, establishing a mechanism for early notification and consultation regarding proposed national technical regulations. Secondly, the New Approach embraced the principle of "reference to standards" within directives, marking a significant departure from the prescriptive detail of the traditional approach. Thirdly, it actively promoted European standardisation, particularly through recognised European standards organisations such as CEN, CENELEC and ETSI, fostering a more collaborative and decentralized approach to technical specifications[281].

The mutual information directive[282], a cornerstone of the New Approach, was strategically designed to mitigate the temporal gap between national regulatory actions and the slower pace of EC-wide harmonisation and standardisation. By mandating Member States to notify the Commission and other Member States of new draft national technical regulations and industrial standards prior to their enactment, the directive introduced a crucial element of transparency and early intervention. This notification process empowered the Commission to request a standstill period, typically up to one year, during which the proposed national measure would be suspended, providing an opportunity for the Commission to prepare an EC directive addressing the issue or to facilitate discussions and amendments aimed at minimising potential trade barriers.

The principle of "reference to standards" was established in the Council Resolution of 7 May 1985 on a new approach to technical harmonisation and standards[283] and represents a fundamental paradigm shift in the New Approach, aiming to simultaneously dismantle technical trade barriers and cultivate a cohesive European doctrine in areas of paramount public interest, such as safety, health, and consumer protection. While Member States retain their fundamental responsibility for safeguarding these objectives, the New approach seeks to circumvent trade impediments by harmonising the essential safety requirements that products must meet. This harmonisation is deliberately limited to the adoption of these essential requirements through directives (and later, regulations), leaving the formulation of detailed

[281] Jacques Pelkmans, 'The New Approach to Technical Harmonisation and Standardization' (1987) 25 JCMS: Journal of Common Market Studies 249.

[282] 'Directive - 83/189 - EN - EUR-Lex' <https://eur-lex.europa.eu/eli/dir/1983/189/oj/eng> accessed 14 August 2025.

[283] Council Resolution of 7 May 1985 on a new approach to technical harmonisation and standards 1985.

technical specifications to European standardisation organisations. This division of labour is critical: directives become less prescriptive and more strategically focused on outcomes, while standardisation bodies, possessing specialised technical expertise and industry representation, are entrusted with developing practical and readily adaptable standards. Compliance with these European standards, established through mandates from the Commission to CEN and CENELEC, and operating under the oversight of a Standing Committee, creates a presumption of conformity. This crucial legal construct signifies that products demonstrably compliant with relevant European standards are presumed to meet the essential requirements of the corresponding directive, thereby guaranteeing free market access throughout the Community[284]. Significantly, the New approach also retains a degree of flexibility for producers: it explicitly acknowledges that producers can still trade freely even if they choose not to adhere to European standards, provided they can demonstrate, through alternative means, that their products nevertheless fulfil the fundamental essential requirements. This provision for alternative routes to conformity recognises the potential for innovation and technological advancement beyond codified standards, and avoids unduly restricting market access to only those products that precisely adhere to standardised specifications.

The proactive promotion of European standardisation constitutes the third pillar of the New approach. The EC Commission actively fosters European standardisation through the issuance of mandates to CEN, CENELEC, and ETSI, ensuring a symbiotic relationship between standardisation processes and harmonisation efforts. These mandates explicitly direct standardisation activities towards the essential requirements defined in directives, thereby ensuring alignment between regulatory objectives and technical specifications.

The New Approach offered a range of potential advantages over the Old Approach. Perhaps most fundamentally, it promised greater coherence between the policy and legal regimes governing technical harmonisation, and a more effective conveyance between European harmonisation and standardisation processes. By combining total harmonisation of essential safety and health objectives with a more flexible and decentralised approach to the technical means of achieving these objectives through standardisation, the New Approach aimed to address the rigidities and inefficiencies of its predecessor. Improved information exchange, facilitated by the mutual information directive, further enhanced the preventative capacity of the new framework. Specifically, the New approach directly targeted and addressed many of the acknowledged drawbacks of the Old Approach. The time-consuming nature of directive adoption was expected to be mitigated by delegating detailed technical aspects to standardisation bodies and by accelerating the pace of European standardisation. The problem of excessive uniformity was addressed by decoupling the harmonised objectives from the variability of technical means, allowing for greater product diversity and innovation. A discernible increase in the tempo and volume of European standardisation was reasonably anticipated as a direct consequence of the new strategic focus and organisational reforms within standardisation bodies.

However, the New Approach also presented certain limitations and unresolved challenges. While it held considerable promise for improving the efficiency and effectiveness of technical harmonisation, certain critical areas remained inadequately addressed, or posed new potential difficulties. Notably, despite its focus on standards, the New approach initially offered little concrete guidance or initiative regarding testing and certification procedures, a significant omission given the increasing prevalence and importance of conformity assessment in ensuring market access. This lack of clarity on certification mechanisms represented a serious weakness that required prompt attention to avoid perpetuating trade

[284] Philippe Portalier, 'Myths and Realities of the Presumption of Conformity' <https://orgalim.eu/insights/myths-and-reality-presumption-conformity>.

barriers through divergent national certification requirements. For this reason, certification and conformity assessment procedures became a central focus of the NLF. Furthermore, the combination of harmonised objectives, European standards, and existing legal frameworks was expected to improve the application and implementation of directives within Member States, the practical challenges of transposition, enforcement, and consistent interpretation across diverse national legal systems remained significant. Finally, while the focus on total harmonisation of essential requirements and comprehensive product groups was anticipated to garner greater political interest from Member States, the inherent sensitivities surrounding national regulatory autonomy and the potential for resistance to Community-level interventions could still impede the successful adoption and implementation of directives[285].

Despite these recognised challenges and potential pitfalls, which have been later addressed in the New Legislative Framework, the New approach represented a qualitative advancement over the Old Approach. Its strategic emphasis on essential requirements, reliance on flexible standardisation mechanisms, and proactive information exchange mechanisms offered a more pragmatic and adaptable framework for achieving technical harmonisation. Furthermore, the New Approach directly addressed the fundamental objective of enhancing the competitiveness of European industry. Standardisation, when strategically aligned with quality assurance and international acceptance, demonstrably enhances market penetration and yields overall economic benefits[286]. Moreover, it can foster market creation and anticipatory effects, particularly in dynamic sectors such as telecommunications and information technology, by establishing common platforms and facilitating interoperability. The reduced workload on Commission services, resulting from the delegation of technical specification development to standardisation bodies, also allowed for a reallocation of resources to more pressing and strategic policy priorities.

3. The New Legislative Framework

The New Legislative Framework constitutes the latest development in the project of establishing a robust European internal market. The NLF represents a horizontal, general approach to harmonizing the conditions for product circulation, aiming to simultaneously ensure a high level of protection of public interests and facilitate the free movement of goods. At its core, the NLF is not a singular legislative instrument, but rather a cohesive structure built upon a constellation of legal acts. Among these, three legislative instruments stand out as particularly formative in defining the NLF's scope, operational mechanisms, and overarching objectives. These are Decision No 768/2008/EC of the European Parliament and of the Council on a common framework for the marketing of products[287], Regulation (EC) No 765/2008 of the European Parliament and of the Council setting out the requirements for accreditation and market surveillance relating to the marketing of products[288], and Regulation (EU) 2019/1020 of the European Parliament and of the Council on market surveillance and compliance of products[289]. These three acts, while distinct in their specific focus, collectively provide the essential building blocks for the NLF, contributing synergistically to its effectiveness and comprehensiveness.

[285] Pelkmans (n 281).

[286] Tassey (n 190).

[287] Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC (Text with EEA relevance) 2008 (OJ L).

[288] Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011 (Text with EEA relevance.) 2019 (OJ L).

[289] ibid.

Decision No 768/2008/EC functions as the conceptual and structural backbone of the NLF. Its primary contribution lies in establishing a set of common principles and reference provisions intended for application across all sectoral legislation pertaining to product marketing. This decision operates on a horizontal plane, providing a coherent template for the revision or recasting of existing directives and regulations within specific product sectors. Crucially, Decision 768/2008/EC introduces standardised definitions for key concepts, most notably for the various categories of economic operators involved in the product supply chain, including manufacturers, importers, distributors, and authorised representatives. By clarifying these roles and delineating their respective obligations, the decision addresses a critical ambiguity prevalent in earlier legislative instruments which often lacked precise definitions, leading to inconsistencies in interpretation and implementation across different sectors and Member States. Furthermore, this decision lays out a modular system of conformity assessment procedures, meticulously detailed in Annex II, offering a range of options tailored to different product types and risk profiles. This modular approach ensures that conformity assessment, a central mechanism for demonstrating product compliance, is both rigorous and appropriately scaled, avoiding undue burdens where risks are minimal while maintaining strict oversight for high-risk products. The decision also codifies the rules governing CE marking, a visible declaration by the manufacturer that a product conforms to the applicable requirements, and sets out the prerequisites for bodies tasked with conformity assessment to be notified to the European Commission, ensuring their competence and impartiality. In essence, Decision 768/2008/EC provides the common lexicon, the procedural framework, and the structural blueprint upon which sectoral product legislation within the NLF is constructed, thereby fostering consistency and coherence across the regulatory landscape.

Complementary to Decision 768/2008/EC's definitional and procedural contributions, Regulation (EC) No 765/2008 establishes the crucial mechanisms for accreditation and market surveillance within the NLF. This regulation directly addresses the operational imperative of ensuring that products circulating within the EU market not only comply with legislative requirements on paper but also in practice. Accreditation, as defined and regulated by this instrument, is the formal recognition that a conformity assessment body is competent to carry out specific conformity assessment tasks. By establishing harmonised requirements for accreditation bodies and their operation, Regulation 765/2008 seeks to guarantee the reliability and credibility of the conformity assessment process itself, thereby addressing one of the points that had remained open since the inception of the New Approach. However, the regulation's most impactful contribution arguably lies in its provisions concerning market surveillance. It mandates Member States to establish and implement systematic approaches to market surveillance, empowering them to monitor and verify product compliance after products have been placed on the market. This proactive element is critical for ensuring that the NLF is not merely a framework for ex-ante compliance but also incorporates robust ex-post control. Regulation 765/2008 outlines the powers and responsibilities of market surveillance authorities, enabling them to take appropriate measures when products are found to present a risk to public interests. These measures can range from requiring economic operators to take corrective action to restricting or even prohibiting the placing of non-compliant products on the market. The regulation explicitly aims to ensure a high level of protection of public interests, such as health and safety, while simultaneously ensuring that the fundamental principle of the free movement of products is not unduly restricted. This delicate balance underscores the inherent tension within the NLF, a tension the framework constantly seeks to navigate: the imperative to protect public interests while preserving the economic benefits of a seamless internal market.

Building upon the foundations laid by Decision 768/2008/EC and Regulation (EC) No 765/2008, Regulation (EU) 2019/1020 further refines and strengthens the market surveillance and product compliance aspects of the NLF. Enacted a decade after its predecessors, this regulation reflects the evolving landscape of product markets, particularly the growth of e-commerce and the increasing complexity of global supply chains. A key innovation introduced by this regulation is the reinforcement of market surveillance in the context of online sales. Recognising the increasing volume of products sold directly to consumers via online platforms, Regulation 2019/1020 introduces provisions to enhance the effectiveness of market surveillance in the digital sphere, addressing challenges related to identifying responsible economic operators and ensuring traceability of products sold online. Moreover, it explicitly strengthens cooperation and information exchange between market surveillance authorities in different Member States, acknowledging that effective market surveillance in a borderless internal market necessitates collaborative and coordinated action. By updating and expanding the market surveillance framework established by Regulation 765/2008, Regulation 2019/1020 contributes to ensuring that the NLF remains fit for purpose in the face of contemporary market realities, reinforcing its capacity to protect public interests and foster fair competition.

Collectively, Decision 768/2008/EC, Regulation (EC) No 765/2008, and Regulation (EU) 2019/1020 represent the tripartite legislative bedrock of the NLF. Decision 768/2008/EC provides the overarching structure, defining key terms, establishing procedural models, and setting the stage for sectoral legislation. Regulation 765/2008 operationalises the framework by establishing the crucial mechanisms of accreditation and market surveillance, ensuring ex-post compliance and enforcement. Regulation 2019/1020 adapts and strengthens these mechanisms to address contemporary challenges, particularly in the realm of online commerce. The impetus behind the development and implementation of the NLF stems from a multifaceted set of objectives. Firstly, consistency across sectoral legislation was paramount. Prior to the NLF, a patchwork of directives and regulations governed different product sectors, often leading to inconsistencies and overlaps, hindering clarity and increasing compliance costs for businesses operating across multiple sectors. Secondly, clarity was a central goal. Past legislation frequently employed terms without clear definitions, leading to divergent interpretations and implementation difficulties across Member States, thereby undermining the intended harmonisation. Thirdly, the NLF sought to firmly establish the responsibility of economic operators for ensuring product compliance, thereby shifting the focus from purely national regulatory control to a shared responsibility model, ultimately aiming to guarantee a uniformly high level of protection for public interests and fair competition within the entire Union market. Finally, and crucially, the NLF was designed with competitiveness in mind. By streamlining regulatory requirements, clarifying obligations, and promoting a consistent application of rules, the framework sought to avoid creating unnecessary burdens, particularly for small and medium-sized enterprises, while simultaneously maintaining the required level of protection for public health, safety, consumers, and the environment. This balancing act between regulatory efficiency and robust protection underscores the enduring complexity inherent in crafting effective market harmonisation legislation within a diverse and evolving economic and political landscape.

While the NLF represents a significant advancement towards a more coherent and effective regulatory environment for products in the EU, its ongoing success hinges upon consistent implementation across Member States, continuous adaptation to emerging market trends, and a sustained commitment to its core objectives of public interest protection and fair competition within a thriving internal market.

3.1. Blue Guide on the implementation of EU product rules

Within this intricate regulatory framework of the NLF, the "Blue Guide on the implementation of EU product rules"[290] emerges as a pivotal guidance document. The 2022 iteration of this guide, which updates the previous version dated 2016, serves as a crucial instrument intended to foster a more consistent and unified comprehension of EU product regulations spanning diverse sectors and throughout the single market.

At its core, the Blue Guide elucidates the multifaceted elements of the NLF and the mechanisms of market surveillance in exhaustive detail. Its overarching objective is to refine the understanding of existing legislation, thereby contributing to the formulation of legislation that is not only more comprehensive and coherent but also proportionate to the regulatory challenges it seeks to address. While the guide's immediate purview is primarily legislation concerning specific product categories, exemplified by fertilising products and unmanned aircraft systems, its relevance extends beyond these specific domains. Indeed, the Blue Guide explicitly states that fundamental elements, including definitions, standards, conformity assessment protocols, accreditation procedures, and market surveillance strategies, possess broader applicability. These foundational aspects are deemed pertinent to other areas of Union harmonisation legislation, signifying the Blue Guide's intended function as a horizontal instrument that transcends sector-specific nuances.

The target audience of the Blue Guide is deliberately broad, reflecting its ambition to serve as a central reference point for all stakeholders involved in the EU product regulatory ecosystem. It is explicitly addressed to Member States, signifying its role in guiding national implementation and enforcement. Furthermore, it extends its reach to a diverse array of parties deemed essential for the effective functioning of the product market, encompassing trade and consumer associations, standardisation bodies, manufacturers, importers, distributors, conformity assessment bodies, and even trade unions. This inclusive approach underscores the recognition that ensuring the free circulation of products and maintaining a high level of protection is a shared responsibility requiring concerted action across a spectrum of actors. The guide, in essence, attempts to bridge the knowledge gap across these diverse groups, fostering a common operational understanding.

Crucially, the Blue Guide operates within clearly defined parameters regarding its legal standing. It is unequivocally presented as a guidance document, explicitly stating that only the legally adopted text of the Union harmonisation act itself holds legal force. This caveat is further reinforced by the acknowledgment that discrepancies may exist between the provisions of a Union harmonisation act and the interpretative content of the guide. Moreover, the ultimate authority for the binding interpretation of EU legislation is unequivocally vested in the Court of Justice of the European Union. This series of disclaimers underscores a fundamental principle of EU law: guidance documents, however authoritative or comprehensive, cannot supersede the enacted legislative text or the jurisprudence of the Court. While this clarity is legally necessary and prudent, it simultaneously introduces a layer of complexity for users. Stakeholders must remain cognisant that the Blue Guide serves as an interpretive aid, not a definitive legal source. This necessitates a nuanced approach to its application, requiring users to consistently verify its guidance against the primary legal texts and to understand the potential for interpretative divergence. A critical perspective may question the degree to which this inherent non-binding nature might undermine the guide's practical effectiveness, particularly in situations where legal certainty is paramount.

[290] Commission notice The 'Blue Guide' on the implementation of EU product rules 2022 (Text with EEA relevance) 2022/C 247/01 2022.

The substantive content of the Blue Guide is structured around key themes central to EU product regulation. It offers a historical overview of the evolution of approaches to regulating the free movement of goods, tracing the trajectory from the Old Approach to the New Approach, as discussed in section 2.1 and 2.2 of this chapter. This historical contextualisation provides a valuable framework for understanding the rationale underpinning the current regulatory paradigm. The guide then delves into the intricacies of the 'New Legislative Framework' itself, which constitutes the cornerstone of modern EU product regulation.

A significant portion of the Blue Guide is dedicated to delineating the roles and obligations of the various actors within the product supply chain. This comprehensive mapping encompasses manufacturers, authorised representatives, importers, distributors, fulfilment service providers, and end-users. By clearly articulating the responsibilities incumbent upon each actor, the guide seeks to foster a culture of accountability and shared responsibility throughout the supply chain. This is particularly pertinent in increasingly complex and globalised supply chains, where ensuring product compliance necessitates the active engagement of multiple intermediaries. Furthermore, the Blue Guide rigorously outlines product requirements, encompassing essential requirements, traceability protocols, technical documentation standards, the EU declaration of conformity, and marking requirements. These elements represent the tangible benchmarks against which product compliance is assessed and verified, forming the operational core of the harmonisation process.

The processes of conformity assessment and accreditation are afforded significant attention within the Guide. It details various modules for conformity assessment, offering a structured approach to evaluating product adherence to regulatory requirements. The guide also clarifies the role and function of conformity assessment bodies, the entities entrusted with conducting these evaluations. Moreover, it elucidates the system of accreditation and the European accreditation infrastructure, highlighting the importance of ensuring the competence and impartiality of conformity assessment bodies. This emphasis on conformity assessment and accreditation underscores the EU's commitment to robust and verifiable mechanisms for ensuring product safety and compliance. Building upon this, the Blue Guide thoroughly addresses market surveillance, detailing the roles and responsibilities of market surveillance authorities in monitoring and enforcing product regulations within the market. This section is critical for understanding the practical mechanisms through which regulatory compliance is actively maintained and infringements are addressed post-market entry.

Finally, the Blue Guide extends its perspective to the international dimension of EU legislation on products. It examines agreements on Conformity Assessment and Acceptance of industrial products and mutual recognition agreements, highlighting the EU's engagement in international regulatory cooperation and the facilitation of trade through the recognition of conformity assessment procedures with third countries. This international outlook acknowledges the globalised nature of product supply chains and the need for international regulatory coherence to facilitate trade and ensure product safety on a global scale.

In conclusion, the Blue Guide on the implementation of EU product rules constitutes a comprehensive and indispensable resource for navigating the complexities of EU product regulation. Serving as a detailed exposition of the NLF and market surveillance mechanisms, it aims to enhance understanding, foster uniformity, and promote effective implementation across Member States and a wide array of stakeholders. While its non-legally binding status and selective scope necessitate a nuanced and critical approach to its application, the Blue Guide undeniably serves as a foundational instrument in ensuring the free movement of goods, maintaining a high level of protection, and facilitating the proper functioning of the EU internal market. Its continued relevance is underscored by its iterative revisions,

reflecting its ongoing adaptation to the evolving regulatory landscape and its enduring commitment to providing clarity and guidance in a complex and dynamic field. However, ongoing evaluation of its practical impact and effectiveness in achieving its stated goals remains crucial to ensure its continued efficacy as a cornerstone of EU product regulation.

4. The role of conformity assessment and notified bodies

4.1. Conformity assessment rules in the EU regulatory framework

The integrity of the harmonised regulatory framework within the European Community hinges upon demonstrable product conformity with essential requirements mandated by specific legislation and related standard. At the heart of this system lies the CE marking[291], a declaration by the responsible economic operator that a product aligns with the stipulated levels of protection deemed necessary for collective interests, encompassing public health, safety, and environmental preservation. This marking transcends a mere label; it serves as a visible testament that the manufacturer has diligently executed all requisite evaluation procedures as prescribed by Community law. Its presence on a product signals to market surveillance authorities and end-users alike that the item in question has navigated the necessary conformity assessment processes. Indeed, the CE marking operates as a critical juncture in the product's lifecycle, functioning as both an entry ticket to the unified European market and an ongoing pledge of sustained compliance. As we shall see in Chapter 4, such a critical role of CE marking can be found in the AI Act as well[292].

Conformity assessment, in this context, emerges as a pivotal mechanism employed by public authorities to rigorously ensure that products circulating within the market adhere to applicable requirements. This process is particularly concentrated on safeguarding user and consumer health and safety, reflecting the precautionary principle that underpins much of European regulatory philosophy. To operationalise this complex evaluative endeavour, a modular approach has been instituted. This system intelligently segments the conformity assessment process into discrete modules, meticulously covering both the design and production phases of products. This modularity is not arbitrary; it is deliberately structured to offer manufacturers a spectrum of choices when seeking to provide public authorities with assurances regarding the safety and conformity of their offerings. The legislation and related standards strategically enumerate the permissible modules applicable to a given product or sector, factoring in a nuanced constellation of considerations. These encompass the inherent type of product, the nature and magnitude of risks it potentially poses, the existing economic infrastructure of the sector, and the prevailing modes of production employed. This multi-faceted consideration ensures that the selection of conformity assessment procedures is contextually relevant and proportionate to the risks involved. In this manner, the European regulatory approach incorporates the conformity assessment mechanisms analysed in Section 5 of Chapter 2.

The main objective of the modular approach is to empower manufacturers with a range of options, thus acknowledging the heterogeneity of industrial structures and production methodologies across different sectors. Directives, therefore, meticulously define the conditions under which a manufacturer

[291] Grace Ballor, 'The Origins of CE Marking: Standards, Business, and the European Market in the 1980s–1990s' Harvard Business School - Working Paper 21-142 <https://www.hbs.edu/faculty/Pages/item.aspx?num=60460> accessed 20 June 2024.

[292] Victoria Hendrickx and Wannes Ooms, 'Commentary to Article 48 of the EU AI Act. CE Marking' (Social Science Research Network, 16 April 2024) <https://doi.org/10.2139/ssrn.4890939> accessed 21 February 2025.

is entitled to exercise discretion in module selection, always under the proviso that the chosen pathway effectively substantiates compliance with the essential requirements. This carefully calibrated flexibility is designed to avoid imposing overly burdensome modules that might disproportionately encumber economic operators relative to the overarching objectives of the directive. Such an approach recognises the imperative of fostering economic dynamism whilst upholding rigorous safety and regulatory standards. Within this flexible framework, judicious allowance is made for the adoption of supplementary modules or procedural variations, particularly when sector-specific exigencies or directive-specific nuances warrant such adaptations. This inherent adaptability underscores the nuanced and context-sensitive nature of the conformity assessment regime.

The modular system itself is articulated through a series of distinct modules, each representing a specific pathway to demonstrating conformity. Module A, termed "Internal Production Control", represents the most streamlined option, placing the onus squarely on the manufacturer. Under this module, the manufacturer assumes responsibility for ensuring and formally declaring that its products meet the requirements of the corresponding legislation. Crucially, the manufacturer affixes the CE marking and generates a written declaration of conformity, effectively self-certifying compliance. Module B, "EC Type-Examination", introduces third-party verification into the process. Here, a notified body, an organisation designated by a Member State and possessing recognised competence, steps in to conduct an "EC type-examination". This involves the notified body meticulously examining a product specimen to ascertain whether it is representative of the intended production and whether it demonstrably satisfies the directive's requirements. Upon successful examination, the notified body issues an EC type-examination certificate, providing formal attestation of conformity for the examined specimen. Module C, "Conformity to Type", builds upon Module B. The manufacturer operating under Module C must ensure that its production rigorously conforms to the type described in the EC type-examination certificate issued under Module B and, furthermore, that these products also independently meet the directive's stipulations. The manufacturer, in this instance, affixes the CE marking and issues a declaration of conformity, affirming ongoing adherence to the established type.

Modules D and E introduce Quality Assurance systems into the conformity assessment landscape. Module D, "Production Quality Assurance", mandates that the manufacturer operates an approved quality system specifically tailored for production, final product inspection, and testing. This system is subject to ongoing surveillance by a notified body, whose identification symbol accompanies the CE marking, signifying third-party oversight of the manufacturer's quality assurance processes. Module E, "Product Quality Assurance", similarly necessitates an approved quality system, albeit focused on final product inspection and testing rather than the entirety of the production process. Again, the CE marking in Module E is linked to the identification symbol of the notified body responsible for ongoing surveillance of this quality system. Module F, "Product Verification", adopts a more direct approach to conformity verification. The manufacturer, under Module F, diligently checks and attests that each product individually meets the directive's requirements. A notified body then undertakes appropriate examinations and tests, potentially on a statistical basis, to independently verify conformity. Module G, "Unit Verification", is applied to single, unique products. Here, the manufacturer ensures and formally declares that the individual product conforms to the directive's requirements, predicated upon a certificate issued by a notified body for that specific unit. Finally, Module H, "Full Quality Assurance", represents the most comprehensive approach. It requires the manufacturer to operate an approved quality system encompassing design, manufacture, and final product inspection and testing. As with Modules D and E, the CE marking under Module H is accompanied by the identification symbol of the notified body responsible for ongoing surveillance of this holistic quality system.

4.2. Notified bodies in the EU regulatory framework

This framework of conformity assessment under the European regulatory approach assigns a crucial role to notified bodies. These independent certification organisations, designated by a Member State's Competent Authority, function as crucial intermediaries in the conformity assessment process, tasked with evaluating whether products or systems meet the stringent requirements stipulated for market entry within the European Economic Area. The act of notification, therefore, signifies a formal conferral of authority upon a certification body, empowering it to conduct conformity assessment procedures under the aegis of a specific EU legislation. The sheer number of notified bodies, exceeding 1500 across the Union, underscores their pervasive influence and essential role in upholding product standards across diverse sectors[293].

The core function of a Notified Body resides in the provision of conformity assessment services mandated by EU legislation. These services are not merely procedural formalities; rather, they constitute a critical contribution to the overall assurance of product conformity. In fulfilling this mandate, Notified Bodies engage in a multifaceted approach that encompasses verifying products through rigorous sampling and testing methodologies, certifying and auditing the quality management systems implemented by manufacturers, meticulously assessing technical documentation to ensure comprehensive compliance, and critically auditing the operations of subcontractors deemed essential to the production process. This comprehensive suite of services is delivered in accordance with the specific directives governing the product in question, diligently considering all relevant documents and established industry standards.

The operational efficacy of Notified Bodies is intrinsically linked to a system of continuous oversight and evaluation by Competent Authorities. This dynamic relationship is maintained through ongoing communication and periodic audits, ensuring that Notified Bodies consistently adhere to the prescribed standards of competence and impartiality. Regular communication channels are established, facilitating the exchange of information pertaining to ongoing assessments, critical findings, issued certifications, and any significant organisational modifications within the Notified Body itself. Indeed, a proactive obligation rests upon Notified Bodies to promptly inform Competent Authorities of any substantial changes in their organisational structure or operational protocols, reflecting a commitment to transparency and accountability.

Furthermore, the operational integrity of Notified Bodies is underpinned by stringent personnel and resource management requirements. Central to their credibility is the necessity for personnel to possess demonstrably sufficient knowledge and practical experience directly relevant to the products and conformity assessment procedures they undertake. This expertise extends beyond mere product familiarity, encompassing sound technical training in pertinent regulatory requirements, active engagement with standardisation activities, in-depth understanding of relevant technologies and production methods, proficiency in verification procedures, and a nuanced comprehension of normal product usage conditions. Coupled with qualified personnel, Notified Bodies must exhibit robust resource management capabilities, demonstrating control and responsibility for the performance of all resources deployed, while meticulously maintaining comprehensive records documenting the suitability and qualifications of all staff employed. Crucially, access to appropriate facilities is paramount, necessitating the capacity to conduct testing and re-testing within the EU, thereby enabling the notifying authority to effectively verify their ongoing competence and operational capabilities.

[293] 'EUROPA – European Commission – Growth – Regulatory Policy - SMCS' <https://webgate.ec.europa.eu/single-market-compliance-space/notified-bodies> accessed 3 January 2026.

However, the operational framework of Notified Bodies is deliberately structured to preclude potential conflicts of interest and ensure unwavering impartiality. A fundamental tenet of their operation is the prohibition from assuming the role of manufacturer, authorized representative, supplier, or competitor to their clients. This stringent separation is designed to eliminate any potential bias that might arise from commercial affiliations. Extending this principle of impartiality, Notified Bodies are explicitly barred from providing consultancy services pertaining to products or quality systems that they are concurrently auditing, further safeguarding their objective assessment function. Moreover, the principle of independence is rigorously enforced, requiring all personnel to formally declare any potential conflicts of interest, thereby ensuring that economic considerations do not unduly influence decisions regarding the issuance of CE certificates. Confidentiality constitutes another cornerstone of their operational ethos, precluding the disclosure of client information or proprietary product/technology details to any entity other than the relevant Competent Authority, to whom reporting is explicitly mandated. Finally, the scope of their operational competence is strictly delineated, limiting their conformity assessment activities solely to those product categories and procedures for which they have been formally notified, preventing any unauthorized expansion of their remit.

Beyond these operational constraints, Notified Bodies are bound by a comprehensive set of roles and responsibilities that further define their function within the regulatory ecosystem. Paramount among these is the unwavering commitment to maintaining the confidentiality of all information acquired during their assessment activities. Complementing this, a robust professional liability insurance policy is mandatory, providing adequate financial coverage for their activities, with the scope and value of such insurance commensurate with the inherent risks associated with their specific sector of operation. Furthermore, Notified Bodies are obligated to proactively provide pertinent information to their notifying authority, market surveillance authorities, and fellow notified bodies. Information sharing with other notified bodies is particularly focused on instances of refused conformity assessment attestations, specifically identifying the product and manufacturer in question, facilitating a coordinated approach to non-compliant products across the Union. Operationally, Notified Bodies are expected to function competently, non-discriminatorily, transparently, neutrally, independently, and impartially, upholding the highest standards of professional conduct. Crucially, they remain accountable to national competent authorities, acknowledging their responsibilities in areas of significant public interest, thereby embedding their function within a broader framework of public accountability. In terms of market accessibility, Notified Bodies are explicitly free to offer their conformity assessment services to any economic operator, irrespective of their establishment location within or outside the European Union, fostering a level playing field for global manufacturers seeking access to the EEA market. To maintain currency and relevance, Notified Bodies are expected to actively participate in relevant standardisation activities or, at minimum, remain comprehensively informed of such developments, ensuring their assessment practices align with the evolving landscape of industry standards. Transparency and information dissemination are further emphasised through the requirement to inform the notifying authority about all certificates issued, refused, restricted, suspended, or withdrawn due to safety-related non-conformities, as well as providing comprehensive reports on other conformity assessment activities undertaken. Moreover, a collaborative framework is fostered through the obligation to provide other bodies notified under the same Union harmonisation legislation with pertinent information regarding both negative and, upon request, positive conformity assessment results, promoting consistency and shared learning within the notified body network. Finally, the responsibility extends to market surveillance, with Notified Bodies required to furnish market surveillance authorities, both domestically and across Member States, with relevant information facilitating effective market monitoring and enforcement. It is critical to note that

this principle of independence permeates the entire Notified Body organisation, extending to bodies affiliated with business associations or professional federations, ensuring that structural affiliations do not compromise their objective assessment function.

In instances where specialised expertise or capacity is required, Notified Bodies are permitted to subcontract specific tasks associated with conformity assessment. However, this delegation of responsibility is carefully circumscribed, stipulating that the Notified Body retains full and ultimate responsibility for all tasks performed by subcontractors, and must rigorously ensure that any subcontractor fully adheres to all regulatory requirements. Subcontracting arrangements must be formalized through legally binding contracts, guaranteeing transparency and fostering confidence in the integrity of the Notified Body's overall operations. Furthermore, these contractual agreements should explicitly prohibit any further levels of subcontracting, maintaining a clear and manageable chain of responsibility. Critically, the Notified Body bears the onus of ensuring that subcontractors and their personnel meet all regulatory requirements as if the tasks were executed by their own employees, upholding consistent standards regardless of task delegation. It is imperative to emphasise that certificates and attestations of conformity are invariably issued under the name and explicit responsibility of the Notified Body, irrespective of any subcontracted activities, reinforcing their ultimate accountability.

Notified Bodies constitute an indispensable component of the European regulatory framework, serving as independent assessors critical to ensuring that products circulating within the EU market meet stringent standards for safety, health, and performance. Acting as supposed impartial arbiters of conformity, they provide a vital layer of assurance, fostering confidence in the conformity of products and contributing significantly to the overall integrity and safety of the European marketplace. However, their role is not without its challenges, which are explored in the following section.

4.3. Issues related to notified bodies

While the reliance of notified bodies under the EU regulatory framework is intended to streamline trade and ensure safety, this system gives rise to significant and systemic issues[294]. Chief among these challenges are persistent concerns regarding the independence and demonstrable competence of notified bodies, crucial for unbiased evaluations[295]. This is exacerbated by a pervasive problem of fragmented oversight and accountability within the broader European certification regime, characterised by a lack of transparency and limited public scrutiny[296]. These interconnected issues collectively raise profound questions about the system's effectiveness, its susceptibility to economic pressures over safety imperatives, and its overall democratic legitimacy, demanding a critical re-evaluation to ensure genuine public protection

The advent of the New Approach and the NLF have inadvertently fostered a competitive market among notified bodies, a dynamic that, while potentially driving down costs for businesses seeking certification, also introduces complexities. Companies are afforded the liberty to select any European notified body for conformity assessment, a choice that, in theory, promotes efficiency and

[294] Jean-Pierre Galland, 'Standards, Certification, and Accreditation: Indispensable Tools for European Safety Regulations?' in Jean-Christophe Le Coze and Benoît Journé (eds), *The Regulator–Regulatee Relationship in High-Hazard Industry Sectors: New Actors and New Viewpoints in a Conservative Landscape* (Springer Nature Switzerland 2024) <https://doi.org/10.1007/978-3-031-49570-0_8> accessed 30 April 2024.

[295] Jean-Pierre Galland, 'The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies' (2013) 4 European Journal of Risk Regulation 365 <https://doi.org/10.1017/S1867299X00002634>.

[296] Christian Frankel and Jean-Pierre Galland, 'Market, Regulation, Market, Regulation' (Wissenschaftsverlag Mainz Aachen, Germany 2015) <https://enpc.hal.science/hal-01275604> accessed 24 May 2024.

responsiveness. However, there is a persistent challenge of ensuring the independence and demonstrable competence of these entities[297]. This concern encompasses not only their operational autonomy from product manufacturers but also their structural and functional separation from undue influence by Member States. Furthermore, disconcerting fragmentation exists within the European certification regime, characterised by a lack of cohesive oversight and interconnectedness. A comprehensive, cross-cutting analysis of notified bodies, encompassing their legal structures, public or private status, revenue streams, and staffing levels, remains conspicuously absent. This opacity is compounded by the inherently discreet nature of their evaluations, which primarily unfold within the confines of direct, often opaque, relationships with product producers.

Emerging from this landscape are the Big Third-Party Certifiers (BTCP), a cohort of certification bodies that have achieved global reach and increasingly function as de facto advisors to regulatory bodies[298]. These BTPCs, leveraging their intermediary position between regulators and regulated entities, strategically pursue expansion across diverse sectors, extend their global footprint, and exert subtle but significant influence on the regulatory trajectory. Their expansionist strategies often involve advising regulators to adopt system-based regulations that conveniently align with their pre-existing international networks, thereby solidifying their market dominance. Crucially, their market penetration extends beyond traditional safety certifications into the domains of social and environmental standards, capitalizing on established relationships with producers to capture new certification markets. Indeed, BTPCs increasingly assume quasi-regulatory roles, authoring best practice certification guides and actively participating in the shaping of auditable standards, further blurring the lines between private certification and public regulation.

Illustrative case studies underscore the complexities and potential vulnerabilities inherent in this system. Perhaps the most salient illustration of systemic weakness in this system is the Poly Implant Prothese (PIP) scandal, a watershed event involving a French company's fraudulent use of unauthorized silicone in breast implants[299]. A German notified body had certified these products, raising profound questions about the efficacy of prevailing certification procedures and the depth of due diligence exercised by notified bodies[300]. The scandal exposed a regulatory culture that, at times, prioritised procedural paperwork over substantive product examination. The scandal also highlighted the problematic practice of notified bodies utilizing subsidiaries with potentially insufficient expertise, further compromising the rigor of the certification process. Moreover, the inherent ambiguities within the regulatory language are often exploited by various stakeholders to advance their own vested interests, further complicating effective regulatory enforcement[301].

These instances, alongside broader systemic observations, reveal significant weaknesses and challenges within the New Approach framework. Galland points to a "democratic deficit" in the implementation of this regulatory model[302]. From the initial drafting of essential safety requirements to the final stages of conformity certification, public scrutiny is demonstrably limited. Public consultation, when it occurs, is often relegated to surveys conducted by informal groupings, lacking genuine

[297] Galland, 'The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies' (n 295).

[298] Jean-Pierre Galland, 'Big Third-Party Certifiers and the Construction of Transnational Regulation' (2017) 670 ANNALS of the American Academy of Political and Social Science 263.

[299] Peter Rott, 'Certification of Medical Devices: Lessons from the PIP Scandal' in Peter Rott (ed), *Certification – Trust, Accountability, Liability*, vol 16 (Springer International Publishing 2019) <https://doi.org/10.1007/978-3-030-02499-4_9> accessed 29 April 2024.

[300] Barend Van Leeuwen, 'PIP Breast Implants, the EU's New Approach for Goods and Market Surveillance by Notified Bodies' (2014) 5 European Journal of Risk Regulation 338 <https://doi.org/10.1017/S1867299X0000386X>.

[301] Rott (n 299).

[302] Galland, 'The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies' (n 295).

participatory mechanisms. Furthermore, the reliance on accreditation as a solution to ensure the competence and reliability of notified bodies has not consistently prevented regulatory failures, as evidenced by the PIP scandal. The decentralized and loosely coupled nature of the system, characterised by a diffuse network of actors with potentially divergent interests, creates opportunities for economic considerations to overshadow paramount safety imperatives. Consequently, the systematic reliance on standardisation and certification procedures, while ostensibly promoting safety and trade facilitation, may inadvertently lead to an over-reliance on these mechanisms, potentially at the expense of more robust and publicly accountable forms of regulatory oversight.

These critical remarks suggest that while the NLF has undeniably contributed to trade facilitation and the pursuit of safety objectives within the EU, it is not without its inherent contradictions and vulnerabilities. The framework, intended to be a generic model applicable to emerging safety and security challenges, such as data protection, artificial intelligence, and cybersecurity, may inadvertently introduce new complexities[303]. Adding layers of safety procedures to already high-hazard systems can paradoxically obfuscate responsibilities and dilute accountability.

5. The European standardisation system

5.1 The standardisation regulation

The evolution of the legal framework from the Old Approach to the NLF reveals how European standardisation constitutes a linchpin in the project of economic and political integration within the Union. It transcends the purely technical domain, embedding itself deeply within the fabric of public policy and legislative support, assigning to technical standards the role of effectively operationalising and concretizing broad policy directives. Alongside the facilitation of the free movement of goods and services across the internal market through the implementation of European Union harmonisation legislation, standardisation concurrently aims to ensure sustainable development, characterised by elevated levels of safety and quality.

Given the important role of technical standardisation in the European context, the legislative framework descried in the previous sections has been enriched by other initiatives that aim to define the roles and responsibilities of the ESOs and their relationship with public authorities. The Regulation (EU) No 1025/2012 on European standardisation constitutes a cornerstone of the European Union's approach to standardisation[304].

This legislative instrument is demonstrably ambitious and multifaceted in its objectives, aiming to streamline and fortify the collaborative framework underpinning the creation and application of European standards. A meticulous examination of the regulation reveals its intent to harmonise the roles of diverse actors within the European standardisation landscape, encompassing European standardisation organisations, national standardisation bodies, Member States, and the European Commission itself. At its core, the regulation seeks to establish a robust and adaptable system for the

[303] Galland, 'Standards, Certification, and Accreditation' (n 294).

[304] Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council Text with EEA relevance 2012 (OJ L).

development of European standards and European standardisation deliverables. These instruments are explicitly designed to bolster Union legislation and policies, effectively serving as technical specifications that can be referenced in various regulatory and policy contexts across the European single market. Furthermore, the regulation evinces a forward-looking perspective by incorporating specific provisions for the identification and utilisation of ICT technical specifications. This inclusion underscores the EU's recognition of the pivotal role of ICTs in the contemporary economic and social landscape and the necessity for interoperability and common technical grounds in this rapidly evolving domain. Beyond the operational aspects of standard development and application, Regulation 1025/2012 also addresses the crucial elements of financing European standardisation activities and ensuring meaningful stakeholder participation in these processes. This dual emphasis on resource allocation and inclusive governance mechanisms highlights the regulation's holistic approach to fostering a vibrant and effective European standardisation system.

To fully grasp the scope and implications of Regulation 1025/2012, it is imperative to delineate its foundational definitions. The regulation provides a clear conceptual architecture by defining key terms such as "standard", "European standardisation deliverable", and "technical specification". A "standard", as articulated in the regulation, is understood as a technical specification adopted by a recognised standardisation body for repeated or continuous application, with the critical caveat that compliance remains voluntary. This voluntary nature is a defining characteristic of standards within the European system, distinguishing them from legally binding regulations and directives. The regulation further categorises standards into international, European, harmonised, and national variants, reflecting the multi-layered governance of standardisation at different geographical and jurisdictional levels. In addition to formal standards, the regulation recognises 'European standardisation deliverables', a broader category encompassing any technical specification, other than a European standard, adopted by a European standardisation organisation for repeated application, again, with voluntary compliance. This broader category acknowledges the evolving landscape of technical specifications and the need to accommodate diverse forms of standardised outputs beyond traditional norms. The overarching concept of a 'technical specification' is defined as a document that prescribes technical requirements to be fulfilled by a product, process, service, or system. This broad definition underscores the pervasive applicability of standardisation across diverse sectors and economic activities. Within the technologically salient domain of ICTs, the regulation specifically defines "ICT technical specification" as a technical specification directly pertinent to information and communication technologies. Finally, the regulation formally identifies "European Standardisation Organisations" by listing CEN, CENELEC, and ETSI in Annex I, thereby providing concrete institutional anchors for the European standardisation system. Similarly, the concept of "National Standardisation Body" is defined as a body notified to the Commission by a Member State, establishing a clear link between national and European level standardisation efforts. These meticulously defined terms provide the essential vocabulary for understanding the mechanisms and procedures outlined within Regulation 1025/2012.

A critical element embedded within Regulation 1025/2012 is the emphasis on transparency and inclusive stakeholder participation. The regulation mandates that both ESOs and NSBs establish and publicise annual work programs, detailing their standardisation activities. This requirement for public disclosure, facilitated through readily accessible websites and notifications to relevant bodies and the Commission, ostensibly aims to foster transparency and accountability within the standardisation system. Furthermore, the regulation explicitly prohibits NSBs from objecting to a standardisation subject being considered at the European level, promoting a centralised and harmonised approach to standardisation within the Union. In terms of stakeholder engagement, the regulation stipulates that ESOs should actively

encourage the representation and participation of a diverse array of stakeholders, including SMEs, consumer organisations, and environmental and social interest groups. This reflects a policy objective to democratise the standardisation process, moving beyond purely industry-led initiatives to incorporate a broader spectrum of societal interests. NSBs are similarly tasked with encouraging and facilitating SME access to standards and standards development processes, recognising the vital role of SMEs in innovation and economic growth, and ensuring they are not unduly disadvantaged by standardisation procedures. Moreover, Member States are encouraged to foster the participation of public authorities in national standardisation activities, thereby aligning public sector interests with broader standardisation efforts and potentially leveraging public procurement as a mechanism to promote standard adoption. These provisions related to transparency and stakeholder participation are ostensibly designed to enhance the legitimacy, relevance, and ultimately the effectiveness of European standards.

The strategic deployment of European standards in support of Union legislation and policies represents a core operational objective of Regulation 1025/2012. To this end, the regulation empowers the Commission to adopt an annual Union work program for European standardisation. This program serves as a strategic planning document, identifying priority areas and indicating the European standards and standardisation deliverables that the Commission intends to request from ESOs. This proactive approach allows the Commission to steer European standardisation efforts towards areas of strategic policy relevance, ensuring that standards are developed in a manner that directly supports the achievement of Union-wide policy goals. The Commission is further empowered to formally request ESOs to draft European standards or European standardisation deliverables, effectively initiating the standard development process in areas deemed strategically important. A particularly significant provision concerns harmonised standards. Where a harmonised standard demonstrably satisfies the requirements of Union harmonisation legislation, the Commission is mandated to publish a reference to it in the Official Journal of the European Union. This publication confers a significant status upon the harmonised standard, creating a presumption of conformity with the essential requirements of the relevant Union legislation for products or services that comply with the standard. This mechanism provides a streamlined and efficient pathway for demonstrating regulatory compliance and facilitating market access within the single market. To further enhance the system's responsiveness and market relevance, the Commission has established a notification system for stakeholders. This system is purportedly designed to ensure proper consultation throughout the standard development process, fostering a more iterative and feedback-driven approach to standard creation and ensuring that the resulting standards are genuinely fit for purpose and reflective of market needs.

Recognising the unique and rapidly evolving nature of the ICT sector, Regulation 1025/2012 includes specific provisions pertaining to ICT technical specifications. The regulation acknowledges that not all relevant technical specifications in the ICT domain originate from formal standardisation bodies at the national, European, or international level. Therefore, it empowers the Commission to identify ICT technical specifications that, while not formally designated standards, nonetheless meet specific criteria outlined in Annex II of the regulation. These criteria, though not explicitly detailed in the provided notes, likely pertain to the technical rigor, market relevance, and interoperability potential of the ICT specifications. A significant outcome of this identification process is that these recognised ICT technical specifications can then be referenced in public procurement procedures. This provision is particularly salient given the substantial public sector expenditure on ICT goods and services and the potential of public procurement to act as a powerful lever for promoting innovation and interoperability in the ICT market. By allowing public procurement to reference a broader range of technically sound ICT specifications, including those beyond traditional standards, the regulation aims to foster greater flexibility

and responsiveness to the dynamic pace of innovation in the ICT sector. This approach potentially facilitates cross-border service provision, encourages greater competition among ICT providers, and ultimately promotes enhanced interoperability and innovation within the European digital economy.

The financial dimensions of European standardisation are explicitly addressed within Regulation 1025/2012. The regulation acknowledges that the development and maintenance of a robust standardisation system require dedicated financial resources and accordingly provides for the possibility of Union financing for various standardisation activities. Specifically, the Union may financially support ESOs for activities such as the development and revision of European standards, ensuring the quality and consistency of standards, conducting preliminary research and preparatory work, managing central secretariat functions, translating standards into multiple languages, and implementing technical assistance programs aimed at promoting standard adoption and implementation. This broad range of eligible activities underscores the comprehensive nature of Union financial support, encompassing the entire lifecycle of standard development and dissemination. Furthermore, financing may also be extended to NSBs and other bodies demonstrably contributing to these aforementioned standardisation activities. This acknowledges the distributed nature of the European standardisation system and the need to support relevant actors at both the European and national levels. Crucially, the regulation also contemplates the financing of European stakeholder organisations, provided they meet specific criteria yet to be fully elaborated upon in the provided notes. This potential funding stream for stakeholder organisations, particularly those representing SMEs, consumers, and environmental interests, could serve as a significant mechanism to enhance their effective participation in the standardisation process, ensuring that their perspectives are adequately represented and considered throughout the standard development lifecycle. The provision of Union financing is thus presented as an enabling mechanism, designed to ensure the financial sustainability and operational effectiveness of the European standardisation system, supporting its overarching objectives of fostering competitiveness, innovation, and societal well-being.

In terms of governance and adaptability, Regulation 1025/2012 incorporates both delegated and implementing powers vested in the Commission. The Commission is empowered to adopt delegated acts, a form of secondary legislation that allows for the amendment of certain non-essential elements of the regulation. Specifically, delegated acts can be used to amend annexes to the regulation, such as updating the list of recognised ESOs in Annex I and adapting the criteria for European stakeholder organisations eligibility for Union financing. This delegated power provides a mechanism for the regulation to remain current and responsive to evolving circumstances, allowing for adjustments to institutional arrangements and eligibility criteria without requiring lengthy and complex legislative revisions. In addition to delegated powers, the Commission is also assisted by a committee in implementing the regulation. The specific composition and mandate of this committee are not detailed in the provided notes, but its existence suggests a procedural framework for the Commission to consult with Member States and other relevant stakeholders in the practical application and enforcement of the regulation. This committee structure likely plays a role in ensuring consistent and harmonised implementation of the regulation across the diverse national contexts of the European Union.

To ensure accountability and facilitate ongoing evaluation of the effectiveness of Regulation 1025/2012, the regulation mandates reporting and review mechanisms. ESOs and European stakeholder organisations are required to submit annual reports to the Commission, providing insights into their activities and progress in relation to standardisation efforts. These reports serve as crucial inputs for monitoring the functioning of the European standardisation system and identifying areas for potential improvement or further regulatory action. Furthermore, the Commission is obligated to present a

comprehensive report to the European Parliament and the Council on the overall implementation of the regulation. This high-level reporting requirement ensures parliamentary oversight and political accountability, providing a mechanism for the legislative bodies of the European Union to assess the impact and effectiveness of Regulation 1025/2012 in achieving its stated objectives. This periodic review process is essential for ensuring that the regulatory framework remains relevant, effective, and aligned with the evolving needs of the European economy and society.

The anticipated effects and overall impact of Regulation 1025/2012 are demonstrably wide-ranging, extending across economic competitiveness, consumer welfare, and the addressing of broader societal challenges. In terms of competitiveness, the regulation is expected to facilitate the free movement of goods and services within the single market by promoting network interoperability, common means of communication, and fostering technological development and innovation. Economically, the regulation is projected to promote greater economic interpenetration within the internal market, encourage the development of new products and markets by reducing technical barriers to trade, improve supply conditions through standardised quality and technical specifications, intensify competition by levelling the playing field for market entrants, and ultimately contribute to lowering costs for businesses and consumers alike. For consumers, the regulation aims to maintain and enhance product and service quality by establishing clear benchmarks, provide readily accessible information about product characteristics through standardised labelling and documentation, ensure interoperability and compatibility between different products and systems, and ultimately increase overall product safety and value. Beyond purely economic considerations, Regulation 1025/2012 is also positioned as a tool to address major societal challenges. The regulation's emphasis on standards relevant to areas such as climate change mitigation, sustainable resource use, promotion of innovation, and the integration of people with disabilities underscores its broader societal relevance and its potential contribution to achieving wider European policy objectives beyond market integration. Furthermore, the regulation's specific provisions designed to enhance SME participation in standardisation are anticipated to empower smaller businesses to actively contribute their innovative technology solutions to standardisation efforts, fostering a more inclusive and dynamic innovation ecosystem. Finally, by enabling public procurement to reference a wider range of ICT technical specifications, the regulation seeks to improve the efficiency and effectiveness of public spending on ICT goods and services, responding to the rapid evolution in the ICT sector and promoting cross-border service provision, competition, and innovation in public procurement markets.

In essence, Regulation (EU) No 1025/2012 represents a significant evolution in the European approach to standardisation. It moves beyond a purely reactive or sector-specific approach towards a more strategic and integrated framework, explicitly designed to support Union policies, enhance competitiveness, and address multifaceted societal challenges. By focusing on enhanced cooperation between key actors, promoting transparency and inclusive stakeholder participation, and providing targeted financial support, the regulation seeks to create a more effective, relevant, and responsive European standardisation system.

5.1.1. Evaluation and revision of the standardisation regulation

In 2024, the Regulation on European standardisation underwent an evaluation[305] to provide an informed, evidence-based analysis of its performance across several criteria: effectiveness, efficiency,

[305] 'European Standardisation – Evaluation' <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13446-European-standardisation-evaluation_en> accessed 16 August 2025.

coherence, EU added value, and relevance. This evaluation also aimed to identify its strengths and weaknesses to address them in the forthcoming revision of the Regulation[306]. The impetus for this assessment was a significantly altered political and economic landscape, marked by the dual green and digital transformations, emergent geo-economic challenges, and a decade of profound jurisprudence from the Court of Justice of the European Union.

The aggregate findings of this evaluation indicate that while the Regulation has yielded certain improvements, particularly in the velocity and inclusiveness of the standard-setting process, its framework is increasingly struggling to satisfy contemporary market, policy, and legal exigencies. A more granular examination of its performance across the established evaluation criteria reveals the nature and extent of these challenges.

With respect to effectiveness, the Regulation has proven to be only moderately successful in reducing the development time for harmonised standards. The process is still widely perceived as excessively slow, with an average duration of six years, of which three are typically consumed by drafting and consensus-building, a timeline misaligned with the rapid pace of legislative requirements, market needs, and global competition. These protracted delays, combined with procedural complexities and systemic inefficiencies, demonstrably hamper the potential positive impact of harmonised standards on the competitiveness of EU enterprises. The root causes of these impediments are multifaceted, attributable to a confluence of factors including persistent procedural hurdles, recurrent quality deficiencies in draft standards, and the imposition of new safeguards and control mechanisms required by CJEU case law (e.g., the Elliot case). Concurrently, ESOs face mounting pressure from both legislation and the market to deliver in a more timely fashion, a challenge compounded by persistent communication and coordination deficits and the insufficient deployment of digital technologies. While ESOs and National Standardisation Bodies possess a strong intrinsic incentive to develop harmonised standards due to their legal value, they may lack a compelling impetus to adapt or accelerate their processes. This is because the Regulation exclusively permits standardisation requests to be addressed to the three designated ESOs, with a conspicuous absence of viable alternatives when harmonised standards are not delivered on schedule. In terms of stakeholder participation, the Regulation has been moderately effective in augmenting the involvement of SMEs and civil society actors. Nevertheless, their substantive impact is often circumscribed by a lack of the necessary resources, institutional capacity, and technical expertise required for effective engagement in complex and time-consuming standardisation activities at the European, national, and international levels. Finally, the Regulation's effectiveness in promoting the use of ICT specifications has been limited, a deficiency partly mitigated by the advent of new digital legislation that itself relies on harmonised standards.

From the perspective of efficiency, the evaluation encountered significant challenges in data collection, particularly for the cost-benefit analysis, which had to rely on public consultations, surveys, and interviews rather than comprehensive quantitative data. The direct reporting obligations imposed upon ESOs, NSBs, Annex III organisations[307], and the European Commission were generally deemed proportionate, in part due to ongoing simplification efforts. Although robust data on indirect compliance costs are lacking, the cost-benefit analysis suggests that the benefits accrued from the Regulation outweigh the costs incurred by business and civil society stakeholders, notwithstanding the significant

[306] 'Standardisation Regulation – Revision' (*European Commission - Have your say*) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14511-Standardisation-Regulation-revision_en> accessed 16 August 2025.

[307] These are stakeholders eligible for participation within the ESOs as recognised by Regulation 1025/2012. hese bodies are: ANEC (Association Normalization Européenne pour les Consommateurs) for consumers' interests, ECOS (Environmental Coalition on Standards) for environmental interests, ETUC (European Trade Union Confederation) for workers' interests and SBS (Small Business Standards) for SMEs.

costs associated with expert participation. While ESOs have progressively deployed digital tools to facilitate standards development, their rate of adoption lags behind that of other international standards-development organisations. Furthermore, the formal objections system is widely considered to be excessively long, cumbersome, and inefficient. A central paradox has emerged wherein an initial objective of simplification has been superseded by the introduction of new administrative and adoption procedures over the last decade, largely as a consequence of CJEU case law[308]. This jurisprudence has clarified the legal status of harmonised standards as an integral part of EU law and affirmed the Commission's central supervisory role. While these procedural steps are legally necessary, they have rendered the overall process more burdensome and less efficient. Opportunities for simplification, such as improved reporting and monitoring, procedural streamlining, and accelerated development processes, are constrained by the inherent rigidity of the Regulation and its complete reliance on the standardisation capabilities of the three designated ESOs.

The Regulation was found to possess a high degree of internal coherence. It has contributed to greater consistency in the application of harmonised standards across the EU legislative acquis, most notably within legal instruments governed by the NLF. Some minor inconsistencies persist, for instance in the terminology used in the General Product Safety Regulation. It is also noteworthy that in certain policy domains, legislators have for justifiable reasons opted for alternative standards-development systems. Examples include the Corporate Sustainability Reporting Directive, which established a separate standard-setting body, and the Regulation on the European Health Data Space, which empowers the Commission to adopt common specifications via implementing acts.

A clear consensus emerged among stakeholders that the Regulation provides distinct EU added value and that European standardisation is highly appreciated. A key objective that has been successfully achieved is the systematic removal of conflicting national standards, a critical step for the functioning of the internal market. A hypothetical withdrawal of the Regulation and its replacement with disparate national laws would undoubtedly impose a significant burden on intra-EU trade and prove detrimental to the Single Market.

However, the relevance of the Regulation has been called into question by the evolving landscape. The framework's weaknesses have been thrown into sharp relief by a series of external factors: accelerated technology cycles, particularly in the digital and green domains; an aging population of technical experts; shifting geo-economic dynamics; the increased assertiveness of non-EU countries in international standardisation bodies; and the aforementioned impact of CJEU case law. Consequently, the existing framework struggles to deliver effectively in emerging technology areas that are critical to the EU's policy ambitions and the competitiveness of its industries. The EU's capacity to act as a global standard-setter is increasingly challenged, necessitating more agile coordination and faster positioning within international standardisation fora. The fact that the majority of harmonised standards currently cited in the Official Journal of the European Union are more than five years old is a stark indicator of a system that needs to enhance its responsiveness to innovation and improve its anticipation of future standardisation needs, including at the pre-normative research stage. Moreover, the jurisprudence of the CJEU, while clarifying legal status, has also created procedural complexities and highlighted an urgent need for greater transparency and accessibility of harmonised standards. This confluence of challenges

[308] For an extensive discussion of key cases, see Annalisa Volpato, 'The Legal Effects of Harmonised Standards in EU Law: From Hard to Soft Law, and Back?' *The Legal Effects of EU Soft Law* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781802208917/book-part-9781802208917-16.xml> accessed 4 January 2026. More recently, Olia Kanevskaia, 'Is It Really All about the Money? The Future of European Standardization after PublicResourceOrg' (2025) 16 European Journal of Risk Regulation 344 <https://doi.org/10.1017/err.2024.64>.

suggests that the alternative routes to standardisation, already being explored in specific sectors, may need to be considered more broadly to meet future requirements.

5.2. The Vademecum on European Standardisation

As mentioned in previous sections, the principle of reference to standards is a central tenet of the European standardisation system since the New Approach. The Regulation on European Standardisation provides guidance on how this reference to standards can be achieved through cooperation among a diverse number of actors, primarily the ESOs and the EC. In this context, standardisation requests play the crucial role of initiating the entire mechanism to concretely realize the interplay between technical standards and EU law. Given the complexities of this process, the European Commission has promulgated the Vademecum on European Standardisation[309]. This document ostensibly functions as a comprehensive guide, aiming to cultivate a shared interpretive framework concerning standardisation mandates and the intricate interplay of roles within the European standardisation system. Addressed to a broad spectrum of stakeholders, ranging from Commission officials and public authorities within Member States and the EFTA countries to ESOs themselves, organisations identified in Annex III of Regulation 1025/2012, National Standardisation Bodies, and various industry stakeholders, the Vademecum's declared purpose is to underpin the application of Union legislation and policies pertaining to both products and services through the utilisation of European standards or European standardisation deliverables.

The principal objective of the Vademecum, as explicitly articulated, is to elucidate the roles and responsibilities inherent in the sequential processes of planning, preparing, and executing standardisation requests. This objective underscores the document's instrumental function in fostering clarity and procedural coherence within the European Standardisation system. Initially published in 2003 and subsequently revised in 2009, the contemporary iteration of the Vademecum reflects the evolving landscape of European standardisation policy. This historical evolution is noteworthy, signifying a responsiveness to the dynamic nature of the Union's standardisation policy and a commitment to maintaining the Vademecum's relevance within this evolving context. While the inaugural version primarily addressed standardisation requests issued under Union harmonisation legislation for products, the revised Vademecum adopts a markedly broader purview. This expansion in scope indicates a strategic shift towards leveraging European standards in support of a wider array of Union policies, extending beyond the traditional confines of product harmonisation. This broadening of scope is a significant development, potentially amplifying the influence of European standards across diverse sectors and policy domains, though it simultaneously raises questions regarding the document's capacity to maintain specificity and actionable guidance across this expanded remit.

The Vademecum is partitioned into three distinct parts, each strategically targeted at specific segments of its intended audience. Part I, concerning the role of the Commission's standardisation requests to the ESOs, is broadly addressed to Commission officials and all actors within the ESS. This initial section lays the foundational groundwork by defining the conceptual underpinnings of standardisation requests within the European Union's policy framework. Part II, focusing on the preparation and adoption of these requests, is specifically directed at Commission officials. This targeted

[309] 'Vademecum on European Standardisation in Support of Union Legislation and Policies. PART I - Role of the Commission's Standardisation Requests to the European Standardisation Organisations' <https://ec.europa.eu/docsroom/documents/13507/attachments/1/translations> accessed 4 May 2023.

approach suggests an emphasis on equipping Commission departments with the procedural knowledge necessary for effectively initiating and formalising standardisation requests. Finally, Part III, providing guidelines for the ESOs' execution of standardisation requests, is primarily addressed to the ESOs and their technical bodies. This section underscores the Vademecum's commitment to guiding the operational phases of standardisation, ensuring that ESOs possess a clear understanding of the Commission's expectations and the procedural norms for developing, publishing, and revising requested deliverables. This tripartite structure, therefore, reflects a deliberate attempt to cater to the differentiated needs and responsibilities of the various stakeholders within the ESS, providing tailored guidance at each stage of the standardisation request lifecycle.

Underpinning the Vademecum are several key principles, notably inclusiveness, transparency, and compliance with the stipulations outlined in a standardisation request. These tenets are presented as foundational to the legitimacy and effectiveness of the ESS. The emphasis on inclusiveness suggests an aspiration towards a participatory standardisation process, ideally accommodating the diverse interests and perspectives of relevant stakeholders. Transparency, similarly, aims to foster trust and accountability within the system, implying open access to information and clear procedural pathways. Compliance, as a principle, highlights the imperative of adhering to the established requirements embedded within a standardisation request, thereby ensuring the integrity and reliability of the resulting European standards. Furthermore, the Vademecum is presented as establishing a legal basis for issuing requests in domains extending beyond Union harmonisation legislation on products. This provision significantly expands the Commission's capacity to utilise standardisation as a policy instrument, allowing for requests in support of Union policies even in the absence of specific Union legislation. This expanded remit, while potentially enhancing the policy-supporting function of European standards, may also introduce complexities regarding the legal and political justifications for such requests, requiring careful consideration of subsidiarity and proportionality principles.

Within Part I of the Vademecum, the explicit objectives are to foster a common understanding of the role of standardisation requests and to clarify the roles and responsibilities of the various actors involved in their planning, preparation, and execution. This section grounds the discussion by situating European standardisation within the broader context of the European Union, referencing key documents that establish and implement the Union's standardisation policy and acknowledging the formal recognition of ESOs within this framework. The document proceeds to define the concept of a standardisation request, characterising it as an implementing act, a designation that carries significant legal and procedural implications. The Vademecum also acknowledges the potential ramifications of non-compliance with a standardisation request, although the specifics of these implications warrant further scrutiny. A detailed delineation of the actors recognised by the Regulation follows, encompassing the roles of the Commission, Member States and national public authorities, ESOs, ESO members (including NSBs), and industry stakeholders. This comprehensive mapping of actor roles is crucial for establishing a shared understanding of the distributed responsibilities within the ESS, albeit the precise nature of the power dynamics and potential for conflicts of interest between these actors could benefit from further critical examination. The conditions under which a standardisation request may be issued, encompassing annual planning, basic conditions, preliminary and ancillary actions, and the specific areas in which requests may or must be used, are subsequently addressed. This section provides practical guidance on the initiation and justification of standardisation requests, offering a structured framework for Commission officials. Finally, Part I delves into the scope, work programme, and validity of standardisation requests.

Part II of the Vademecum shifts focus to the procedural aspects of preparing and adopting Commission standardisation requests, aiming to clarify the principles governing drafting and adoption. This section is intrinsically procedural, outlining the responsibilities of Commission departments, the overall workflow, the setting of target dates, and the logistical considerations of translations. A significant emphasis is placed on consultations, encompassing consultations with ESOs, Annex III organisations, Member States' sectoral experts, and other relevant stakeholders. This consultative approach ostensibly aims to ensure that standardisation requests are well-informed, technically robust, and reflect the diverse perspectives of relevant parties. The Vademecum also addresses Union financing of European standardisation activities, a crucial practical consideration for the implementation of standardisation requests. General guidance for drafting standardisation requests is provided, covering aspects such as addressees, clarity in indicating the legal framework and justifying a request, defining the scope of requirements, referencing legal requirements, indicating legal requirements intended to be covered by European standards, and managing significant changes in European standards supporting Union legislation. This detailed guidance suggests a commitment to promoting well-structured and legally sound standardisation requests. The provision of a model structure for standardisation requests, outlining the typical components of a Commission implementing decision, further enhances the practical usability of the Vademecum for Commission officials. Finally, Part II addresses the acceptance and execution phases, including ESOs' acceptance, agreement on the work programme, updates, Commission follow-up, informing the Committee on Standards, and establishing common principles for deadlines. This procedural granularity demonstrates the Vademecum's ambition to provide comprehensive operational guidance throughout the entire lifecycle of a standardisation request.

Part III, directed towards ESOs, provides guidelines for the execution of standardisation requests, aiming to establish coherent practices for developing, publishing, and revising Commission-requested deliverables. This section introduces general principles for execution, encompassing project planning, the establishment and agreement of requested work programmes, updates, reporting, and coherence in the standardisation work itself. Specific guidance is offered for developing harmonised standards, addressing measures to ensure compliance with a request, developing provisions, selecting normative references, indicating legal requirements, and considering harmonised standards developed by other bodies. This detailed technical guidance is crucial for ensuring that ESOs develop harmonised standards that effectively support Union legislation. The Vademecum further addresses the adoption and revision of requested European standards and European standardisation deliverables by ESOs, including considerations of validity, reflecting the state of the art, restricting scope, managing significant changes, standstill periods, obsolete standards, and cooperation with other bodies. The section concludes by emphasising liaison with the Commission and the consideration of international relevance, highlighting the interconnectedness of European standardisation with broader global standardisation efforts.

In summary, the Vademecum on European Standardisation emerges as a meticulously structured and comprehensive guide, designed to clarify the roles, responsibilities, and procedural intricacies within the European Standardisation System. Divided into three strategically targeted parts, the document addresses distinct segments of its stakeholder audience, offering tailored guidance on the initiation, preparation, execution, and compliance assessment of standardisation requests. Its emphasis on inclusiveness, transparency, and compliance, coupled with its expanded scope beyond product harmonisation, underscores its ambition to serve as a foundational document for ensuring the effectiveness and legitimacy of European standards in supporting a wide spectrum of Union legislation and policies. However, the ultimate efficacy of the Vademecum rests not merely on its textual comprehensiveness, but also on its practical application and its capacity to adapt to the ever-evolving dynamics of the European

regulatory landscape and the inherent complexities of multi-stakeholder governance in standardisation processes. Further empirical research would be necessary to fully assess the Vademecum's impact on fostering a truly common understanding and achieving enhanced procedural coherence within the ESS, particularly in light of its broadened scope and the diverse and potentially divergent interests of the actors it seeks to guide

5.2 Open issues related to the European standardisation system

5.2.1. Legal status of harmonised standards

The legal status of technical standards within the European Union represents a complex and multifaceted issue[310]. The precarious location of harmonised standards in the EU law, positioned ambiguously between the realms of binding law and voluntary guidelines, generates significant theoretical and practical challenges for the EU legal order. Technical standards, ostensibly conceived as instruments of soft law, occupy a peculiar niche, exerting considerable influence despite lacking the formal binding force typically associated with legal instruments. This section will explore the nuanced legal standing of technical standards, dissecting their function, integration within EU law, and the attendant issues of judicial review, legitimacy, and accountability. Through a critical examination of these elements, it will become evident that the seemingly straightforward dichotomy of soft versus hard law proves insufficient to encapsulate the intricate reality of technical standards and their operational impact within the European legal framework.

Initially, it is crucial to delineate the conceptual terrain of soft law and hard law within the EU context[311]. Article 288 of the Treaty on the Functioning of the European Union (TFEU) explicitly differentiates between binding instruments, i.e., regulations, directives, and decisions, and non-binding instruments, such as recommendations and opinions. Technical standards, in their conventional articulation, generally fall into the latter category, aligning with the characteristics of soft law[312]. This designation implies a normative framework that, while not directly enforceable through judicial sanctions, nevertheless wields influence by shaping expectations and guiding behaviour. However, to confine technical standards solely to the realm of non-binding soft law would be to oversimplify their operational reality. The distinction between these categories is far from absolute[313]. Soft law instruments, including technical standards, can indeed engender tangible legal effects. Principles of legitimate expectations and equal treatment, cornerstones of EU administrative law, can imbue soft law with a quasi-legal force, compelling adherence by administrative bodies and influencing judicial interpretation. Furthermore, technical standards frequently function as benchmarks in administrative and judicial review processes,

---

310 Eliantonio and Cauffman (n 19).

311 Oana Andreea Stefan and others, 'EU Soft Law in the EU Legal Order: A Literature Review' (Social Science Research Network, 4 March 2019) <https://doi.org/10.2139/ssrn.3346629> accessed 4 January 2026; Fabien Terpan, 'Soft Law in the European Union—The Changing Nature of EU Law' (2015) 21 European Law Journal 68 <https://doi.org/10.1111/eulj.12090>.

312 Hans-W Micklitz, 'Soft Law, Technical Standards and European Private Law' *Research Handbook on Soft Law* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781839101939/book-part-9781839101939-19.xml> accessed 4 January 2026.

313 Annalisa Volpato, 'The Legal Effects of Harmonised Standards in EU Law: From Hard to Soft Law, and Back?' *The Legal Effects of EU Soft Law* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781802208917/book-part-9781802208917-16.xml> accessed 5 November 2023.

effectively serving as yardsticks against which actions and products are evaluated for compliance and conformity[314].

The integration of technical standards into the formal architecture of EU law is most prominently manifested through European Harmonised Standards (hEN). These standards are meticulously crafted by ESOs, operating as private entities, yet fulfilling a public function delegated by the EU. Their purpose is to operationalise the 'essential requirements' stipulated by EU legislation across a spectrum of sectors. The critical juncture in their legal ascendance occurs when the European Commission formally publishes references to hEN in the Official Journal of the European Union. This act of publication bestows upon these standards a 'presumption of conformity'. This presumption is of paramount practical significance: economic operators who adhere to hEN are afforded a legally defensible position, presumed to be in compliance with the mandatory essential requirements established by EU law. The EU Commission, in its role as a regulatory architect, strategically employs 'standardisation requests' to assign specific technical tasks to these private organisations. However, the ostensible intent is to maintain a clear separation between technical specification and political discretion. These requests are ostensibly designed to be purely technical, avoiding any delegation of substantive political judgement to the ESOs, thereby maintaining the Commission's prerogative in policy formation.

The European Court of Justice has offered a significant jurisprudential pronouncement, stating that harmonised technical standards are to be considered "part of EU law" insofar as they constitute measures implementing or applying an act of EU law[315]. This judicial interpretation elevates the legal standing of hEN beyond mere voluntary guidelines, integrating them into the fabric of the EU legal order. However, this integration is not without its inherent tensions and paradoxes. The legal recognition of copyright protection for technical standards presents a notable contradiction with their purported status as "law."[316] Fundamental legal principles dictate that laws should be publicly accessible and readily available to all citizens. Conversely, technical standards, often developed and maintained by private entities, are frequently subjected to commercial copyright, necessitating payment for access. This creates a potential impediment to transparency and accessibility, values central to the rule of law and democratic accountability. Furthermore, the private genesis of technical standards may inadvertently create regulatory gaps. Private technical standards, falling outside the formal framework of hEN, may not be subject to the same rigorous judicial review mechanisms or the consumer protections afforded by other instruments[317].

Because of these issues, the precise legal status of harmonised standards remains a subject of ongoing debate, contingent on factors such as their incorporation into EU law through mechanisms like hEN, judicial interpretation by the ECJ, and their practical application in diverse contexts. Issues of accessibility, transparency, and, most fundamentally, legitimacy and accountability remain central to any critical assessment of their evolving role and pervasive impact within the European Union. The inherent

---

[314] Oana Andreea Stefan and others, 'EU Soft Law in the EU Legal Order: A Literature Review' (Social Science Research Network, 4 March 2019) <https://doi.org/10.2139/ssrn.3346629> accessed 17 August 2025.

[315] Carlo Colombo and Mariolina Eliantonio, 'Harmonized Technical Standards as Part of EU Law: Juridification with a Number of Unresolved Legitimacy Concerns?: Case C-613/14 James Elliot Construction Limited v. Irish Asphalt Limited, EU:C:2016:821' (2017) 24 Maastricht Journal of European and Comparative Law 323 <https://doi.org/10.1177/1023263X17709753>.

[316] Olia Kanevskaia, 'The Saga of Copyrighted Standards: A Perspective on Access to Regulation' *Research Handbook on Law and Technology* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781803921327/chapter20.xml> accessed 23 April 2024.

[317] Mariolina Eliantonio, 'Judicial Control of the EU Harmonized Standards: Entering a Black Hole?' (2017) 44 Legal Issues of Economic Integration <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\LEIE\LEIE2017022.pdf>.

tension between their function as quasi-regulatory instruments and their private genesis necessitates continuous scrutiny and a nuanced understanding of their place within the EU's multi-layered legal architecture.

5.2.2. Legitimacy issues of the European standardisation system

This section undertakes a multifaceted analysis of the legitimacy challenges inherent in the development and application of technical standards within the EU context, drawing upon the established conceptual framework of input, output, and throughput legitimacy[318], and further considering the perspectives offered by private law[319]. This exploration reveals a complex and often contested terrain, where concerns regarding democratic accountability, procedural fairness, and substantive quality intertwine with the exigencies of effective regulation and technological advancement[320].

Concerns regarding input legitimacy, which centres on the representative and participatory nature of rule-making processes, are particularly pertinent to the realm of technical standardisation. Input legitimacy, in essence, assesses the extent to which the EU's regulatory outputs are responsive to citizen concerns through mechanisms of public participation and representation. In the context of standard setting, several factors give rise to apprehension. A fundamental challenge stems from the inherently international and often private character of many standard-setting bodies. International standardisation is often pursued through private forums composed primarily of industry practitioners, raising questions regarding balanced geographical representation and the inclusion of broader societal interests. This is acutely evident in the realm of ICT, where pivotal standards for the internet and the World Wide Web are frequently formulated within private ICT forums and consortia, operating largely outside the formal participatory and transparency requirements typically associated with the ESOs. This detachment from established democratic procedures raises concerns about the representativeness and accountability of these influential standard-setting processes.

Moving beyond input considerations, an analysis of output legitimacy focuses on the substantive outcomes of standardisation processes, primarily evaluating policy outputs against criteria of efficiency and efficacy. While the reliance on technical expertise within the European standardisation organisation is intended to enhance the scientific robustness and hence the output legitimacy of its standards, weaknesses can emerge that potentially undermine its legitimacy. These include concerns regarding the transparency of expert selection procedures, a lack of openness in the phases leading to the formulation of scientific reports, and the absence of remuneration for experts. This latter point raises critical questions regarding the potential for bias and the erosion of expert neutrality, especially when participation becomes reliant on voluntary contributions, potentially favouring experts with institutional backing or vested interests. These shortcomings, even within a system designed to prioritise efficacy, underscore the inherent challenges in achieving robust output legitimacy.

Throughput legitimacy, the third facet of this analytical framework, directs attention to the procedural dimensions of standard development, scrutinising the design of the process itself. It evaluates the efficacy, accountability, transparency, inclusiveness, and openness to consultation inherent within the standard-setting process. In this domain, the judicial scrutiny exercised by the CJEU offers a particularly nuanced

---

[318] Vivien A Schmidt, 'Democracy and Legitimacy in the European Union Revisited: Input, Output and "Throughput"' (2013) 61 Political Studies 2 <https://doi.org/10.1111/j.1467-9248.2012.00962.x>.

[319] Rodrigo Vallejo, 'The Private Administrative Law of Technical Standardization' (2022) 40 Yearbook of European Law 172 <https://doi.org/10.1093/yel/yeab011>.

[320] Tartaro, 'Regulating by Standards' (n 20).

perspective. The CJEU appears to subject harmonised European standards, which are formally non-binding, to closer legal scrutiny compared to other standards incorporated by reproduction into EU law, which carry a binding character. This differential judicial oversight suggests a sensitivity to the procedural legitimacy of harmonised standards, even in their formally non-binding guise, potentially reflecting a recognition of their practical normative impact. Furthermore, the precision and scope of the European Commission's standardisation requests to ESOs are crucial determinants of throughput legitimacy. These requests, alongside overarching regulations and sectoral legislation, delineate the tasks assigned to ESOs. To ensure procedural legitimacy, these requests must be carefully crafted to avoid delegating political discretion to private standard-setting bodies. The Commission's mandate should be circumscribed to technical tasks, preventing the encroachment of political decision-making into ostensibly technical processes[321].

Several factors are identified as significantly influencing the overall legitimacy of technical standards. Enhanced transparency and broader public participation are widely recognised as potentially mitigating information costs and the social costs associated with undue industry lobbying[322]. Greater openness can foster a more balanced representation of interests and contribute to a more robustly legitimate standard-setting process. However, the inherent hybridity embedded within standardisation requests from the Commission to ESOs remains a persistent challenge. While the legal nature of these requests as instruments of delegation is relatively clear, the content often exhibits a hybrid character, blending technical specifications with implicit policy choices. To safeguard legitimacy, standardisation requests should rigorously avoid any delegation of political discretion, ensuring they remain firmly anchored in the realm of technical expertise and preclude the surreptitious introduction of political considerations into ostensibly technical mandates.

In conclusion, the legitimacy of technical standards within the European Union is a multifaceted and continuously evolving issue. Concerns spanning input, output, and throughput legitimacy arise from the inherent characteristics of standard-setting processes, including the significant involvement of private actors, the technical complexity of the subject matter, and the potential for standards to be influenced by specific sectoral or commercial interests. Navigating this complex terrain requires ongoing vigilance and a commitment to enhancing transparency, participation, and accountability within standard-setting processes, while simultaneously recognising the crucial role of technical standards in facilitating effective regulation and fostering innovation within the EU. The inherent tension between the pursuit of technical efficacy and the imperative of democratic legitimacy remains a central challenge in ensuring that technical standards serve the broader public interest within the European Union and beyond.

5.2.3. Further issues in standard development in support of EU policy and legislation

While the issues analysed above concern legal challenges identified by legal scholars and jurisprudence, other practical concerns have been raised by standards practitioners. The ETSI Technical Report (TR) 103 880[323] illustrates these problems, focusing particularly on a confluence of factors that impede the timely and effective production of standards deemed acceptable by the EC for citation in the

[321] Pierluigi Cuccuru, 'Regulating by Request: On the Role and Status of the Standardisation Mandate under the New Approach' in Mariolina Eliantonio and Caroline Cauffman (eds), *The Legitimacy of Standardisation as a Regulatory Technique* (Edward Elgar Publishing 2020) <https://doi.org/10.4337/9781789902952.00008> accessed 5 June 2024.

[322] Michael Faure and Niels Philipsen, 'Standardisation from a Law and Economics Perspective' in Mariolina Eliantonio and Caroline Cauffman (eds), *The Legitimacy of Standardisation as a Regulatory Technique* (Edward Elgar Publishing 2020) <https://doi.org/10.4337/9781789902952.00013> accessed 4 April 2024.

[323] ETSI, 'Study into the Challenges of Developing Harmonised Standards in the Context of Future Changes to the Environment in Which Products Are Being Developed and Operated'.

OJEU. These impediments, as meticulously documented in the ETSI TR, span procedural inefficiencies, evolving interpretations of regulatory requirements, and complexities arising from the expanding scope of standardisation itself. This analysis will dissect these core issues, critically evaluating their implications for ETSI, the broader European standardisation landscape, and the competitiveness of European industries.

A central tenet of the ETSI report is the critical importance of timeliness in the development and citation of harmonised standards. The report unequivocally states that delays in this process directly translate into increased costs for manufacturers and impede the velocity of innovation cycles. This assertion resonates with established economic principles concerning the costs of regulatory uncertainty and delayed market entry. The absence of readily available harmonised standards creates a regulatory lacuna, necessitating provisional compliance strategies and potentially hindering investment in new product development. This situation is particularly acute in dynamic technological sectors where rapid iteration and time-to-market are paramount. The report highlights that these delays are not merely procedural inconveniences but represent tangible economic burdens that can erode the competitive edge of European enterprises.

Furthermore, the report draws attention to the escalating complexity inherent in defining functional requirements, especially in the increasingly critical domains of software and security. This escalation is partly driven by legislative mandates such as the essential requirements in the Radio Equipment Directive and emerging regulations like the AI Act, which demand more nuanced and comprehensive standards beyond traditional physical attributes. The transition from evaluating primarily physical characteristics to encompassing functional and performance-based criteria, particularly in software-driven systems, introduces significant methodological and philosophical challenges. The report correctly identifies the inherent difficulty in testing and verifying "subjective" requirements and, more problematically, in demonstrating the absence of vulnerabilities, particularly in complex systems potentially leveraging artificial intelligence. The very notion of "proof of negatives" in cybersecurity, as the report notes, is fundamentally unattainable. Security analysis can only probabilistically assess risk mitigation strategies rather than providing absolute guarantees of invulnerability. This tension between the aspirational goals of comprehensive security legislation and the inherent limitations of technical verification poses a considerable hurdle for standardisation bodies.

A particularly contentious issue, as highlighted by the report, revolves around the EC's evolving interpretation of "legal certainty." This concept, ostensibly influenced by the legal ramifications of the Elliott case, has introduced a new layer of scrutiny to harmonised standards, extending beyond their technical adequacy to encompass their legal robustness. This shift has resulted in significant delays in the citation process, as standards are now subjected to a level of legalistic examination that some stakeholders perceive as disproportionate to their primary function as technical specifications. While the pursuit of legal certainty is undoubtedly a legitimate objective for regulatory frameworks, the report suggests a potential disconnect between the EC's current approach and the practical realities of standards development. Harmonised standards are, at their core, technical documents developed through consensus-based processes involving diverse technical expertise. Imposing excessively stringent legalistic criteria risks undermining their global relevance and usability, potentially prioritising legalistic precision over technical practicality. This tension necessitates a careful recalibration of expectations and a clearer articulation of the precise legal requirements that harmonised standards are expected to fulfill.

Process inefficiencies within the standardisation approval mechanisms further compound these challenges. The report identifies the EN Approval Procedure, specifically the "all or nothing" acceptance/rejection paradigm for the standardisation requests within a constrained timeframe, as

inherently problematic. This inflexibility, coupled with the absence of mechanisms for iterative feedback or conditional acceptance, can lead to significant bottlenecks and rework. Late-stage comments from the EC during the EN approval process exacerbate these issues, disrupting established timelines and potentially necessitating substantial revisions to already developed standards. Moreover, the report raises concerns about the Harmonised Standards Consultant process, where standards deemed acceptable by HAS consultants are subsequently rejected by the EC. This inconsistency undermines the perceived value of the HAS process and creates uncertainty regarding the ultimate criteria for EC acceptance, potentially diverting resources and effort towards navigating perceived EC preferences rather than focusing solely on technical excellence and consensus building.

Inter-organisational dynamics also emerge as a critical factor influencing the efficacy of harmonised standards development. The report highlights tensions in the relationship between ETSI and the EC, particularly concerning the level of detail specified in SRs. The EC's tendency to delve into granular technical requirements, traditionally considered within the purview of ESOs like ETSI, is perceived by some stakeholders as an encroachment on ESO's technical domain and potentially indicative of a lack of trust in their technical expertise. This perception, coupled with instances of perceived changing of minds by the EC regarding previously accepted standards, contributes to an atmosphere of uncertainty and potentially undermines collaborative efforts. The report implicitly calls for a clearer delineation of roles and responsibilities, fostering a more collaborative and less prescriptive relationship between the EC and ESOs, predicated on mutual respect for their respective expertise and mandates.

Furthermore, the report acknowledges the increasing need for enhanced collaboration not only within ETSI itself, across various Technical Committees and Industry Specification Groups, but also with other ESOs such as CEN and CENELEC. The convergence of radio and non-radio technologies necessitates a more integrated and holistic approach to standardisation. However, the report points to limitations in inter-connectivity which may hinder effective cross-organisational collaboration. Addressing these structural and procedural limitations is crucial for ensuring that standardisation efforts keep pace with the evolving technological landscape and effectively address the increasingly complex and interdisciplinary nature of modern products and systems.

Finally, the report touches upon the broader implications of the EU's new Standardisation Strategy, including concerns about potential undue influence from foreign actors and the role of national delegates in decision-making processes. These governance considerations reflect a growing awareness of the strategic importance of standardisation in shaping global technological trajectories and safeguarding European interests. The exclusion of ETSI from certain SRs, such as the one related to the AI Act[324], serves as a concrete example of the potential impact of these evolving strategic considerations. Additionally, the ongoing discussion around access to standards for SMEs, a key focus of the EU strategy, underscores the need to ensure equitable access to the standardisation process and its outputs, preventing the creation of barriers to entry for smaller businesses and fostering broader participation in the standardisation ecosystem.

The ETSI TR 103 880 provides a valuable and insightful assessment of the challenges confronting harmonised standards development within the European context. The report meticulously documents a range of interconnected issues, spanning procedural inefficiencies, evolving regulatory interpretations, and complex stakeholder dynamics. These challenges, if left unaddressed, pose a significant threat to the timely availability of harmonised standards, potentially hindering innovation, increasing costs for

[324] Marta Cantero Gamito, 'The Role of ETSI in the EU's Regulation and Governance of Artificial Intelligence' (2024) 37 Innovation: The European Journal of Social Science Research 1425 <https://doi.org/10.1080/13511610.2024.2349627>.

European industries, and ultimately eroding European competitiveness in the global market. The report's recommendations, focusing on enhanced collaboration, process optimisation, and clearer communication between ETSI and the EC, offer a constructive roadmap for mitigating these challenges and fostering a more efficient and effective standardisation ecosystem. Moving forward, a concerted and collaborative effort involving the EC, ESOs, and industry stakeholders is essential to navigate these complexities and ensure that the European standardisation system remains robust, responsive, and fit for purpose in a rapidly evolving technological and regulatory landscape. Further research could explore specific mechanisms for enhancing EC-ESO collaboration, optimising the EN approval process, and addressing the challenges of legal certainty and functional requirements verification in the context of increasingly complex and software-defined systems.

6. The standardisation strategy and the future of European standardisation

While the European standardisation system has had and will continue to have a crucial role in EU policy and legislation, the issues identified above place it at a crossroads[325]. Interventions are necessary to strengthen its legitimacy and improve its overall effectiveness. The European Union's Strategy on Standardisation delineates a vision to achieve this objective[326].

The European Union's strategic approach to standardisation explicitly positions standards as foundational instruments for realising a resilient, ecologically sustainable, and digitally advanced single market, simultaneously seeking to bolster the global stature of the ESS. This system, demonstrably productive over preceding decades through the generation of over 3,600 harmonised standards facilitating legal compliance, interoperability, safety, and environmental protection, nevertheless confronts a critical juncture. The perceived underappreciation of standards' strategic significance has demonstrably impacted the Union's erstwhile leadership in global standard-setting arenas. Consequently, the EU strategy constitutes a comprehensive response aimed at reinvigorating the ESS to meet contemporary and future exigencies. Standards operate as indispensable infrastructural elements within the single market, establishing equitable conditions for enterprises and enhancing consumer trust; however, the operational environment is increasingly defined by global competition, wherein rival jurisdictions leverage standardisation assertively for market access and technological diffusion. Europe's competitiveness, technological sovereignty, capacity to mitigate strategic dependencies, and the safeguarding of its fundamental values are intrinsically linked to the efficacy of European stakeholders within international standardisation bodies. Achieving this necessitates the cultivation of robust standardisation competencies across industrial and academic sectors, coupled with a systemic transformation towards greater agility, flexibility, and anticipatory capacity to expedite high-quality standard delivery. The prevailing system's frequent inability to produce timely standards for nascent and disruptive technologies results in the forfeiture of critical 'first-mover' advantages, undermining strategic positioning.

A resilient single market and the successful twin, green and digital, transition of European industry are contingent upon a standardisation system meticulously aligned with overarching EU policy priorities. European standards are paramount for attaining climate neutrality, fostering circularity, and building

[325] Pierre Larouche and Justus Baron, 'The European Standardisation System at a Crossroads' (4 May 2023) <https://doi.org/10.2139/ssrn.4466316> accessed 9 April 2024.

[326] European Commission, 'An EU Strategy on Standardisation - Setting Global Standards in Support of a Resilient, Green and Digital EU Single Market' <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0031>.

economic resilience, encompassing domains such as testing methodologies, management systems, and interoperability solutions. Digital leadership equally hinges on the Union's capacity to mould international standards for digital products, processes, and services into globally recognised benchmarks. Failure to secure this influence risks the subordination of EU policy ambitions to standards defined by external regions. Addressing this, the Commission identifies critical standardisation urgencies demanding immediate attention to preclude strategic dependencies and affirm EU leadership in pivotal green and digital technologies. These exigencies encompass standards essential for overcoming obstacles in COVID-19 vaccine and medicine production, supporting the recycling of critical raw materials (CRM), enabling the clean hydrogen value chain, facilitating low-carbon cement production due to its significant emissions-saving potential, ensuring chip security, authenticity, and reliability certification, and enhancing data interoperability, sharing, and re-use to underpin Common European Data Spaces. According to the Strategy, to meet these immediate challenges and anticipate future requirements, the Commission is enacting a suite of measures. Immediate action on identified urgencies involves launching standardisation requests, engaging stakeholder communities, and providing targeted financing, while mandating ESOs to prioritise swift delivery. Furthermore, a High-Level Forum has been established, uniting representatives from Member States, ESOs, National Standardisation Bodies, industry, civil society, and academia. This Forum advises on priority-setting and future needs, coordinate the effective representation of European interests internationally, and ensure ESS activities support a greener, digital, fair, and resilient EU economy, including bolstering technical expertise and facilitating political concertation through an annual high-level event.

The strategy also aims to strengthen the position of the ESOs while simultaneously increasing their legitimacy and accountability. The privileged status of ESOs (CEN, CENELEC, ETSI) under Regulation 1025/2012, as the sole entities eligible for Commission standardisation requests, carries significant responsibilities, especially as standards increasingly embody core EU democratic values, interests, and socio-ecological principles beyond mere technical specifications. Concerns persist, however, regarding disproportionate corporate influence within ESO decision-making processes, particularly at ETSI. To uphold the integrity, inclusiveness, and accessibility of the ESS, the Commission advocates for strengthened administrative and good governance principles when ESOs handle European requests and develop compliance standards. Key measures include a legislative proposal amending Regulation 1025/2012, as discussed above, to ensure delegates of National Standardisation Bodies hold decisive power throughout the development of Commission-requested standards, thereby enhancing openness, transparency, and inclusiveness through balanced representation, including societal stakeholders. ESOs are called upon to propose governance modernisation plans ensuring full representation of public interest, SMEs, civil society, and users, and facilitating access to standards, potentially including free access. Insufficient progress may trigger a regulatory revision. Complementing this, a peer review process among Member States and National Standardisation Bodies will disseminate best practices for SME-friendly conditions and broader societal involvement. Recognising the increasing use of common specifications as a fallback solution in recent legislation, the Commission will develop a horizontal framework for their criteria and processes to prevent fragmentation, particularly when standardisation is delayed or blocked.

Ultimately, standards serve as instrumental tools embedded within broader policy imperatives encompassing industrial competitiveness, the free movement of goods, innovation, safety, consumer and worker protection, environmental sustainability, open strategic autonomy, and the transition to a climate-neutral, resilient, circular economy. Ensuring the usability, effectiveness, and accessibility of standards across value chains, particularly for SMEs and societal actors, is paramount. The EU's ambition is to

function as a first-mover vanguard, leading international standards-setting through cooperation with like-minded partners in strategic future technologies, committed to rendering the European standardisation system more functional and agile to deliver standards enhancing competitiveness, serving the public interest, promoting sustainability, and reinforcing democratic values. Transparency throughout the standards-making process is crucial for removing bottlenecks and enhancing efficiency, enabling public and private actors to better identify gaps and future requirements. The sustained engagement and contribution of all relevant stakeholders, inter-institutional partners, ESOs, civil society, industry, academia, coupled with robust checks and balances, remain fundamental to the system's efficacy. The publication of an annual dashboard detailing planned, ongoing, and completed standardisation activities alongside key planning documents aims to augment systemic transparency. This multifaceted strategy is thus designed to reassert the EU's position as a global frontrunner in standards development, underpinning its core values and furnishing European industries with a vital competitive edge.

7. Conclusion of the Chapter

This chapter has provided the necessary institutional and legal topography of the European standardisation system, establishing the structural context in which the regulation of Artificial Intelligence is legally situated. By tracing the historical trajectory from the rigidities of the Old Approach to the flexible modularity of the New Legislative Framework, the analysis has illuminated the strategic division of labour that defines the European single market: the reservation of essential public interest requirements to the legislator, and the delegation of detailed technical execution to private standardisation organisations.

The discussion highlighted that this regulatory architecture is not merely a technical facilitation mechanism but a sophisticated governance system underpinned by the "presumption of conformity." As demonstrated, this presumption serves as the critical legal interface connecting voluntary technical specifications with binding legislative mandates. However, our examination of the operational realities, ranging from the role of notified bodies to the governance of ESOs like ETSI, reveals a system currently grappling with profound internal and external pressures. While the framework has historically succeeded in integrating product markets, its application to complex, rapidly evolving technologies has exposed significant fissures. We observed that the "soft law" character of standards is increasingly contested, with CJEU jurisprudence (such as the James Elliott case) and Commission policies progressively treating harmonised standards as part of the corpus of EU law. This shift has precipitated a crisis of legitimacy and efficiency, creating friction between the demand for legal certainty and the practical realities of technical consensus-building.

Furthermore, the analysis of recent evaluations and the 2022 Standardisation Strategy underscores that the system is at a crossroads. The findings indicate that the traditional machinery of standardisation is struggling to keep pace with the velocity of the digital transition, hampered by procedural delays, fragmented oversight of conformity assessment bodies, and unresolved tensions regarding the representation of societal interests versus industry dominance. The friction between the European Commission's assertive political steering, aimed at ensuring standards align with European values and strategic autonomy, and the technical autonomy of bodies like ETSI highlights the difficulty of reconciling democratic accountability with the efficiency required for global competitiveness.

Having thus mapped the general blueprint of the New Legislative Framework and identified its inherent vulnerabilities, the research is now positioned to apply this context to the specific case of

Artificial Intelligence. The following chapters will transition from this broad institutional analysis to a focused examination of the AI Act itself. We will explore how this specific piece of legislation attempts to adapt the general mechanisms of the NLF described here to the unique idiosyncrasies of AI, and whether the standardisation system, with all the imperfections and challenges identified in this chapter, is capable of bearing the regulatory weight of high-risk artificial intelligence.

Chapter 4.
Regulating AI in the European Union

1. Aim of the Chapter

While the previous chapter provided a broader analysis of the European Standardisation System, the present chapter discusses how the AI Act fits within that system. The AI Act, proposed in 2021 and entered into force in August 2024, stands as the world's first comprehensive regulation for artificial intelligence. This legislation is the culmination of an extended process that has progressively made AI a central focus for European regulatory efforts. This chapter begins by tracing the key stages, starting from 2017, that led to increasing calls for AI regulation across Europe. This trajectory involved a critical assessment of AI's growing impact on society. Concerns arose regarding fundamental rights, data privacy, potential market distortions, and the ethical implications of autonomous decision-making. As AI technologies became more widespread across various sectors, a coordinated legislative response became necessary to mitigate potential harms and promote trustworthy innovation, thus preparing the ground for this regulatory action.

Following this historical overview, the analysis shifts to examine several core elements of the AI Act. Given the regulation's extensive and complex nature, this investigation focuses specifically on certain key areas: the material scope (defining which AI systems and applications are covered), the framework for risk categorisation (establishing different requirements based on potential harm), the essential requirements for high-risk AI systems (such as rules for data quality, technical robustness, human oversight, and transparency), and the various pathways for compliance. While important topics like regulatory sandboxes and rules for General Purpose AI systems exist, they are not included in this discussion as they fall outside the specific objectives of this dissertation.

An additional part of the analysis then concentrates on the crucial role of harmonised standards, conformity assessment procedures and the accredited notified bodies that perform them. These mechanisms are vital for ensuring that AI systems available in the EU market or put into service strictly meet the required standards. This helps build user trust and maintain market integrity. The accredited notified bodies act as independent third-party evaluators, tasked with confirming that high-risk AI systems comply with the rules before they are deployed. Their involvement is essential for upholding the strict safety and ethical standards envisioned by the legislation, turning legal principles into practical requirements.

Finally, this chapter addresses some of the unresolved complexities and open questions that continue to challenge the full implementation and interpretation of the regulation. These issues include, but are not limited to, the detailed aspects of defining AI and differentiating between various levels of autonomy in real-world applications, the evolving understanding of unacceptable risk in a rapidly changing technological landscape, and the practical challenges of enforcing rules across national borders and ensuring international consistency of regulatory standards. Additionally, questions remain about the framework's ability to adapt to rapid technological advancements, the sufficiency of resources for effective market surveillance, and the precise mechanisms for seeking compensation in cases where AI causes harm, particularly concerning who holds responsibility. The ongoing discussion surrounding these

ambiguities highlights the dynamic nature of AI governance and the continuous need for adapting and refining regulatory tools.

2. Making AI a regulatory target in the EU

2.1 The EESC opinion (2017)

The European Economic and Social Committee's (EESC) opinion on the ramifications of artificial intelligence across the digital single market, production, consumption, employment, and society[327] constitutes the first significant contribution to bring artificial intelligence to the centre of the discussion in European institutions. Predicated on the fundamental assertion that while AI harbours substantial potential for societal and economic advancement, its swift evolution and widespread adoption simultaneously engender considerable ethical, safety, societal, and economic exigencies, the EESC positions itself as a proactive agent in shaping the trajectory of AI development. This proactive stance is underscored by the fact that this opinion was an own-initiative undertaking, demonstrating a self-directed engagement with a topic deemed to be of paramount importance for the future of Europe. The EESC's intention to actively monitor AI developments from a multifaceted perspective, encompassing not only technical aspects but also critical ethical, safety, and societal dimensions, reflects a holistic approach to understanding the complex implications of this technology.

The EESC's role as a pivotal actor in structuring and promoting the public debate on AI, representing the diverse interests of European civil society, highlights the Committee's commitment to fostering a broad and inclusive dialogue. By aiming to involve a wide spectrum of stakeholders, ranging from policymakers and industry representatives to social partners, consumers, and non-governmental organisations, the EESC seeks to cultivate an informed and balanced understanding of AI, thereby mitigating the risks of both unwarranted alarmism and complacent relativism. This commitment to inclusivity is crucial, as the societal impact of AI will invariably affect a wide array of actors, necessitating a collaborative approach to governance and regulation. Recognising the inherently transnational nature of AI's influence, the EESC rightly emphasises the imperative for the establishment of supra-national policy frameworks. The specific recommendation for the European Union to assume a global leadership role in formulating clear global policy frameworks for AI, grounded in European values and fundamental rights, underscores the EESC's conviction that a coordinated international effort is essential to effectively address the challenges and harness the opportunities presented by AI. This call for global leadership aligns with the EU's historical role in setting international standards and norms in various domains.

The EESC's identification of eleven key areas where AI presents societal challenges provides a comprehensive overview of the potential disruptions and transformations that this technology may precipitate. These areas, spanning ethics, safety, privacy, transparency and accountability, work, education and skills, (in)equality and inclusiveness, law and regulations, governance and democracy, warfare, and superintelligence, demonstrate the breadth and depth of the EESC's analysis. By delineating these specific domains, the EESC provides a structured framework for further discussion and policy intervention. The core principles and recommendations put forth by the EESC offer concrete pathways for navigating these challenges. The emphasis on a "human-in-command" approach, asserting the primacy of human

[327] 'Artificial Intelligence - The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society (Own-Initiative Opinion) | EESC' (11 October 2016) <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and> accessed 19 August 2025.

control over AI systems, reflects a deep-seated concern for maintaining human agency in an increasingly automated world. This principle serves as a foundational ethical guideline, underpinning many of the EESC's subsequent recommendations. The call for a comprehensive code of ethics for the development, application, and use of AI underscores the need to proactively embed human values and fundamental rights into the very fabric of AI systems. This is particularly salient given the potential for AI to perpetuate and even amplify existing societal biases if ethical considerations are not explicitly addressed.

Furthermore, the EESC's advocacy for a robust standardisation system for verifying, validating, and monitoring AI systems highlights the critical need for ensuring the safety, transparency, comprehensibility, accountability, and ethical alignment of these technologies. Such a standardisation system would be instrumental in building public trust and fostering the responsible deployment of AI. The recommendation for the development of a European AI infrastructure, characterised by open-source, privacy-respecting learning environments, real-life test environments, and high-quality data sets, underscores the strategic importance of fostering European capabilities in this critical technological domain. The EESC's recognition of the competitive advantage inherent in responsible European AI systems with established certification and labels suggests a forward-thinking approach that seeks to leverage ethical considerations as a source of economic strength.

The EESC's detailed consideration of the impact of AI on the world of work reflects a keen awareness of the potential for widespread economic and social disruption. The call for joint efforts by the EU, national governments, and social partners to identify affected job sectors and address the multifaceted consequences for employment, the nature of work, social systems, and inequality is crucial. The recommendation to invest in job sectors less susceptible to AI-driven automation and to support the co-creation of complementary AI systems, such as human-machine teams, demonstrates a nuanced understanding of the evolving relationship between humans and machines in the workplace. The emphasis on investing in education and training to equip individuals with the skills necessary to work alongside AI and to cultivate uniquely human capabilities underscores the importance of proactive adaptation to the changing demands of the labour market. The EESC's cautious stance against granting legal personality to robots or AI is a significant point of contention within the broader debate, reflecting concerns about the potential erosion of established legal principles, the creation of moral hazard, and the increased risk of abuse. This position aligns with a more anthropocentric view of legal responsibility and accountability.

The EESC's acknowledgment of the lack of a universally accepted definition of AI and its categorisation into narrow and general AI provides a useful conceptual framework for understanding the current state of the field and its potential future trajectories. The explanation of the recent advancements in narrow AI, attributed to increased computing power, large datasets, and the development of machine learning and deep learning, offers valuable context for appreciating the rapid progress in this domain. The detailed elaboration of the opportunities and threats associated with AI across various sectors further strengthens the EESC's analysis. The identification of specific opportunities in areas such as sustainable agriculture, safer transport, and medicine, alongside the detailed discussion of ethical, safety, transparency, privacy, and governance challenges, provides a balanced perspective on the transformative potential and inherent risks of AI. The EESC's recognition of the potential for AI to exacerbate existing inequalities and the call for exploring potential mitigating measures, such as an AI tax or basic income, demonstrates a commitment to addressing the broader societal implications of this technology.

The EESC's opinion offers a comprehensive and insightful analysis of the multifaceted consequences of artificial intelligence, and anticipates many of the themes that would be central to the European discussion until the AI Act. The emphasis on ethical considerations, the call for robust AI standardisation,

and the focus on a human-centric approach are all themes that remain present in European AI policy discussions, as will be explored further below.

2.2 The AI for Europe strategy & the Coordinated plan on artificial intelligence (2018)

Following the opinion of the EESC, AI became a topic of policy discourse at EC level with the Communication outlining the "Artificial Intelligence for Europe" strategy[328] and a subsequent "Coordinated Plan on Artificial Intelligence."[329] A central tenet pervading these documents is, once again, the recognition of AI as a transformative force, possessing the capacity to reshape both societal structures and economic paradigms on a scale comparable to historical technological revolutions such as the advent of the steam engine and electricity. This perceived potential underpins the European Union's ambition to not merely participate in the global AI landscape, but to actively shape its trajectory, establishing itself as a world leader in the development and deployment of AI that is simultaneously ethical, secure, and fundamentally human-centric.

The strategic imperative for a coordinated European approach emerges as a dominant theme across these policy documents. The European Commission posits that the realisation of this ambitious vision necessitates a concerted effort, transcending the capabilities of individual Member States acting in isolation. This emphasis on collaboration underscores the understanding that only through the pooling of resources, the alignment of strategic priorities, and the fostering of synergies can the EU effectively harness the transformative power of AI. The coordinated plan, in particular, serves as a blueprint for such collaborative action, identifying key areas demanding focused attention. These areas encompass the crucial dimensions of investment in AI research and innovation, the cultivation of centres of excellence and the widespread diffusion of AI technologies across the European economic fabric, the establishment of robust data infrastructures to fuel AI development, the proactive engagement with the societal challenges and opportunities presented by AI, the embedding of ethical considerations into the very fabric of AI systems, and the establishment of a regulatory framework that is both enabling and protective.

The concept of "human-centric AI" occupies a prominent position within the EU's strategic discourse. This approach reflects a deliberate choice to prioritise human well-being, fundamental rights, and ethical values in the development and application of AI technologies. It signifies a commitment to ensuring that AI serves humanity, contributing to societal progress and individual flourishing, rather than being pursued solely for the sake of technological advancement or economic gain. Principles such as "ethics by design" and "security by design" are presented as crucial operationalisations of this human-centric vision, advocating for the integration of ethical and legal considerations, as well as robust cybersecurity measures, from the earliest stages of AI system design and development. This proactive stance contrasts with potentially reactive approaches that might address ethical and security concerns only after technologies have been deployed at scale.

The EU's policy documents articulate a multi-pronged strategy to achieve its AI ambitions. A significant focus is placed on bolstering Europe's technological and industrial capacity in AI. This involves ambitious investment targets, aiming to mobilize substantial public and private funding to support AI research, development, and deployment. The establishment of networks of AI research excellence centres, the creation of world-reference testing facilities, and the strengthening of Digital

[328] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Artificial Intelligence for Europe 2018.

[329] Coordinated Plan on Artificial Intelligence 2018.

Innovation Hubs are proposed as concrete initiatives to foster innovation and accelerate the adoption of AI technologies, particularly among SMEs, which constitute a significant portion of the European economy. Recognising the critical role of data in the AI era, the development of a Common European Data Space is envisioned to facilitate access to and sharing of data, while respecting privacy and security considerations. Furthermore, the importance of nurturing talent in AI is highlighted, with plans for common PhD programs and initiatives to modernize education and training systems to equip the European workforce with the skills needed for an AI-driven future.

Synthesising these perspectives reveals a broad consensus within the EU institutions regarding the transformative potential of AI and the strategic imperative for a coordinated European response. The Commission's communications lay out an ambitious vision focused on technological leadership and economic competitiveness, underpinned by a commitment to ethical principles. Implicit within these discussions are potential tensions, such as the balancing act between fostering innovation and implementing regulations, or the practical challenges of translating ethical principles into concrete design and implementation practices. Furthermore, while the documents articulate a clear strategic direction, questions remain regarding the practical implementation of these ambitious plans, the effectiveness of the proposed measures in addressing the rapid advancements in AI, and the EU's ability to effectively compete in the global AI landscape while adhering to its core values. The 2018 policy documents thus represent a foundational moment in the EU's engagement with artificial intelligence, setting the stage for ongoing policy development, research, and societal debate as this transformative technology continues to evolve.

2.3 Building Trust in Human-centric AI & the Ethics Guidelines for Trustworthy AI (2019)

The articulation of a distinctly European approach to artificial intelligence, centrally framed by the principle of human-centricity, is further developed the EC Communication Building Trust in Human Centric Artificial Intelligence[330], promulgated in April 2019. The Communication posits trust as an indispensable precursor to realising an AI ecosystem dedicated to the augmentation of human well-being. This communication constructs a vision of AI as an instrument designed to serve humanity, predicated on a developmental trajectory that is simultaneously lawful, ethically grounded, and technically robust. The envisioned benefits, according to the Commission, span a considerable spectrum of societal and economic domains, from enhancing healthcare efficacy and optimising energy consumption to fortifying cybersecurity defenses and refining financial risk management protocols. Crucially, the document demonstrates an awareness of the inherent challenges accompanying AI proliferation, notably the transformation of labour markets and the emergence of intricate legal and ethical quandaries. Acknowledging risks stemming from data vulnerabilities, algorithmic biases, and potential operational flaws, the Communication proposes a triadic strategy. This approach involves bolstering the EU's technological and industrial capacities, anticipating and managing socio-economic transformations induced by AI, and establishing a pertinent ethical and legal architecture.

The foundation of Europe's approach to trustworthy AI is intrinsically linked to its established values and regulatory landscape. The Union's commitment to human dignity, equality, the rule of law, and fundamental rights, as enshrined in the EU Charter, provides a normative bedrock for AI development. Existing regulations, notably the GDPR, already mandate a high standard for personal data protection,

[330] 'Communication: Building Trust in Human Centric Artificial Intelligence | Shaping Europe's Digital Future' <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence> accessed 19 August 2025.

incorporating principles of data protection by design and default, and granting individuals rights concerning automated decision-making. Complementary legislation, such as the Free Flow of Non-Personal Data Regulation and the Cybersecurity Act, further contributes to a secure and reliable digital environment. However, the unique characteristics of AI, particularly its capacity for autonomous learning and decision-making, introduce novel challenges. The potential for AI systems to generate unreliable outcomes due to incomplete or biased data, or to be susceptible to malicious manipulation, underscores the imperative for a human-centric development paradigm. This perspective posits that AI applications must not only adhere to legal statutes but also align with profound ethical principles, designed to avert unintended harm and promote inclusivity by ensuring diversity in development and accessibility for persons with disabilities.

In response to these exigencies, the Commission convened a high-level expert group on AI, tasked with formulating comprehensive ethics guidelines that would apply uniformly across the internal market. The Ethics Guidelines for Trustworthy AI[331] are the main outcome of the high-level expert group on AI and constitute a seminal framework explicitly designed to promote and operationalise the concept of Trustworthy AI. This framework emerged through a deliberative process incorporating feedback from over five hundred contributors during an open consultation phase on its initial draft, reflecting a concerted effort towards inclusivity and broad societal input. Fundamentally, Trustworthy AI is conceptualised within these guidelines as resting upon three inextricably linked yet distinct pillars: compliance with all applicable laws and regulations (lawful), adherence to established ethical principles and values (ethical), and demonstrable resilience from both technical and social standpoints to mitigate unintentional harm (robust). Crucially, while each component is deemed necessary, none is considered sufficient in isolation; their synergistic interaction throughout an AI system's entire lifecycle is paramount. The guidelines explicitly acknowledge that while legal compliance constitutes a mandatory baseline, ethical adherence possesses the capacity to transcend existing legal strictures, thereby positioning the document's primary focus on securing the ethical and robust dimensions of AI development and deployment, rather than providing exhaustive legal guidance.

Building upon this tripartite foundation, Chapter I of the guidelines delineates the conceptual bedrock of Trustworthy AI, grounding it firmly in the Charter of Fundamental Rights of the European Union and international human rights law. This human-centric approach situates AI as a tool intended to serve humanity and the common good, enhancing human welfare and freedom. From these fundamental rights, four core ethical principles are derived, deemed essential for ensuring AI systems are both ethical and robust: *Respect for human autonomy*, mandating that systems augment rather than unjustifiably subordinate, coerce, or manipulate human decision-making, safeguarding meaningful human choice and oversight; *Prevention of harm*, requiring systems to be safe, secure, and technically robust, actively avoiding causing or exacerbating harm to human dignity, mental or physical integrity, or the natural environment, with heightened consideration for vulnerable populations; *Fairness*, encompassing both substantive dimensions (equitable distribution of benefits and burdens, freedom from unfair bias, discrimination, and stigmatization) and procedural dimensions (contestability and access to redress regarding AI decisions); and *Explicability*, deemed critical for fostering trust, necessitating transparency in processes, clear communication of capabilities and purpose, and explainability of decisions, with traceability and auditability serving as proxies where full explanation proves technically infeasible, provided fundamental rights are upheld. The guidelines astutely recognise that tensions may arise between these principles, advocating for the establishment of accountable deliberation mechanisms to resolve such conflicts, while

[331] AI HLEG, 'Ethics Guidelines for Trustworthy AI' <https://ec.europa.eu/futurium/en/ ai-alliance-consultation>.

simultaneously affirming the absolute nature of certain rights, such as human dignity, precluding their subjection to mere balancing exercises.

Transitioning from abstract principles to tangible implementation, Chapter II articulates seven key requirements instrumental in realizing Trustworthy AI throughout the system lifecycle, applicable to all stakeholders, developers, deployers, end-users, and broader society, each bearing distinct responsibilities. These requirements demand that AI systems ensure *human agency and oversight*, facilitating informed user decisions and incorporating appropriate control mechanisms (e.g., human-in-the-loop, -on-the-loop, or -in-command); exhibit *technical robustness and safety*, including reliable performance, resilience against attacks, security, accuracy, reliability, reproducibility, and established fallback plans; guarantee *privacy and data governance*, ensuring data protection and integrity from collection through processing and generation, governed by secure protocols; uphold *transparency*, achieved through traceability of data and processes, explainability of decisions, and clear communication regarding system interaction and limitations; promote *diversity, non-discrimination and fairness*, actively avoiding unfair biases, ensuring accessibility via universal design, and fostering inclusive stakeholder participation; contribute positively to *environmental and societal well-being*, considering impacts on society, democracy, other sentient beings, and the environment, encouraging sustainable development; and establish clear *accountability*, necessitating auditability of algorithms, data, and design processes, mechanisms for impact minimisation, reporting, and redress, and clear assignment of responsibility. The Guidelines further enumerate a non-exhaustive catalogue of technical methods (e.g., trustworthy architectures, ethics by design, explanation techniques, rigorous testing) and non-technical methods (e.g., regulation, standardisation, certification, education, governance frameworks) to fulfil these requirements, emphasising that achieving Trustworthy AI is an ongoing, iterative process demanding continuous evaluation and refinement.

Consequently, Chapter III seeks to operationalise the preceding framework through the provision of a concrete, albeit non-exhaustive, Trustworthy AI Assessment List. Primarily targeting developers and deployers, this practical tool purports to translate the key requirements of Chapter II into actionable assessment points. Acknowledging its developmental nature, the guidelines outline a piloting phase involving stakeholders to garner practical feedback, with a revised iteration anticipated subsequently. It is explicitly clarified that this assessment list does not constitute legal advice nor focus on legal compliance; its remit remains centred on evaluating and ensuring the ethical and robust dimensions of AI. Crucially, the guidelines stress that employing this list should transcend mere procedural "box-ticking," instead representing a continuous commitment to identifying requirements, evaluating solutions, and improving outcomes throughout the system lifecycle, inherently involving relevant stakeholders. Effective implementation is posited as requiring integration within the organisation's governance structure, engaging both operational and strategic management levels to ensure institutional acceptance and contextual relevance. Specific responsibilities are envisaged for diverse organisational roles, including management, legal/compliance, product development, quality assurance, human resources, procurement, and project managers. Ultimately, the Ethics Guidelines aspire to establish a foundational, horizontally applicable framework for Trustworthy AI, adaptable to specific contexts and use cases, serving as a catalyst for discourse on "Trustworthy AI for Europe" and fostering global deliberations on establishing robust ethical frameworks for artificial intelligence systems.

The Ethics Guidelines for Trustworthy AI has profoundly reconfigured the discourse surrounding artificial intelligence, serving as a pivotal catalyst for a global proliferation of similar documents[332]. These instruments meticulously delineate the aspirational principles that AI systems ought to embody to be

[332] Jobin, Ienca and Vayena (n 14); Corrêa and others (n 14).

deemed trustworthy, human-centric, transparent, or accountable. This widespread adoption of ethical frameworks underscores a global consensus on the imperative to integrate moral considerations into the development and deployment of advanced technological systems.

Nevertheless, concurrently with their salutary impact on public and professional awareness, these early normative frameworks have demonstrably contributed to maintaining the ethical debate at a predominantly high-level principled plane[333]. While the articulation of overarching principles such as fairness, beneficence, non-maleficence, and privacy is undeniably crucial for establishing a foundational moral compass, a substantial body of critical scholarship has underscored the inherent limitations of such an approach[334]. One pervasive critique highlights the profound gap between abstract ethical maxims and their practical operationalisation[335]. These guidelines often lack the requisite specificity to translate broad principles into actionable technical standards or enforceable legal obligations, thereby risking remaining largely ineffective in guiding real-world design choices and deployment scenarios[336].

Furthermore, critics frequently point to the non-binding nature of many of these recommendations, classifying them as forms of soft law that provide insufficient mechanisms for accountability or redress[337]. This inherent lack of enforceability creates a vacuum where compliance becomes discretionary, often leading to what has been termed "ethics washing", a superficial adherence to ethical rhetoric without a genuine commitment to fundamental changes in corporate practice or system design[338]. Such a dynamic can inadvertently obscure the deeper structural and power imbalances that AI systems might exacerbate, rather than compelling organisations to genuinely embed ethical considerations at every stage of the AI lifecycle.

Beyond issues of vagueness and enforcement, a significant body of critical analysis also scrutinises the implicit philosophical underpinnings of many prominent guidelines[339]. Concerns have been raised regarding a potential Western-centric bias in their conceptualization of ethics, which may not adequately account for diverse cultural values, epistemologies, or societal priorities across different global contexts[340]. Moreover, the emphasis on individual principles can, at times, overshadow the systemic impacts of AI, neglecting the collective and societal dimensions of harm or benefit, or the intricate interplay of power dynamics embedded within algorithmic systems. This focus on abstract principles, rather than on robust governance mechanisms, ongoing societal impact assessments, or participatory design processes, has been identified as a significant lacuna.

It is precisely in response to these pervasive limitations that the European discourse has demonstrably pivoted towards increasingly stringent forms of regulatory intervention. This strategic shift reflects a broader recognition that ethical considerations, to be genuinely transformative, must transcend voluntary

---

333 Alessio Tartaro and Mihály Héder, 'Ethics of AI' *Elgar Encyclopedia of Political Communication* (Edward Elgar Publishing 2025) <https://www.elgaronline.com/display/book/9781035301447/vol1.chapter124.xml> accessed 4 January 2026.

334 Brent Mittelstadt, 'Principles Alone Cannot Guarantee Ethical AI' (20 May 2019) <https://doi.org/10.2139/ssrn.3391293> accessed 27 November 2023.

335 Thilo Hagendorff, 'The Ethics of AI Ethics: An Evaluation of Guidelines' (2020) 30 Minds and Machines 99 <https://doi.org/10.1007/s11023-020-09517-8>.

336 Luke Munn, 'The Uselessness of AI Ethics' (2023) 3 AI and Ethics 869 <https://doi.org/10.1007/s43681-022-00209-w>.

337 Anaïs Rességuier and Rowena Rodrigues, 'AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics' (2020) 7 Big Data & Society 2053951720942541 <https://doi.org/10.1177/2053951720942541>.

338 Luciano Floridi, 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical' (2019) 32 Philosophy & Technology 185 <https://doi.org/10.1007/s13347-019-00354-x>.

339 Mihály Héder, 'A Criticism of AI Ethics Guidelines' (2020) 20 Információs Társadalom: Társadalomtudományi Folyóirat 57.

340 Thilo Hagendorff, 'Blind Spots in AI Ethics' [2021] AI and Ethics <https://doi.org/10.1007/s43681-021-00122-8> accessed 7 June 2022.

compliance and be enshrined within binding legal frameworks. This trajectory, indicative of a maturation in AI governance thinking, finds its apotheosis in the AI Act, representing a significant move from soft guidance to hard law[341] within the European Union's regulatory landscape, aiming to imbue ethical precepts with genuine enforcement capabilities and legal recourse

2.4 The White Paper on AI (2020)

The documents analysed thus far highlight the emergence of a European strategy on AI that follows a dual path: fostering an "ecosystem of excellence" on the one hand and cultivating an "ecosystem of trust" on the other. This ambition for global leadership, predicated on both innovation and ethical integrity, is reinforced by the *White Paper on Artificial Intelligence: a European approach to excellence and trust*[342].

The pursuit of an "ecosystem of excellence" builds upon Europe's established strengths in advanced research, a burgeoning start-up landscape, and leadership in robotics and manufacturing. Recognising the foundational role of high-performance computing and the vast, yet underutilised, reserves of public and industrial data, the strategy prioritises substantial investment to close the gap with global competitors. This is concretised through the Coordinated Plan on AI, envisioned as a dynamic framework for enhanced collaboration among Member States, directing investments, and fostering synergistic developments across the AI value chain. A critical dimension of this excellence agenda is the explicit integration of societal and environmental well-being, notably through the lens of the European Green Deal, urging AI solutions to intrinsically consider resource efficiency and environmental impact throughout their lifecycle.

In parallel to the pursuit of excellence, the European approach articulates a robust framework for an "ecosystem of trust," acknowledging the inherent risks associated with AI, which range from opaque decision-making and potential discrimination to privacy intrusions and misuse for illicit purposes. The Commission views the absence of trust as a significant impediment to the broader adoption of AI, necessitating a clear, harmonised European regulatory framework to pre-empt market fragmentation and ensure legal certainty. While existing EU legislation pertaining to fundamental rights, data protection (e.g., GDPR), consumer protection, product safety, and liability theoretically applies to AI, the distinctive characteristics of AI, such as its opacity, complexity, unpredictability, and evolving autonomy, pose considerable challenges to effective enforcement. This necessitates a careful assessment of whether current legislative instruments require adjustments or whether new, purpose-built legislation is warranted. The potential harms span both material dimensions (e.g., safety, physical injury) and immaterial ones (e.g., privacy violations, limitations on freedom of expression, discrimination in access to employment), underscoring the broad societal impact of AI. Empirical evidence from algorithmic applications in criminal recidivism prediction and facial analysis, for instance, illustrates how inherent biases in training data or design can lead to discriminatory outcomes, potentially amplifying human biases at an unprecedented scale.

The regulatory response proposed in the White Paper is a targeted, risk-based approach, distinguishing between AI applications based on their potential to cause significant harm. The regulatory response proposed in the White Paper is a targeted, risk-based approach, distinguishing between AI

[341] Luciano Floridi, 'The End of an Era: From Self-Regulation to Hard Law for the Digital Industry' (2021) 34 Philosophy & Technology 619 <https://doi.org/10.1007/s13347-021-00493-0>.

[342] 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust | European Commission' <https://commission.europa.eu/document/d2ec4039-c5be-423a-81ef-b9e44e79825b_en> accessed 19 January 2023.

applications based on their potential to cause significant harm. This largely anticipates what would later be codified in the AI Act. An AI application is deemed "high-risk" if it cumulatively satisfies two criteria: first, its deployment occurs within a sector where significant risks are generally anticipated (e.g., healthcare, transport, energy, and specific areas of the public sector like asylum or judiciary); and second, its specific use within that sector carries a high probability of generating substantial risks, such as producing legal or similarly significant effects for individuals or companies, or causing injury, death, or material/immaterial damage that cannot be reasonably avoided. Certain applications, regardless of sector, are inherently classified as high-risk due to their profound impact on fundamental rights, including AI used in recruitment, for purposes impacting workers' rights, or for remote biometric identification and intrusive surveillance technologies.

For these identified high-risk AI applications, the White Paper envisions a set of mandatory requirements. These include stringent demands concerning training data, stipulating that datasets must be sufficiently broad, representative, and unbiased to prevent discrimination, while also ensuring the protection of privacy and personal data. Record-keeping and data governance obligations are proposed to enhance traceability and accountability, encompassing detailed records of training data, programming methodologies, and validation processes, to facilitate effective oversight and enforcement. Transparency is further emphasised through requirements for information provision, compelling developers to clearly communicate an AI system's capabilities, limitations, and intended purpose, and obliging deployers to inform citizens when they are interacting with an AI system. The principle of robustness and accuracy mandates that AI systems perform reliably, produce reproducible outcomes, effectively handle errors, and exhibit resilience against both overt attacks and subtle manipulations. Crucially, human oversight is required, ranging from human validation of AI outputs to real-time monitoring and intervention capabilities, ensuring that AI systems augment rather than undermine human autonomy. Lastly, given their profound implications for fundamental rights, specific requirements for remote biometric identification are underscored, acknowledging the strict prohibitions and safeguards under EU data protection law and calling for a broad European dialogue on any exceptional circumstances warranting their use. The proposed regulatory framework will apply to all relevant economic operators providing AI-enabled products or services in the EU, irrespective of their establishment location, and will mandate an objective, prior conformity assessment for high-risk AI applications, which may involve testing and certification of algorithms and data. This assessment would be integrated into existing product conformity mechanisms where possible, or new mechanisms would be established. The system would also provide for ongoing ex-post monitoring and ensure effective judicial redress for affected parties. For AI applications not categorised as high-risk, a voluntary labelling scheme is envisaged, enabling economic operators to signal trustworthiness and adherence to specific benchmarks, thereby fostering broader market uptake. The overarching governance structure would facilitate cooperation among national competent authorities, foster capacity building, and provide expertise on standardisation and certification, integrating existing sectoral structures and ensuring broad stakeholder consultation. Ultimately, the objective is to cultivate an environment where AI, as a strategic technology, fully serves European citizens and society, embodying excellence, trust, ethical principles, and sustainability. This comprehensive approach seeks to harness AI's transformative potential while mitigating its inherent risks, thereby establishing Europe as a global benchmark for human-centric AI development and deployment.

2.5 The AI Act Proposal (2021)

The proposal for an Artificial Intelligence Act[343] represents the legislative culmination of the evolving EU strategy, the genesis of which has been traced in the previous sections. The AI Act proposal emerged as the central pillar to actualise the "ecosystem of excellence" and the "ecosystem of trust" by establishing common rules for AI systems. It was designed as part of a coherent triptych of initiatives, complemented by forthcoming rules on liability (later withdrawn[344]) and revisions to sectoral safety legislation, thereby constituting a staged, multi-pronged approach intended to ensure regulatory coherence and legal certainty.

As evidenced in the impact assessment[345], the decision to pursue this binding legal instrument, rather than rely on the soft law mechanisms previously explored, was predicated on a detailed analysis of the distinct challenges posed by AI and the perceived inadequacy of non-binding solutions. Six principal problems were identified as necessitating a legislative intervention. First and foremost among these was the emergence of new risks to safety and security, as the inherent stochasticity and continuous learning capabilities of many AI systems generate hazards not adequately addressed by existing product safety legislation, which has historically focused on static products and risks present at the time of market placement. Concurrently, the proliferation of AI was seen to create an increased risk of fundamental rights violations, either from novel applications such as real-time remote biometric identification or through opaque and complex systems that frustrate the effective enforcement of existing non-discrimination and privacy laws. A third challenge was the inadequate capacity of competent authorities, whose powers and resources are often insufficient to verify compliance or attribute liability in the face of AI's technical complexity and adaptive nature. For businesses, this technological and regulatory landscape created significant legal uncertainty and complexity, particularly across the distributed AI value chain, which in turn could create a chilling effect on innovation and investment. This uncertainty, coupled with public observation of safety incidents or rights infringements, was identified as a source of growing mistrust in AI, a factor considered a significant obstacle to the technology's uptake and, consequently, to the Union's global competitiveness. Finally, in the absence of a common EU framework, the proliferation of divergent national regulations threatened to fragment the digital single market, creating significant obstacles for cross-border operations and jeopardising the Union's digital sovereignty. In light of these challenges, soft law approaches such as voluntary labelling schemes or industry-led codes of conduct were deemed insufficient. Their impact would be limited to self-selecting participants, failing to address systemic risks from unregulated applications, and they would neither provide the requisite legal certainty for businesses nor ensure consistent enforcement. Therefore, a binding legislative instrument was considered indispensable for establishing uniform rules across the Single Market, providing legal certainty, ensuring a high level of protection for health, safety, and fundamental rights, and thereby fostering the trust necessary to position Europe as a global leader in ethical and trustworthy AI.

The regulatory architecture proposed in the AI Act, particularly the preferred option combining mandatory requirements for high-risk applications with voluntary codes of conduct for others, was widely lauded in impact assessments for its anticipated benefits. The primary advantage identified is the effective mitigation of specific, significant risks to safety and fundamental rights through a targeted, proportionate approach. This, in turn, is expected to foster greater legal certainty by establishing harmonised rules and

[343] European Commission (n 48).

[344] 'Proposed EU AI Liability Rules Withdrawn - Bird & Bird' <https://www.twobirds.com/en/insights/2025/proposed-eu-ai-liability-rules-withdrawn> accessed 4 January 2026.

[345] 'Impact Assessment of the Regulation on Artificial Intelligence | Shaping Europe's Digital Future' (n 37).

clear obligations across the AI value chain, thereby reducing complexity for developers and encouraging investment. By codifying requirements for safety, rights protection, and transparency, the Act aims to cultivate the public and commercial trust deemed essential for widespread AI adoption. Such harmonisation is also crucial for preventing the legislative fragmentation that could impede the functioning of the digital single market, particularly for SMEs. The proposal further seeks to bolster governance and enforcement through a robust system of ex-ante conformity assessments, post-market monitoring, and an EU-level cooperation mechanism, granting national authorities enhanced supervisory powers. Critically, the framework includes explicit measures to support innovation, such as regulatory sandboxes and dedicated support from Digital Innovation Hubs, intended to reduce compliance burdens and facilitate market entry for SMEs and start-ups, thus reinforcing the EU's ambition to assume a position of global leadership in setting standards for human-centric and ethical AI.

Despite these perceived strengths, the proposal has drawn considerable criticism from various stakeholders. A persistent concern relates to the potential for significant administrative burdens and compliance costs, which, as the EESC noted[346], could disproportionately affect micro and small organisations. The very definition and scope of "high-risk" AI have been a focal point of debate, with some stakeholders finding the criteria vague while others warned that the list-based approach in Annex III could inadvertently excluded controversial applications from the high-risk categorisation[347]. Critics also pointed to potential gaps, such as the unclear status of legacy AI systems and the exclusion of AI used in democratic processes from the high-risk category. The proposed definition of AI itself was contested as being either overly broad or too narrow. Beyond definitional issues, there were calls for more robust transparency and explainability requirements, alongside a stronger affirmation of human oversight, with the EESC stressing that certain critical decisions in domains like the judiciary and social services must remain an exclusively human prerogative. Concerns also persist regarding the practical challenges of enforcement, including the need for sufficient resources and expertise at national and EU levels. Finally, stakeholders have called for clearer articulation of the Act's relationship with existing legislation to avoid conflicting obligations, a stronger mandate for third-party conformity assessments for all high-risk systems, and the inclusion of a formal complaints and redress mechanism for individuals harmed by AI.

At its core, the regulatory approach of the AI Act is anchored in a human-centric philosophy, positing that AI should serve societal good in alignment with Union values and fundamental rights. To this end, it adopts a technology-neutral and ostensibly future-proof definition of an AI system, encompassing a broad range of techniques listed in an annex, which is designed to be adaptable to future technological advancements. The central regulatory mechanism is a proportionate, risk-based approach that intervenes only where deemed strictly necessary. This framework is structured as a pyramid of risk. At its apex are a limited set of prohibited AI practices, which are considered to contravene Union values so severely that they are outright banned; these include systems that deploy subliminal techniques to distort human behaviour, exploit vulnerabilities to cause harm, or are used for general-purpose social scoring by public authorities.

Immediately below this tier of prohibited practices are the high-risk AI systems, which are subject to the most stringent mandatory requirements. An AI system is classified as high-risk through one of two

[346] 'Regulation on Artificial Intelligence | EESC' (25 March 2021) <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/regulation-artificial-intelligence> accessed 19 August 2025.

[347] Asress Adimi Gikay and others, 'High-Risk Artificial Intelligence Systems under the European Union's Artificial Intelligence Act: Systemic Flaws and Practical Challenges' (Social Science Research Network, 2 November 2023) <https://doi.org/10.2139/ssrn.4621605> accessed 4 January 2026.

pathways: either it functions as a safety component of a product already covered by specific Union harmonisation legislation listed in Annex II, or it falls into one of the categories enumerated in Annex III, which primarily involve applications with significant implications for fundamental rights, such as systems used in critical infrastructure, education, employment, law enforcement, and justice. Providers of these high-risk systems must adhere to a comprehensive set of obligations throughout the system's lifecycle. These include rigorous requirements for data and data governance to ensure quality and mitigate bias; extensive technical documentation and record-keeping to ensure traceability; transparency and the provision of clear information to users; the design of effective human oversight mechanisms; and ensuring the system's robustness, accuracy, and cybersecurity. Compliance is enforced through a mandatory conformity assessment prior to market placement, integrated into existing procedures for products under Annex II and conducted primarily via internal checks for standalone systems under Annex III. This is supplemented by a duty for providers to establish post-market monitoring systems and to register their standalone high-risk systems in a public EU database to enhance transparency and oversight.

For AI systems that fall outside the high-risk and prohibited categories, the Act imposes minimal transparency obligations. These are designed to ensure that individuals are aware when they are interacting with an AI system, such as a chatbot, or when they are exposed to artificially generated or manipulated content like deepfakes, thereby enabling them to make informed choices. For this broad category of non-high-risk AI, providers are encouraged to voluntarily adopt codes of conduct that foster the application of trustworthy AI principles. This tiered regulatory structure is overseen by a governance system comprising designated national competent authorities in each Member State and a European Artificial Intelligence Board at the Union level, which is tasked with facilitating uniform implementation and cooperation. To further support this framework, Member States are encouraged to establish AI regulatory sandboxes—controlled environments for testing innovative systems under regulatory supervision—to foster innovation and enhance legal certainty, particularly for small-scale providers.

Through the legislative process, the AI Act underwent several modifications that led to the final version adopted by the co-legislators in May 2024. However, the overall regulatory structure remained consistent with the initial proposal. The next section will therefore analyse some crucial aspects of the approved version of the Regulation.

3. The AI Act: analysis of key aspects of the Regulation

3.1 Material scope: what is an AI system?

The determination of the material scope of the AI Act hinges fundamentally on the definition of an "AI system". Initial conceptualisations of this pivotal term, emanating from the European Commission's Communication "Artificial Intelligence for Europe" and subsequently elaborated upon by the High-Level Expert Group on AI, conceived of artificial intelligence systems as constructs exhibiting intelligent behaviour through environmental analysis and autonomous action towards specified objectives. This nascent understanding accommodated systems realised purely in software as well as those embedded within physical hardware, and further articulated the reliance upon data, algorithms, and substantial computational resources as constitutive elements. This early delineation portrayed an AI system as a machine-based entity capable of generating predictions, recommendations, or decisions affecting real or

virtual environments, predicated upon a given set of human-defined goals and operating with variable levels of autonomy.

Expanding upon this foundational understanding, the Rule 57 Opinion from the European Parliament's JURI Committee proposed a more granular definition[348]. This refinement specified an artificial intelligence system as a machine-based entity developed employing a variety of techniques and approaches detailed in the Annex I of the proposal (eventually removed in the adopted legislative text), possessing the capacity to exert influence on its environment through outputs such as predictions, recommendations, or decisions, all geared towards predetermined objectives. The proposed definition underscored the utilisation of machine and/or human-based data inputs for environmental perception (real or virtual), abstraction of these perceptions into models through automated or manual analysis (exemplified by machine learning), and subsequent employment of model inference for formulating outcome options. Crucially, this rendition also affirmed the varying degrees of autonomy inherent in AI systems. Annex I, integral to this more specific definition, provided an illustrative enumeration of techniques including machine learning, optimisation approaches, and deep learning, lending operational specificity to the concept. The JURI Committee, reflecting on the breadth of the Commission's original definition when conjoined with Annex I, voiced concerns about its potentially expansive and undifferentiated reach.

The Council's iterative negotiations manifested in compromise texts that further modulated the definition. The 2021 Progress report iteration[349] defined an AI system by its operational characteristics: receipt of data and inputs (machine or human sourced), inferential processes to achieve objectives utilising techniques outlined in Annex I, and generation of outputs encompassing content (particularly relevant to generative AI systems), predictions, recommendations, or decisions with environmental influence. This modification explicitly incorporated the notion of 'inference' as a defining characteristic and formally acknowledged "generative AI systems". Recital 6 underwent parallel amendment, foregrounding the capabilities of inference and output generation. Further conceptual clarity was introduced with the definition of a "general-purpose AI system," delineated as one intentionally designed to perform universally applicable functions such as image and speech recognition, pattern detection, question answering, and translation. Such systems, inherently versatile, could function in diverse contexts and be integrated into other AI systems. Importantly, general-purpose AI systems were excluded from being considered as possessing an "intended purpose" within the purview of the proposed AI Act, unless formally branded or integrated into a separately regulated system, signaling an effort to circumscribe the scope of regulatory oversight for foundational AI technologies.

The culminating definition enshrined in the final AI Act, article 3, describes an "AI system" as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments". This definition is further clarified through the issuance of the Commission Guidelines on the definition of an artificial intelligence system[350]. These guidelines, promulgated in February 2025, are explicitly intended to assist in the practical application of Article 3(1) of the AI Act, which provides the foundational definition, and to facilitate a common understanding

[348] 'Historic Documents | EU Artificial Intelligence Act' <https://artificialintelligenceact.eu/documents/> accessed 19 August 2025.

[349] ibid.

[350] 'Commission Guidelines on AI System Definition' <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> accessed 19 August 2025.

amongst providers, market stakeholders, and institutional actors concerning the material scope of the regulation. While explicitly stated as non-binding, and acknowledging the ultimate interpretive authority of the Court of Justice of the European Union, the Commission Guidelines offer a crucial interpretative lens through which to examine the core constituents of an "AI system" as conceived within the AI Act. The document meticulously unpacks the seven key elements embedded within the Article 3(1) definition, providing nuanced explanations and illustrative examples for each, thus offering a degree of operational clarity that was arguably absent in the legislative text itself.

According to the Commission Guidelines, the definition of an "AI system" pivots around seven fundamental aspects. Firstly, it must be a "machine-based system," a characteristic the Guidelines interpret expansively to encompass both hardware and software components that enable system functionality. This designation, as the Guidelines elaborate, underscores the computational and operational dependence on machines inherent to AI systems, spanning from training processes to deployment. This emphasis on machine-based operation reinforces the intended differentiation from purely human-driven processes, though the precise boundary, particularly regarding systems augmenting or heavily relying on human input, warrants continued scrutiny.

Secondly, an AI system is "designed to operate with varying levels of autonomy." The Guidelines clarify this element as denoting 'some degree of independence of actions from human involvement and of capabilities to operate without human intervention'. This clarification is critical, as it steers the interpretation away from absolute autonomy, recognising the spectrum of human-machine interaction in AI systems. It implicitly excludes systems that are purely manually operated, but encompasses systems with both direct and indirect human oversight, the latter involving automated controls delegated by humans. This interpretation provides a more pragmatic understanding of autonomy within the regulatory context, recognising the engineered nature of autonomy in AI as opposed to philosophical or ontological conceptions.

The third element, adaptiveness after deployment, is presented as a potentially distinguishing, albeit not strictly necessary, feature. The Guidelines emphasise that the term "may exhibit adaptiveness" signifies that while adaptiveness is a characteristic of many AI systems, particularly self-learning systems, its absence does not automatically disqualify a system from being classified as AI. This nuanced approach addresses concerns about excluding systems based on more established or less overtly "adaptive" AI techniques, while still recognising the dynamism inherent in many advanced AI deployments. This reinforces the Act's intended scope to encompass a broad spectrum of AI methodologies, beyond only those explicitly exhibiting continuous learning in operation.

The fourth element pertains to "explicit or implicit objectives," stating that AI systems operate with predefined aims. These objectives, as per the guidelines, can be explicitly encoded by developers or implicitly derived from system behaviour or underlying assumptions, including those embedded within training data. This acknowledgment of implicit objectives broadens the scope of the "human-defined objectives" requirement articulated in the Act's definition itself, acknowledging that objectives might not always be overtly programmed but rather emergent from design choices or data dependencies. This interpretation is particularly salient in understanding the objectives of complex AI systems, where goals might be multifaceted and not always directly programmable.

A pivotal aspect, consistently highlighted in the Act and further emphasised by the Guidelines, is the capacity to "infer". The Guidelines elaborate that "inference" is a key characteristic, distinguishing AI systems from "simpler traditional software systems or programming approaches". This capability to "infer", as the document clarifies, encompasses both the "process of obtaining outputs" and the "capability of AI systems to derive models or algorithms from inputs or data." This duality in the

definition of inference suggests that it refers not only to the operational phase of generating outputs but also crucially to the "building phase", or developmental stage, of AI systems. This distinction is critical for delineating the material scope as it suggests the AI Act intends to regulate not just the applications but also the underlying developmental methodologies if they employ inference in this defined sense. Crucially, the guidelines link "inference" to "AI techniques" further defined as "machine learning approaches that learn from data" and "logic- and knowledge-based approaches that infer from encoded knowledge." Within machine learning, the document further differentiates between supervised, unsupervised, self-supervised, and reinforcement learning, providing examples of AI systems that exemplify each category. This detailed exposition on machine learning techniques potentially offers more concrete criteria for identifying systems that fall within the "AI system" definition, and could be seen as responding, in part, to industry demands for greater clarity on the scope and applicability of the AI Act. Similarly, the explanation of "logic- and knowledge-based approaches," and their distinction from machine learning approaches that learn from data, clarifies the inclusion of more symbolic AI methods within the regulatory ambit, reflecting the technology-neutral stance of the Act.

However, the Guidelines also explicitly delineate systems that fall *outside* the scope of the AI system definition, offering examples such as systems for improving mathematical optimisation, physics-based simulations enhanced with machine learning, basic data processing systems, systems based on classical heuristics, and simple prediction systems achievable via basic statistical learning rules. These examples are crucial for operationalising the boundary of the definition, suggesting that while machine learning or statistical methods might be employed, the *nature* and *complexity* of the inference, and the *purpose* for which it is deployed, are key differentiators for regulatory purposes. This targeted exclusion suggests an attempt to refine the definition's scope, potentially mitigating industry concerns about over-regulation of conventional software.

Regarding the outputs generated by AI systems, the Guidelines reaffirm the four categories enumerated in the Act: "predictions, content, recommendations, or decisions." These output categories are not mutually exclusive, and, as the Guidelines suggest in the context of "content," can be technically re-conceptualised within the spectrum of "predictions" or "decisions." However, the separate listing emphasises the diversity of AI system applications and the range of potential impacts they may have. The Guidelines note that each category differs in its level of human involvement, hinting at a potential gradient of regulatory scrutiny that might align with the type of output and its inherent risk profile.

Finally, the Guidelines reiterate that AI system outputs must "influence physical or virtual environments." This element underscores the active and consequential nature of AI system operations, distinguishing them from purely passive systems. The "environment" is broadly defined to include both tangible, physical entities (like robot arms) and virtual spaces (digital ecosystems, data flows, software environments). This expansive interpretation reinforces the Act's intention to address the real-world impacts of AI systems across diverse domains, both physical and digital.

In conclusion, the Commission Guidelines on the definition of an "AI system" represent a significant interpretive instrument for understanding the material scope of the AI Act. By dissecting Article 3(1) into seven constituent elements, providing detailed explanations, and offering illustrative examples of both included and excluded systems, the guidelines furnish a more operational and nuanced understanding of the definition. While non-binding, their detailed articulation of "inference" and "AI techniques" as well as the delineation of systems falling outside the scope provide critical insights for both regulatory implementation and further scholarly analysis of the AI Act's definitional contours and practical reach. These guidelines serve to clarify certain ambiguities inherent in the legislative text, though the ultimate

adjudication on the scope and interpretation, as the document itself acknowledges, remains within the purview of the Court of Justice of the European Union.

The legal delineation of an "AI system" within the AI Act has proven to be one of the most contentious elements of the legislative drafting process, precisely because this definition determines the regulation's jurisdictional scope and economic reach. This debate illuminates a fundamental tension between the necessity for strict legal precision and the protean, multi-layered nature of the technology itself.

The initial proposal put forth by the European Commission faced immediate and severe criticism for its potential over-inclusiveness, leading many commentators to disparagingly label the draft a "Software Act" rather than an AI Act[351]. The breadth of the definition threatened to impose disproportionate compliance burdens on SMEs for technologies that posed no novel risks, thereby conflating basic automation with autonomous decision-making. Conversely, the legislative bodies faced the equal and opposite danger of under-inclusiveness[352]. A definition restricted solely to specific techniques, such as machine learning, would violate the principle of technological neutrality and inevitably fail the test of future-proofing. A closed list of techniques would create perverse incentives for developers to employ unlisted methods to circumvent regulatory oversight, rendering the law obsolete as new computational architectures emerged.

To resolve this dichotomy, the final text, aligned with the 2023 OECD update, shifted the definitional focus from a list of techniques to a set of key functional characteristics, specifically "autonomy" and "inference." However, this shift introduced a new layer of indeterminacy. Autonomy is not a binary state but a gradual variable; there remains no consensus on the precise threshold of independence from human control required to distinguish an AI system from a standard automated process. Similarly, the "ability to infer", intended to demarcate AI from basic data processing, relies on conceptually elastic terms that may prove difficult to adjudicate in practice[353].

These definitional difficulties are further complicated by structural and contextual loopholes, particularly regarding GPAI and Foundation Models. The legislative framework was originally designed around the concept of "intended purpose," a standard borrowed from product safety legislation which assumes the manufacturer knows the specific application of their product. This approach proved structurally inadequate for Foundation Models, such as LLMs, which are designed to be malleable and applicable across a vast spectrum of contexts, ranging from low-risk creative writing to high-risk medical triaging. Because the provider of a foundation model cannot predict the downstream application, relying on "intended use" created a regulatory vacuum. This forced a significant pivot during the negotiations to establish a tiered regulatory regime that targets the capabilities and systemic risks of the models themselves, rather than solely their specific application by downstream deployers[354].

Furthermore, the exclusions regarding military and research contexts introduce significant ambiguity due to the dual-use nature of the technology[355]. The Act explicitly excludes AI systems developed or used exclusively for military purposes; however, in the realm of software, the distinction between civil and

[351] Nathalie A Smuha and others, 'How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act' (Social Science Research Network, 5 August 2021) <https://doi.org/10.2139/ssrn.3899991> accessed 4 January 2026.

[352] Carlos Trincado Castán, 'The Legal Concept of Artificial Intelligence: The Debate Surrounding the Definition of AI System in the AI Act' [2024] BioLaw Journal - Rivista di BioDiritto 305 <https://doi.org/10.15168/2284-4503-3000>.

[353] ibid.

[354] Hannah Ruschemeier, 'AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal' (2023) 23 ERA Forum 361 <https://doi.org/10.1007/s12027-022-00725-6>.

[355] ibid.

military application is often defined by the user rather than the code. A computer vision algorithm developed for autonomous driving (civil) can be readily repurposed for target recognition in lethal autonomous weapons systems (military), making the "exclusively" qualifier legally porous. Similarly, early drafts failed to provide robust exceptions for scientific research and open-source software. Without clear safe harbours, the heavy compliance obligations threatened to chill the scientific ecosystem, where experimental AI is often shared freely for peer review and iteration. The final text attempts to remedy this by carving out exceptions for AI specifically developed for scientific research and development, yet the boundary between R&D and commercial deployment remains a potential area of friction[356].

Ultimately, the tension between legal certainty and flexibility creates tangible economic consequences. The absence of a universally accepted technical definition makes it difficult for developers to predict whether their systems fall within the scope of the regulation. Given that compliance costs for high-risk AI systems are estimated to as high as four hundred thousand euros depending on the system's complexity and risk category, this uncertainty is not merely academic but financially existential for smaller market participants[357].

3.2 Risk categorisation: rational and limitations

The regulatory architecture of the AI Act is fundamentally predicated on a risk-based paradigm, a strategic choice intended to calibrate regulatory intensity with the potential societal impact of diverse AI applications[358]. This legislative framework distinguishes itself through a stratified system, categorising AI systems into discrete risk strata, each corresponding to a differentiated regime of regulatory oversight and mandatory obligations. The explicit objective underpinning this stratification is to ensure that intervention by regulatory authorities is both proportional and meticulously targeted, concentrating scrutiny and mandates upon those AI deployments adjudged to present the most salient and considerable risks to individual welfare and societal values. The foundational premise posits a commitment to restrict legislative intrusion to the minimum necessary requirements to address the risks and problems linked to AI, thereby actively seeking to pre-empt the unwarranted stifling of innovation within sectors characterised by lower potential for deleterious consequences. By discerning AI systems across a spectrum of risk profiles, the Act endeavours to circumvent the imposition of indiscriminate regulatory burdens upon all manifestations of artificial intelligence, irrespective of their prospective ramifications. This targeted calibration is conceived as pivotal in nurturing a climate conducive to technological advancement while concurrently pre-empting the adverse externalities associated with unfettered AI deployment.

A principal justification for this risk-based approach resides in the augmented efficiency it purportedly confers upon regulatory governance. The delineation of risk categories empowers regulatory bodies to strategically allocate their supervisory and enforcement capacities, prioritising sectors where the prospective for societal or individual harm is most pronounced. This judicious deployment of resources is ostensibly more efficacious than undifferentiated regulatory impositions applicable uniformly across

[356] Michèle Finck, 'In Search of the Lost Research Exemption: Reflections on the AI Act' (2025) 74 GRUR International 903 <https://doi.org/10.1093/grurint/ikaf100>.

[357] Patrick Grady, 'The AI Act Should Be Technology-Neutral' (2023) <https://itif.org/publications/2023/02/01/the-ai-act-should-be-technology-neutral/> accessed 4 January 2026.

[358] Tobias Mahler, 'Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal' (Social Science Research Network, 30 September 2021) <https://papers.ssrn.com/abstract=4001444> accessed 4 January 2026.

the entirety of the AI landscape. Furthermore, the graded nature of the system is designed to instantiate proportionality within the regulatory domain, aligning the stringency of compliance mandates with the calibrated level of risk posed by the specific AI system under consideration. AI systems deemed to embody minimal or negligible risk exposure encounter correspondingly minimal regulatory obligations, potentially extending to the complete absence of direct mandates beyond overarching product safety norms. Conversely, systems categorised as high-risk become subject to a comprehensive compendium of compulsory requirements encompassing risk management, rigorous data governance frameworks, meticulous technical documentation practices, systematic record-keeping procedures, robust transparency mechanisms, meaningful human oversight, and stringent criteria for accuracy, robustness, and cybersecurity. Crucially, the risk-based architecture is conceptualised as possessing inherent adaptability, capable of accommodating future technological trajectories and innovations within the ever-evolving sphere of AI. By prioritising potential societal and individual impact as the central criterion for categorisation, rather than tethering regulation to specific extant technologies, the AI Act seeks to maintain enduring relevance amidst the dynamic advancements characterising the AI domain. Implicit in the adoption of a risk-based approach is the ambition to cultivate and reinforce public and commercial trust in AI technologies. By proactively addressing and mitigating the multifaceted risks associated with AI, particularly those impinging upon fundamental rights and physical safety, the legislative endeavour aims to engender a climate of trust deemed indispensable for the broad societal embrace and assimilation of AI and the realization of its potential societal benefits.

Despite the ostensibly sound rationales underpinning the AI Act's risk categorisation framework, its operationalisation and conceptual underpinnings are not without significant limitations. A primary challenge pertains to the intrinsic complexity in precisely defining and consistently operationalising the demarcation of "high-risk" AI systems[359]. The inherently broad definitions of both "AI" and the constituent risk categories inherent within the legislation have the potential to engender significant legal uncertainty for both developers and deployers of AI technologies. Concerns regarding overbreadth are further exacerbated by observations that the expansive definition could ensnare numerous systems not traditionally understood or assessed as embodying the archetypal risks normally associated with artificial intelligence in public discourse.

Further complexities arise from the structure and composition of Annex III of the AI Act, which delineates the sectors and use-cases categorised as inherently high-risk[360]. Critical analysis suggests that this enumerated list is susceptible to being simultaneously over-inclusive and under-inclusive in its practical application. Apprehensions have been voiced that the expansive and potentially ambiguous wording employed in Annex III could inadvertently capture applications of AI that are either genuinely low-risk in nature or constitute merely ancillary uses tangential to the core objectives of the enumerated high-risk sectors. Conversely, the static nature of a list-based approach raises concerns about its inherent inability to proactively account for emergent high-risk AI applications that might fall outside the predefined categories. Innovators, perceiving regulatory disincentives within listed sectors, might strategically tailor the presentation and framing of their innovations to fall demonstrably outside the ambit of Annex III, thereby potentially circumventing stringent regulatory oversight, leading to instances of regulatory under-enforcement. The inherently dynamic and rapidly evolving nature of AI technologies

[359] Johann Laux, Sandra Wachter and Brent Mittelstadt, 'Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18 Regulation & Governance 3 <https://doi.org/10.1111/rego.12512>.

[360] Tiago Sérgio Cabral, 'Rethinking the List-Based Approach to High-Risk Systems under the AI Act' (Social Science Research Network, 3 April 2025) <https://doi.org/10.2139/ssrn.5206860> accessed 19 August 2025.

renders static lists particularly vulnerable to obsolescence and susceptible to regulatory arbitrage. The inclusion of specific sectors and use cases within Annex III, moreover, involves an inescapable element of subjective interpretation and potentially reflects underlying political and normative value judgements embedded in the regulatory process. The designation of what constitutes a "significant risk" and the identification of particular sectors or applications as inherently "high-risk" are not purely technical exercises susceptible to purely objective quantification[361]. Rather, these determinations inherently involve subjective evaluation and potentially incorporate politically mediated categorisations of high-risk AI systems[362]. The inherent absence of an objectively quantifiable and universally accepted methodology for evaluating and measuring risks, particularly those related to amorphous and conceptually contested domains like fundamental rights, further compounds the complexity and potential contestation surrounding the process of risk classification.

The presence within the AI Act of exemptions, particularly Article 6 pertaining to AI systems intended for "improvement" or "preparatory" tasks preceding high-risk uses, also constitutes a locus of critical scrutiny[363]. Concerns have been raised regarding the potential for these exemptions to create unforeseen regulatory loopholes, introducing ambiguity in practical implementation and hindering robust enforcement. The legislative framework stipulates that providers of systems falling under these exemptions bear the responsibility for self-assessing and documenting their classification as non-high-risk, subsequently registering this self-assessment within the EU-wide database. This self-assessment mechanism inherently carries the risk of inconsistent application and raises questions about the efficacy of oversight and the potential for biased or self-serving interpretations by providers. Moreover, the prevailing paradigm of risk assessment embedded within the AI Act, especially in the context of conformity assessment procedures detailed in Article 9, has been critiqued for potentially overemphasising the AI product in isolation, thereby neglecting the critical contextual embedding of AI systems within broader socio-technical systems and operational environments. The explicit limitation of risk mitigation obligations to risks "reasonably mitigated or eliminated through the development or design of the high-risk AI system" is argued by some commentators to represent a conceptually constricted approach, potentially overlooking or downplaying the broader societal externalities and systemic impacts arising from AI deployments, particularly those residing outside the immediately controllable technical parameters of the system's design[364]. Furthermore, the imperative of ensuring coherence and compatibility with existing EU legal instruments, including data protection regulations like the GDPR and diverse sector-specific directives, presents a formidable and ongoing challenge. Navigating the complex interrelationship between the AI Act and extant legal regimes necessitates meticulous fine-tuning and continuous interpretation to ensure regulatory clarity and avoid inadvertent legal conflicts or overlaps.

The subsequent introduction of specific obligations applicable to providers of GPAI models, particularly those deemed to present "systemic risk," introduces a further layer of definitional and operational complexity[365]. The definition of "systemic risk," while explicitly referencing high-impact

[361] 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 <https://doi.org/10.1017/err.2023.57>.

[362] Tartaro, 'Value-Laden Challenges for Technical Standards Supporting Regulation in the Field of AI' (n 55).

[363] 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' (2024) 26 Yale Journal of Law and Technology <https://ora.ox.ac.uk/objects/uuid:0525099f-88c6-4690-abfa-741a8c057e00> accessed 4 September 2024.

[364] Laux, Wachter and Mittelstadt (n 359).

[365] Stefan Larsson and Jockum Hildén, 'Flip or Flop? : The Pacing Problems of European Systemic Risk GPAI Regulation' <http://lup.lub.lu.se/record/98428617-b972-4e15-90e7-cc3de05f83dc> accessed 19 August 2025.

capabilities and potentially negative externalities on public health, safety, fundamental rights, or society at large, and the propensity for wide-scale propagation across value chains, requires more granular elucidation to enable effective operationalisation. The precise criteria and quantitative or qualitative thresholds for triggering the classification of a GPAI model as posing "systemic risk" remain subject to ongoing debate and require more concrete articulation to avoid interpretive ambiguity.

3.3 Essential requirements for high-risk AI systems

AI systems that fall into the high-risk category according to the classification rules in Article 6 are subject to stringent requirements articulated within Articles 8 through 15.

Foremost among these stipulations is the requirement for a comprehensive risk management system[366]. This system, articulated under Article 9, is not envisioned as a static, pro forma compliance exercise, but rather as a dynamic and iterative process embedded throughout the entire lifecycle of the high-risk AI system. It necessitates a systematic, documented, and perpetually maintained framework for the identification, analysis, evaluation, and mitigation of potential risks. The scope of this system extends beyond merely intended usage scenarios, crucially encompassing reasonably foreseeable misuse and integrating post-market monitoring data to adapt and refine risk mitigation strategies over time. This lifecycle-centric approach recognises that risks are not confined to the point of deployment but rather are mutable and may evolve as the system interacts with its operational environment. However, the inherently complex nature of risk assessment in AI, particularly in scenarios involving intricate emergent behaviours, underscores the challenge of establishing universally applicable metrics for what constitutes acceptable residual risk and highlights the potential for subjective interpretation within the regulatory framework, as we shall explore in chapter 5.

Subsequent to the establishment of a robust risk management framework, the EU AI Act places considerable emphasis, in Article 10, on the governance and quality of data used in the development process, particularly for systems employing machine learning paradigms. This provision mandates rigorous data governance and management practices encompassing the entire data pipeline, from initial design choices and collection methodologies to preparation protocols, formulation of underlying assumptions, and a proactive assessment of both data availability and inherent suitability for the intended application. A critical facet of this data governance framework is the imperative to rigorously scrutinize datasets for potential biases and to proactively identify and address data gaps which may compromise the system's performance or fairness. The Act further mandates transparency concerning the provenance, scope, and intrinsic characteristics of the datasets utilized, including detailed documentation of acquisition and selection methodologies, labelling procedures, and data cleaning processes, thereby fostering a culture of accountability and auditability around data-driven AI. While the aspiration to employ datasets that are 'free of errors' and 'complete' is laudable, scholars have pragmatically observed the inherent difficulties, if not outright impossibility, of attaining such absolutist standards in real-world applications[367]. The suggestion that a more pragmatic phrasing focusing on the statistical negligibility of errors might be more realistically achievable underscores a necessary tension between regulatory ambition and practical feasibility. Moreover, the critical recognition of context-specificity embedded in Article 10, mandating consideration of the intended geographical, behavioural, or functional deployment context,

[366] Jonas Schuett, 'Risk Management in the Artificial Intelligence Act' [2023] European Journal of Risk Regulation 1 <https://doi.org/10.1017/err.2023.1>.

[367] 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>.

reflects a nuanced understanding that data quality cannot be divorced from the specificities of the operational environment[368]. The permissibility, under strictly necessary circumstances and with robust safeguards, of processing sensitive data reflects an attempt to navigate the complex intersection between data protection regulations and the imperative to address discriminatory biases that might be perpetuated by incomplete or insufficiently representative datasets, a particularly acute challenge in domains such as biometric identification and algorithmic fairness[369].

To ensure verifiable compliance with these multifaceted requirements, the Act mandates the meticulous creation and diligent maintenance of technical documentation, as stipulated in Article 11. This documentation serves as a comprehensive evidentiary repository, demonstrably illustrating adherence to the substantive requirements outlined in Chapter 2 of Title III and providing national competent authorities and notified bodies with the requisite information for rigorous assessment. The mandated content, delineated in Annex IV, is exhaustive, encompassing not only a general description of the AI system, including its intended purpose and key operational characteristics, but also delving into the granular details of its developmental trajectory. This includes explicit elucidation of design specifications, critical design choices and their rationale, system architecture, computational resource utilisation, and a meticulous exposition of data requirements, extending to the very provenance, scope, and characteristics of datasets employed in training, validation, and testing phases. Crucially, the technical documentation must transparently articulate any trade-offs made during the design and development process, acknowledging the inherent compromises and optimisations that invariably occur in complex technological development. Moreover, detailed expositions concerning the monitoring mechanisms, operational dynamics, control functionalities, performance capabilities and limitations, and foreseeable unintended outcomes are required, alongside comprehensive articulation of human oversight mechanisms and specifications for input data. This rigorous documentation is not merely a procedural formality, but rather constitutes a critical instrument for fostering transparency, facilitating interpretability, and underpinning effective post-market monitoring activities. However, it also raises concerns regarding the potential burden and associated intellectual property sensitivities implicated by the requisite level of detail, particularly with respect to the 'general logic' of the AI system, which can be perceived as a vaguely defined and potentially commercially sensitive requirement. The inherent tension between the imperative for regulatory transparency and the justifiable need to protect proprietary technological innovation remains a salient point of ongoing discussion.

Supplementing the technical documentation, Article 12 mandates the intrinsic capacity for automatic record-keeping within high-risk AI systems. This requirement emphasises the fundamental importance of automated logging of pertinent events. Such logs are instrumental in facilitating traceability, enabling rigorous post-market surveillance, and empowering comprehensive incident analysis, thereby fostering a regime of accountability and auditability. The precise specifications of logging protocols and data retention periods necessitate further elaboration and standardisation, yet the underlying principle of systematically capturing operationally relevant events underscores the regulatory intent to ensure demonstrability of compliance and to provide mechanisms for retrospective investigation and rectification should untoward incidents occur. This emphasis on automaticity further suggests a design philosophy that embeds compliance considerations directly into the system architecture itself rather than relying solely on external audits or ex-post facto evaluations.

---

[368] Håkan Burden and Susanne Stenberg, *Fit for Purpose – Data Quality for Artificial Intelligence* (RISE Research Institutes of Sweden 2024) <https://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-74929> accessed 19 August 2025.

[369] Marvin van Bekkum, 'Using Sensitive Data to De-Bias AI Systems: Article 10(5) of the EU AI Act' (2025) 56 Computer Law & Security Review 106115 <https://doi.org/10.1016/j.clsr.2025.106115>.

Building upon the pillars of documentation and record-keeping, Article 13 of the AI Act mandates proactive transparency and the provision of comprehensive information to deployers. This requirement is not limited to mere procedural adherence but is explicitly focused on ensuring deployers possess the requisite knowledge and understanding to effectively navigate their obligations and responsibilities under the Act. Instructions for use, articulated in appropriate digital or accessible formats, are prescribed as the primary instrument for conveying concise, complete, accurate, and clear information directly relevant, accessible, and readily comprehensible to deployers. These instructions are not envisioned as generic manuals but are rather tailored to furnish deployers with context-specific understanding. They are required to detail not only the intended purpose of the system, but also to explicitly articulate the system's performance envelope, encompassing its capabilities and limitations, including levels of accuracy, robustness, and cybersecurity validation metrics alongside foreseeable contextual factors influencing these performance parameters. Furthermore, instructions must proactively address known or foreseeable misuse scenarios and potential consequential risks to health, safety, or fundamental rights stemming from such misuse. Where applicable, the performance behaviour of the system in relation to specific persons or demographic groups, specifications for input data, and anticipated system modifications must be disclosed. Crucially, information on human oversight mechanisms and the system's expected lifecycle, maintenance protocols, and stipulations regarding intended and precluded uses must be unequivocally conveyed. This comprehensive information provision is not merely about information disclosure but is rather strategically designed to empower deployers to engage in informed decision-making regarding system deployment, operation, and management, thereby fostering a responsible and knowledgeable ecosystem of AI deployment[370]. It is notable that this transparency mandate for high-risk systems is distinct in both scope and purpose from the transparency obligations imposed upon certain other AI systems, such as generative AI models, under Article 50, the latter being primarily focused on user awareness of interaction with an AI entity, whereas Article 13 emphasises deployer-centric, operational understanding and responsibility.

Central to mitigating the inherent risks associated with high-risk AI systems is the mandate for human oversight, meticulously articulated in Article 14[371]. This provision stipulates that high-risk systems must be designed and developed to enable effective oversight by natural persons throughout their operational lifecycle. The objective is not merely nominal human presence but rather demonstrably effective human involvement in preventing or mitigating risks to health, safety, and fundamental rights that may eventuate from autonomous or semi-autonomous AI system operations. Effective human oversight necessitates that designated individuals possess a thorough understanding of the AI system's capacities and inherent limitations. Moreover, oversight mechanisms must actively counter the well-documented phenomenon of "automation bias", the proclivity for humans to uncritically accept or over-rely upon automated system outputs. Crucially, human oversight must equip individuals to accurately interpret AI system outputs and, where circumstances warrant, to override or reverse the system's automated conclusions. Furthermore, the provision for human intervention extends to enabling decisive intervention, potentially including a "stop button" mechanism, to interrupt or halt system operation in critical situations. However, in complex autonomous systems where real-time decision-making is paramount, concerns exist that overly prescriptive or excessive human intervention may, paradoxically, undermine system accuracy and

[370] Kasia Söderlund, 'High-Risk AI Transparency? On Qualified Transparency Mandates for Oversight Bodies under the EU AI Act' (2025) 2025 Technology and Regulation 97 <https://doi.org/10.71265/6bedar76>.

[371] Lena Enqvist, '"Human Oversight" in the EU Artificial Intelligence Act: What, When and by Whom?' [2023] Law, Innovation and Technology <https://www.tandfonline.com/doi/abs/10.1080/17579961.2023.2245683> accessed 21 August 2025.

efficacy[372]. This raises critical questions about the optimal calibration between human control and system autonomy in different operational contexts, highlighting the need for context-sensitive approaches to human oversight design. The conceptual distinction between oversight as review and evaluation versus active intervention as a distinct modality of human engagement further underscores the nuanced spectrum of human roles in governing AI systems.

Finally, Article 15 of the EU AI Act addresses the critically important technical characteristics of accuracy, robustness, and cybersecurity for high-risk AI systems, mandating the achievement of appropriate levels across these dimensions throughout the system's lifecycle. This trio of requirements forms the bedrock of technical reliability and resilience. Accuracy, pertaining to the fidelity and correctness of system outputs, requires that systems attain a level of precision commensurate with their intended purpose and associated risks. Robustness, in turn, necessitates system resilience in the face of errors, faults, inconsistencies, and unanticipated or anomalous operational situations. This includes resilience against malicious attempts by unauthorized actors to subvert or manipulate system performance, potentially necessitating 'fail-safe' mechanisms and proactive safeguards. For systems that inherently incorporate continuous learning after deployment, the robustness requirement is particularly salient, necessitating mechanisms to maintain integrity and prevent detrimental drift or degradation of performance over time. Cybersecurity, as the third pillar, mandates system resilience against cyberattacks and exploitation of vulnerabilities across the entire lifecycle. This encompasses a comprehensive spectrum of cybersecurity threats, including data poisoning, model evasion, and model inversion attacks, requiring proactive security measures and potentially involving techniques such as red-teaming and transparent communication to users regarding security coverage and update protocols. The term "appropriate level", consistently employed throughout Article 15, deliberately underscores the context-dependency of these technical requirements. The acceptable thresholds for accuracy and robustness are not absolute metrics but must be determined relative to the intended purpose and the constellation of potential risks associated with specific high-risk applications. As we shall see in the next Chapter, this presents significant challenges for AI standardisation.

The articulation of these essential requirements collectively constitutes a robust framework designed to foster responsible innovation and deployment of high-risk AI systems. Nevertheless, their implementation is not without challenges for regulated entities. As will be discussed in Chapter 5, the requirements for high-risk AI systems introduce value-laden challenges[373]. Technical standards are intended to assist regulated entities in navigating these challenges, but their effectiveness is not guaranteed.

4. The role of harmonised standards in the AI Act

4.1 Three main functions of harmonised standards

Within the regulatory architecture of the AI Act, technical standards, and in particular harmonised standards referenced in the OJEU, assume a pivotal role in operationalising the Regulation's requirements and ensuring their effective implementation across the European Union. These standards function as the

[372] Andreas Holzinger, Kurt Zatloukal and Heimo Müller, 'Is Human Oversight to AI Systems Still Possible?' (2025) 85 New Biotechnology 59 <https://doi.org/10.1016/j.nbt.2024.12.003>.

[373] Tartaro, 'Value-Laden Challenges for Technical Standards Supporting Regulation in the Field of AI' (n 55).

principal mediators between the abstract, often technologically neutral, legal obligations set forth in the AI Act and the concrete, rapidly evolving realities of AI system development and deployment. Their significance is multifaceted, encompassing at least three aspects: the provision of technical solutions for compliance, the facilitation of a presumption of conformity, and the underpinning of both internal and external conformity assessment procedures[374].

The provision of technical solutions for compliance constitutes perhaps the most immediate and practical function of harmonised standards and common specifications under the AI Act. The Regulation itself is intentionally drafted to be technology-neutral and future-proof, articulating requirements in terms of outcomes, rather than prescribing specific technical means. In this context, harmonised standards, developed within the ESOs, serve to translate these high-level legal requirements into actionable, state-of-the-art technical specifications. For providers of AI systems, adherence to such standards offers a clear, authoritative pathway to compliance, reducing interpretive ambiguity and the risk of divergent implementation across Member States. For example, a harmonised standard might specify the minimum criteria for dataset quality and representativeness to mitigate bias, or delineate protocols for logging and traceability to support transparency and accountability. In doing so, these standards not only facilitate compliance with the Regulation but also promote the diffusion of best practices and technological excellence throughout the European AI ecosystem. The dynamic nature of standardisation processes further ensures that these technical solutions remain responsive to advances in AI research and development, thereby maintaining alignment with the evolving state of the art.

Equally significant is the function of harmonised standards in enabling a presumption of conformity with the AI Act's requirements. Under Article 40 of the AI Act, providers who demonstrate that their AI systems comply with relevant harmonised standards are afforded a legal presumption that their systems meet the corresponding regulatory obligations, particularly those relating to health, safety, and the protection of fundamental rights. This presumption operates as a powerful compliance facilitation mechanism, substantially simplifying the evidentiary burden on providers[375]. Rather than being required to independently substantiate conformity with each requirement, an often complex and resource-intensive process, especially for high-risk AI systems, providers can rely on adherence to established standards as prima facie evidence of compliance. In the absence of harmonised standards, the Commission is empowered to adopt common specifications, which serve an analogous function and ensure that the presumption of conformity remains available even in areas where standardisation has not yet kept pace with technological innovation. This mechanism not only streamlines the compliance process for providers but also enhances legal certainty and predictability, both of which are essential for fostering innovation and investment in the European AI sector.

The role of technical standards extends further to the support of conformity assessment procedures, which are central to the Regulation's risk-based approach, particularly for high-risk AI systems. Conformity assessment under the AI Act encompasses both internal and external modalities. Internally, providers are required to undertake rigorous self-assessment processes to verify that their systems meet all applicable requirements prior to market placement or commissioning. Harmonised standards provide the methodological foundation for these internal assessments, offering detailed criteria and procedures for testing, documentation, and risk management. By following these standards, providers can structure their internal compliance activities in a systematic and comprehensive manner, thereby reducing the likelihood of oversight or non-compliance. Externally, where third-party conformity assessment is mandated, such as for certain categories of high-risk AI systems, harmonised standards serve as the

[374] Tartaro, 'Regulating by Standards' (n 20).
[375] Portalier (n 284).

benchmark against which notified bodies evaluate the conformity of AI systems. This ensures consistency, objectivity, and transparency in the assessment process, while also facilitating mutual recognition of conformity assessments across Member States. The integration of technical standards into both internal and external conformity assessment procedures thus underpins the credibility and effectiveness of the AI Act's regulatory regime, supporting its overarching objectives of safeguarding fundamental rights, ensuring safety, and promoting trustworthy AI.

The strategic decision to rely so heavily upon standardisation is a direct application of the New Approach regulatory model and the NLF, explored in Chapter 3. Within the context of the AI Act, this approach is intended to provide the legal certainty necessary to foster investment and innovation, while simultaneously preventing the fragmentation of the internal market that would arise from divergent national rules. The ultimate goal is the establishment of a single market for lawful, safe, and trustworthy AI. By making conformance with harmonised standards the most straightforward and cost-effective route to compliance, the legislation incentivises their widespread adoption, thereby promoting a high level of protection as a market norm.

Despite its centrality to the EU AI Act, the reliance on technical standardisation as a primary regulatory mechanism is fraught with significant and multifaceted challenges[376]. These problems call into question the suitability of the chosen regulatory model, raise profound concerns about democratic legitimacy, and present considerable hurdles to practical implementation. A principal critique centres on the arguably unsuitable transposition of the "New Approach" framework to the domain of artificial intelligence[377]. This regulatory model, conceived in the 1980s for ensuring the health and safety of tangible goods in the physical marketplace, is ill-suited to the complexities of the online, digital environment, particularly for a technology as protean and impactful as AI. The transposition overlooks the fundamental differences between conventional products and AI systems, whose use cases are often unpredictable and whose applications are deeply embedded within complex socio-technical systems. This makes it exceedingly difficult for developers to foresee, let alone mitigate, all potential impacts on fundamental rights. The core logic of the New Approach, which prioritises market unity and speed of innovation, sits uneasily with the non-negotiable and indefeasible character of fundamental rights, such as privacy, non-discrimination, human dignity, and the right to an effective remedy, which require careful and deliberate balancing rather than expedient technical fixes.

This ill-fitting transposition is compounded by a severe democratic deficit stemming from the delegation of de facto rulemaking authority to private standardisation bodies[378]. The AI Act effectively outsources the formulation of detailed rules for high-risk AI systems to the ESOs, which are private, often bureaucratic entities dominated by industry actors. This delegation of what amounts to legislative power to bodies lacking direct democratic accountability is deeply problematic. ESOs are frequently criticised for being ill-equipped to adjudicate on sensitive ethical and fundamental rights matters, as their technical experts do not necessarily possess the requisite jurisprudential or ethical expertise. The process is further skewed by the structural exclusion of non-technical stakeholders. Civil society organisations, consumer groups, and even SMEs are often de facto marginalised due to a lack of resources, time, and the specific technical knowledge required to participate meaningfully, allowing large commercial interests to exert disproportionate influence. Social stakeholders may be granted observer status but are typically

[376] Sybe de Vries, Olia Kanevskaia and Rik de Jager, 'Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation' (2023) 2023 8 European Papers - A Journal on Law and Integration 583610 <https://doi.org/10.15166/2499-8249/677>.

[377] ibid.

[378] Christine Galvagna, 'Discussion Paper: Inclusive AI Governance' <https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/>.

denied voting rights, further cementing industry control and transforming ESOs into de facto AI regulators whose priorities may not align with broader European values.

The regulatory framework is also undermined by significant conceptual and definitional ambiguities that pervade the legal text. Foundational concepts such as "human-centric," "trustworthy," and "ethical AI" lack clear legal contours and often devolve into nebulous references to fundamental rights without providing concrete guidance, despite scholarly efforts to formalize some of these concepts such as trustworthiness[379]. Furthermore, the frequent use of vague and unspecified language—such as requiring "appropriate" statistical properties, an "examination in view of possible biases" without defining bias, or systems to be "sufficiently transparent"—creates a wide margin of interpretation[380]. This ambiguity threatens to dilute the regulatory effect of the Act and could lead to inconsistent obligations for providers. This issue is magnified by the nature of emerging standards from international bodies like ISO/IEC, which are often "socio-technical" rather than purely technical. These standards attempt to define legal concepts like transparency and accountability but are often too abstract to be certifiable under the New Approach, blurring the line between technical specification and normative lawmaking[381].

Beyond these structural and conceptual failings, the practical implementation and adaptability of the standardisation process face immense challenges. The development of requisite standards has proven to be slower than anticipated, hampered by a mismatch between the ambitious scope of work and the limited resources, in terms of both expert availability and funding, allocated to the task[382]. This deliberate pace of standardisation is fundamentally at odds with the rapid velocity of AI development, creating a persistent regulatory lag.

The approach to standardisation raises significant geopolitical and international coordination issues as well. The conspicuous exclusion of the ETSI from the initial standardisation request, reportedly due to concerns over the influence of non-European members, signals a protectionist turn that is incongruous with the global nature of AI development and the need for international cooperation[383]. This risks fragmenting the standardisation landscape, creating siloed standards that fail to account for the interconnectedness of AI and telecommunications, and ultimately diminishing the EU's capacity to shape global norms. A further misalignment exists with ongoing international efforts, as bodies like ISO/IEC and IEEE have historically developed standards that prioritise the objectives of the organisations deploying AI rather than the protection of individuals, clashing with the AI Act's rights-based orientation[384].

Finally, as we will discuss in Chapter 5, the manner in which the AI Act's essential requirements are framed hinders their transposition into technical specifications usable by AI providers for compliance purposes. As we shall see, this is due to ethical and epistemological challenges that impose unavoidable limits on the standardisation of AI.

---

379 Piercosma Bisconti and others, 'A Formal Account of Trustworthiness: Connecting Intrinsic and Perceived Trustworthiness' *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2024) <https://ojs.aaai.org/index.php/AIES/article/view/31624> accessed 20 April 2026.

380 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>.

381 Michael Veale, 'Value-Laden Areas for Standardisation in the AI Act' (*michael veale*, 5 September 2022) <https://michae.lv/value-laden-areas-in-the-ai-act/> accessed 25 April 2024.

382 Tartaro and Panai (n 117).

383 Marta Cantero Gamito, 'The Role of ETSI in the EU's Regulation and Governance of Artificial Intelligence' (2024) 0 Innovation: The European Journal of Social Science Research 1 <https://doi.org/10.1080/13511610.2024.2349627>.

384 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.

4.2 The standardisation request

The European Commission's issuance of a standardisation request to CEN and CENELEC on 22 May 2023 represents a decisive step in the Union's strategy to establish a robust technical infrastructure for the governance of artificial intelligence[385]. This initiative is designed to advance technical harmonisation in the field of AI and to create the necessary preconditions for the effective implementation of the AI Act. The Commission's approach underscores the centrality of standards as instruments for translating legislative and policy objectives into operational requirements, thereby ensuring not only a high level of protection for safety and fundamental rights but also the creation of a level playing field for AI system development across the Union, with particular attention to the needs of small and medium-sized enterprises.

The standardisation request delineates a comprehensive programme of work for CEN and CENELEC for the drafting of new European standards or standardisation deliverables in ten key technical domains. These domains correspond closely to the requirements articulated in the Artificial Intelligence Act proposal and encompass risk management systems, data governance, record keeping, transparency, human oversight, accuracy, robustness, cybersecurity, quality management systems, and conformity assessment. The selection of these areas reflects a holistic understanding of the technical and organisational dimensions of trustworthy AI, as well as the multifaceted risks that AI systems may pose to individuals and society.

In elaborating the general requirements for these standards, the Commission has mandated that CEN and CENELEC ensure alignment with the generally acknowledged state of the art, thereby minimising risks to health, safety, and fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union and relevant EU law. This requirement is not merely aspirational but is intended to ensure that the resulting standards are both technologically current and normatively robust, capable of addressing the evolving threat landscape and the dynamic nature of AI technologies. Furthermore, the standards must be consistent with the Union's legal framework and international obligations, safeguarding EU values and fundamental rights while also facilitating interoperability with global standards regimes.

A further layer of complexity is introduced by the requirement that these standards provide technology-based, process-based, or methodology-based technical specifications for AI system design and development, including verification, validation, and testing procedures. These specifications must be objectively verifiable and implementable, thereby enabling both providers and regulators to assess compliance in a transparent and reproducible manner. The use of a common set of terms, ideally building on international terminology, is intended to promote clarity and facilitate cross-border cooperation, while the consideration of both horizontal risks common to all AI systems and vertical, sector-specific risks ensures that the standards are sufficiently granular to address the diverse applications of AI across the economy.

The standardisation request also requires CEN and CENELEC to take into account the work of the ETSI, particularly in relation to security aspects, vertical specifications, and testing and validation procedures. This integrative approach is designed to avoid duplication of effort, ensure consistency with existing and future European and sectoral standards, and leverage the expertise of established standardisation bodies. The specificities and cost structures of SMEs are to be given due consideration,

[385] Commission implementing decision 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 - C(2023)3215.

particularly in relation to quality management systems and conformity assessments, so as to avoid imposing disproportionate burdens that could stifle innovation or exclude smaller actors from the market.

At the level of individual technical domains, the standardisation request sets out detailed requirements. For risk management systems, the standards must specify a continuous, iterative process that spans the entire lifecycle of the AI system, with the aim of preventing or minimising risks to health, safety, or fundamental rights. This process should be integrable with existing risk management systems for AI systems that function as safety components or products, and should be usable by both operators and market surveillance authorities. In the domain of data and data governance, the standards are to address the full spectrum of data management activities, from generation and collection to preparation and bias detection, as well as the quality attributes of datasets used for training, validation, and testing. Record keeping standards are to specify requirements for automatic logging of events, thereby ensuring traceability and facilitating post-market monitoring.

Transparency and information provision standards are to focus on design solutions that ensure operational transparency and provide clear instructions for use, tailored to the needs of different user profiles. Human oversight standards must specify measures and procedures that enable users to understand, monitor, interpret, assess, and influence the operation of AI systems, with particular safeguards for remote biometric identification systems, where separate verification and confirmation by at least two natural persons is required before any action or decision is taken. Accuracy and robustness standards are to define the relevant metrics and tools for assessing the performance and resilience of AI systems, taking into account the intended purpose and potential sources of error or environmental interaction. Cybersecurity standards are to provide both organisational and technical solutions to ensure resilience against malicious attempts to alter system behaviour or compromise security, including measures to prevent and control attacks such as training data poisoning or adversarial examples, and to align with essential requirements for products with digital elements.

Quality management system standards are to specify the requirements for continuous compliance with the full range of regulatory obligations, with particular attention to the integration of these systems with existing quality management frameworks and the specific needs of SMEs. Conformity assessment standards are to provide detailed procedures for both provider-led and third-party assessments, including criteria for assessing the competence of those performing the assessments and covering the full range of scenarios envisaged by the Regulation.

The governance of the standardisation process itself is subject to rigorous oversight and reporting requirements. CEN and CENELEC are required to prepare a detailed work programme, indicating all requested standards, responsible technical bodies, and a timetable, and to submit this to the Commission within four months of notification. The work programme must include measures to ensure appropriate stakeholder representation, including SMEs and societal stakeholders, and to guarantee conformity with Union law on fundamental rights and data protection. Regular reporting every six months is mandated. Any major concerns, requests for additional time, or scope adjustments must be promptly communicated to the Commission, which retains access to the overall project plan.

This comprehensive standardisation effort is thus positioned as a cornerstone of the Union's strategy to ensure that AI systems placed on the market or put into service are not only technologically advanced but also safe, rights-respecting, and consistent with Union values. By embedding public interest considerations at every stage of the process, the Commission seeks to foster an environment conducive to investment, innovation, competitiveness, and growth, while maintaining a vigilant focus on the protection of fundamental rights and the broader societal implications of AI deployment. The success of

this initiative will depend not only on the technical quality and relevance of the resulting standards but also on the inclusiveness, transparency, and adaptability of the standardisation process itself, particularly as the covered areas may be subject to future standardisation requests under the evolving framework of the AI Act.

5. The role of conformity assessment and notified bodies in the AI Act

Since the AI Act relies on the tripartite standard regime[386], consisting of standards, certification, and accreditation discussed in Chapter 3, Notified Bodies are essential to its enforcement. These government-recognised entities act as independent third-party auditors, responsible for verifying that specific categories of high-risk AI systems meet the necessary conformity requirements before entering the market[387]. As already mentioned in Chapter 3, the designation of these entities is contingent upon their fulfilment of a stringent set of prerequisites designed to ensure their competence, independence, and integrity. The AI Act specifies requirements related to notified bodies in Chapter III, Section 4[388].

To qualify for designation, a body must possess legal personality and satisfy extensive organisational, quality management, resource, and procedural requirements, including the implementation of suitable cybersecurity measures. Foremost among these are the principle of independence and impartiality, a cornerstone of the entire conformity assessment regime. NBs are required to maintain complete structural and financial independence from the providers of the AI systems they assess, a mandate intended to prevent any conflicts of interest that could compromise the objectivity of their evaluations. However, this principle confronts a structural tension, as the financial relationship inherent in a fee-for-service model, where providers pay for the assessment, creates the potential for regulatory capture and raises persistent concerns about the absolute neutrality of these bodies[389]. This fundamental requirement for independence is buttressed by the demand for demonstrable technical competence and adequate resources. NBs must possess and maintain deep, multidisciplinary expertise across a formidable range of fields, including the specific AI technologies under review, data science, cybersecurity, and, crucially, the legal frameworks governing data protection and fundamental rights. The acquisition and retention of personnel with such a sophisticated and rare combination of skills represent a significant operational and financial impediment for these organisations.

These core competencies must be embedded within a robust organisational framework. A formal quality management system, supported by meticulously documented procedures, is mandated to ensure that all assessment activities are performed with consistency, rigour, and complete independence. This procedural scaffolding is complemented by strict obligations of confidentiality, which protect the sensitive intellectual property and trade secrets to which NBs gain access during their assessments. To underwrite their professional risk, NBs must also secure appropriate liability insurance, unless liability is assumed directly by the Member State. In their operational conduct, NBs are further guided by considerations of proportionality; they are instructed to avoid imposing unnecessary administrative or financial burdens on providers, with a particular sensitivity to the resource constraints of SMEs. The certifications they issue are not indefinite but are granted for a limited period, typically up to five years,

386 Allison Loconto, John Stone and Lawrence Busch, *Tripartite Standards Regime* (2012).

387 Galland, 'Standards, Certification, and Accreditation' (n 294).

388 Juan Antonio Carrillo Donaire, 'Notifying Authorities and Notified Bodies (Chapter III, Section 4). Standards, Conformity Assessment, Certificates, Registration (Chapter III, Section 5)' in Alejandro Huergo Lora and Gustavo Manuel Díaz González (eds), *The EU regulation on artificial intelligence: a commentary*.

389 Galland, 'The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies' (n 295).

requiring re-evaluation to ensure ongoing compliance. This system is made accountable through mandatory reporting obligations to the relevant notifying authority and the provision of a formal appeals process for providers who contest certification decisions.

Despite this highly structured framework, NBs face profound challenges, particularly in adapting their own internal processes to the technology they are meant to regulate. A significant paradox emerges from their low rate of AI adoption for their own assessment tasks, even as they acknowledge its potential benefits for efficiency and accuracy[390]. This reluctance is shaped by a series of salient, conflicting demands that hinder innovation. A primary tension exists between the institutional pressure to innovate and the restrictive nature of regulatory frameworks that prioritise stability and prescriptive compliance, compelling NBs to default to established, legally safer methods and leading to isolated, informal experimentation with AI by individual professionals rather than systemic integration. This is intertwined with the classic organisational tension between exploration and exploitation; NBs are hesitant to invest scarce resources in exploring novel AI-driven assessment tools, with their attendant implementation challenges and regulatory uncertainties, when exploiting existing, compliant manual processes offers a path of lower risk and guaranteed acceptability.

Furthermore, a paradox of collaboration versus isolation may hinder progress. While effective assessment would benefit from close collaboration between stakeholders, the strict legal requirements for impartiality and confidentiality erect functional barriers, preventing the kind of direct engagement with manufacturers that could foster co-development of more sophisticated assessment methodologies. This is compounded by a tension between efficiency and quality. NBs face mounting workloads from new regulations, creating an impetus to adopt AI for efficiency, yet they fear that doing so could compromise the quality and standardisation of their assessments, particularly if it involves opaque algorithms that are difficult to audit. Finally, an internal tension between autonomy and control pits the desire of individual expert assessors to flexibly implement new tools against the organisation's need for centralised control to ensure consistency and legal defensibility across all assessments. These latent paradoxes become particularly salient when activated by external triggers, such as the implementation of new, complex regulations like the AI Act, leading to a state of cautious inertia where significant opportunities for innovation in conformity assessment are postponed or missed entirely.

Beyond these internal operational challenges, the efficacy of Notified Bodies is contingent upon their integration within a broader ecosystem of coordination and oversight. Article 38 of the AI Act explicitly mandates coordination among NBs to prevent divergent interpretations of the Regulation's requirements, which could otherwise lead to an uneven application of the law and create opportunities for "forum shopping" by providers. This coordination, typically facilitated by the Commission, is essential for harmonising assessment practices and fostering a common interpretive community across the Union. The role of NBs is also complementary to that of Market Surveillance Authorities, the national bodies responsible for ex-post enforcement. While NBs conduct pre-market assessments, MSAs monitor products for compliance once they are in use, creating a crucial feedback loop. MSAs are vested with enhanced powers to oversee high-risk AI systems, including the authority to access data and source code, often acting on information and incident reports provided by NBs, users, and their own investigations. This regulatory architecture is further supervised by the European Commission, which retains the ultimate authority to oversee NBs and can, if a national authority fails in its duties, instigate the suspension or withdrawal of a body's designation. The dynamic and adaptive nature of AI technology underscores

[390] Ingela Mauritzon, Magnus Holmén and Lars Nordgren, 'Paradoxical by Design: Notified Bodies (Nbs) and Artificial Intelligence (Ai) Adoption in Conformity Assessment' (Social Science Research Network, 18 July 2025) <https://doi.org/10.2139/ssrn.5356928> accessed 7 August 2025.

the particular importance of these ex-post controls. Because an AI system's behaviour and risk profile can evolve after deployment, the initial ex-ante conformity assessment is inherently insufficient, rendering robust post-market monitoring and incident reporting paramount for ensuring long-term safety and compliance

Once designated, the primary operational obligation of a notified body is to verify the conformity of high-risk AI systems. For systems that are components of products already regulated under existing Union harmonisation legislation, as listed in Annex I, Section A, the conformity assessment bodies notified under those legislative acts may also assess compliance with the AI Act's requirements, provided they meet its specific criteria for independence and competence. For certain high-risk AI systems enumerated in Annex III, such as those for remote biometric identification, the provider is presented with a choice between two conformity assessment pathways. The first is an internal control procedure where the provider self-assesses its quality management system and technical documentation. The second, more stringent procedure involves the direct participation of a notified body.

In this second procedure, the provider must submit an application to a notified body, which then undertakes a comprehensive assessment of the provider's QMS against the criteria set out in Article 17, as well as a thorough examination of the system's technical documentation as specified in Annex IV. This scrutiny may involve requesting further evidence from the provider or conducting its own tests. Critically, should it be deemed necessary and after other avenues have been exhausted, the notified body is empowered to request extensive access to the provider's data, including full training, validation, and testing datasets, and potentially even to the training and trained models themselves, including source code, while respecting intellectual property rights and trade secrets. If the system is found to be compliant, the notified body issues a Union technical documentation assessment certificate, valid for a maximum of four years. If it is non-compliant, a refusal is issued with detailed reasoning. The notified body's involvement does not cease upon certification; it is required to conduct periodic surveillance audits of the approved QMS and may perform additional ad hoc tests of the certified AI system.

To ensure consistency and cooperation, the Commission facilitates coordination among all notified bodies, typically through a dedicated sectoral group, and participation is mandatory. While a notified body may subcontract specific tasks or use a subsidiary, it remains fully responsible for the work performed and must ensure the subcontractor meets all regulatory requirements. The designation of a notified body is not permanent; notifying authorities must monitor their bodies for ongoing compliance and have the power to restrict, suspend, or withdraw a designation if a body no longer meets the requirements or fails in its obligations. Such actions have a direct impact on the validity of certificates issued by that body, which may be suspended, withdrawn, or, under certain conditions, transferred to another notified body. The Commission itself can also intervene and, via an implementing act, withdraw a designation if it has justified doubts about a body's competence and the responsible Member State fails to take adequate corrective action. Notified bodies are also subject to extensive information obligations, requiring them to report on their certification activities to their notifying authority, market surveillance authorities, and other notified bodies, thereby contributing to the broader enforcement and market surveillance ecosystem established under Regulation 2019/1020. In setting their fees for these assessments, notified bodies are explicitly required to consider the specific needs of SMEs and start-ups, reducing fees proportionately to their size and market position.

6. Conclusion of the Chapter

This chapter has provided a comprehensive overview of the European Union's journey towards regulating AI, tracing the progressive evolution of EU policy from initial ethical guidelines and strategic communications to the final legislative proposal and its subsequent entry into force. This trajectory revealed a deliberate pivot from voluntary, 'soft law' approaches towards a binding legal framework, driven by a growing recognition of AI's transformative societal impact and the imperative to safeguard fundamental rights, ensure safety, and foster trust. The analysis then delved into the core regulatory mechanisms of the AI Act. It began with an examination of its foundational, yet complex, definition of an 'AI system' and the crucial interpretative role of the Commission's subsequent guidelines. The chapter then thoroughly examined the Act's hallmark risk-based approach, identifying its rationales for proportionality and efficiency while simultaneously scrutinizing its inherent limitations, such as the definitional ambiguities of 'high-risk' and the challenges posed by the static nature of Annex III in a dynamic technological landscape. Furthermore, the stringent essential requirements for high-risk AI systems, spanning risk management, data governance, technical documentation, record-keeping, transparency, human oversight, accuracy, robustness, and cybersecurity, were dissected, highlighting both their ambition and the practical implementation challenges they present for regulated entities.

The chapter then highlighted the instrumental role of harmonised standards and accredited notified bodies in the operationalisation and enforcement of the AI Act. Harmonised standards, functioning as critical bridges between abstract legal requirements and concrete technical specifications, were shown to facilitate compliance through providing technical solutions and enabling a presumption of conformity. However, a critical assessment revealed significant challenges, including concerns regarding the suitability of transposing the 'New Approach' framework to AI, a persistent democratic deficit in standard-setting due to the dominant influence of private bodies, and the slow pace of standard development relative to rapid technological advancement. Similarly, while notified bodies are posited as independent third-party evaluators vital for pre-market conformity assessment of high-risk AI systems, their capacity to effectively adapt their own internal processes and acquire the requisite multi-disciplinary expertise presents considerable practical and paradoxical challenges to their intended function.

Finally, despite the comprehensive nature of the AI Act, the chapter concluded by addressing several enduring complexities and open questions that will inevitably shape its future implementation and effectiveness. These include the persistent definitional nuances distinguishing various levels of AI autonomy in real-world applications, the evolving understanding of unacceptable risk in a rapidly changing technological landscape, and the formidable practical challenges associated with cross-border enforcement, resource allocation for effective market surveillance, and ensuring the framework's adaptability to future technological breakthroughs. Furthermore, the precise mechanisms for establishing legal responsibility and seeking compensation in cases of AI-induced harm remain areas requiring further clarity and refinement.

In sum, the AI Act represents a pioneering, albeit complex, regulatory undertaking that signifies a pivotal moment in global AI governance. Its design, rooted in a human-centric philosophy and a proportionate, risk-based approach, attempts to balance innovation with robust protection for fundamental rights and safety. However, the multifaceted challenges identified within this chapter, ranging from definitional and interpretive ambiguities to the practicalities of implementation, standardisation, and third-party oversight, underscore that the journey towards effective and trustworthy AI governance is an ongoing, dynamic process. These foundational insights into the AI Act's structure, rationale, and inherent complexities provide the necessary groundwork for the next chapters of this

dissertation, which will delve deeper into the specific limitation of its operationalisation via harmonised standards.

Chapter 5.
What standards can't do: an evaluation of the regulatory effectiveness of technical standards for the regulation of AI

1. Aim of the Chapter

As discussed in the previous chapter, technical standards have assumed an increasingly prominent role in AI regulation, positioned as essential instruments for translating high-level legal principles into actionable technical and operational requirements. Within the specific context of the AI Act, there exists a clear legislative expectation that harmonised standards will provide the necessary technical specifications for the effective implementation of the essential requirements. This chapter, however, aims to critically interrogate this foundational assumption. Through a metaethical analysis of several key provisions of the Act, it becomes evident that many of its core obligations are inherently characterised by a significant degree of vagueness and profound contestability. The mechanism of delegation to technical standards established by the AI Act implies that value-laden issues embedded within the Act's requirements are effectively transferred to standardisation bodies, which are tasked with operationalising these high-level requirements. As argued in this chapter, however, a range of factors make it highly unlikely that standards can deliver unequivocal answers to such complex problems. The inherent diversity of perspectives among stakeholders, the ambiguity of certain normative concepts, and the limitations of technical methodologies all contribute to this persistent indeterminacy, suggesting that technical standards alone cannot fully resolve the ethical and societal challenges posed by AI regulation.

This inherent vagueness and contestability consequently generate critical decision problems for regulatees, who are nevertheless called to implement the AI Act requirements to release compliant systems on the EU market. These entities are forced to resolve fundamental value-laden questions, for instance on acceptable level of risk, that are not uniquely determined by either the regulatory text or the accompanying technical standards. This indeterminacy effectively delegates a considerable margin of discretionary judgment to the actors involved in the conformity assessment process, whether providers themselves or third parties notified bodies. Consequently, the analytical focus must necessarily shift towards the crucial role of these conformity assessment procedures, which emerge as the de facto institutional locus where these contested value questions will be pragmatically resolved. This delegation raises significant open problems concerning accountability, consistency, and the potential for divergent interpretations, all of which merit extensive exploration in future research.

2. Understanding regulatory effectiveness in the context of AI standards

Given that the central objective of this chapter is to analyse the regulatory effectiveness of technical standards within the domain of artificial intelligence, it is first necessary to establish a precise conceptual

framework for what "regulatory effectiveness" entails. At its most fundamental level, regulatory effectiveness is concerned with the extent to which a given regulatory intervention succeeds in contributing to the achievement of its intended effects[391]. Since regulatory interventions are typically designed to channel and modify the conduct of a target population the primary measure of their effectiveness lies in whether the intervention impacts the behaviour of regulatees in the manner envisioned by the regulator.

A more granular examination of regulatory effectiveness reveals that its assessment hinges upon two principal components[392]. The first is the demonstrable impact of the intervention, whether this impact is articulated in terms of outputs, the direct results of regulatory action, or outcomes, the broader societal or economic consequences. The second component is the degree to which this impact aligns with the stated regulatory purposes, be they articulated as aims, targets, or specific objectives.

The attainment of regulatory effectiveness is contingent upon a confluence of favourable conditions, the absence of which can lead to regulatory failure. A regulatory environment is deemed effective when the intervention is demonstrably fit for its intended purpose. Conversely, ineffectiveness can emanate from three primary sources. The first is regulator failure, which occurs when regulators lack the requisite integrity, intelligence, or resources to execute their functions in a manner that serves the regulatory objectives[393]. The second source is regulatee resistance, wherein the subjects of the regulation fail to comply, for any number of reasons[394]. This form of failure is often symptomatic of a lack of consensus among regulatees regarding the appropriateness of the intervention and can manifest proactively, through lobbying for loopholes or exemptions during the standard-setting phase, or reactively, through non-compliance after the rules are established. The third source involves disruptive externalities or the intervention of third parties, such as a rival regulator, which introduces circumstances beyond the control of both the primary regulator and the regulatees[395]. Consequently, a regulatory environment is poised for effectiveness when regulators are competent and adequately resourced, regulatees are predisposed to compliance, and the broader context is free from disruptive external forces. These conditions must hold throughout the entire regulatory cycle, from the initial setting of rules to the subsequent monitoring of compliance and the application of corrective measures for non-compliance.

A significant methodological challenge in the study of regulatory effectiveness lies in establishing an appropriate threshold for success[396]. If the threshold for effectiveness is set at the absolute and complete achievement of all regulatory objectives, virtually every regulatory intervention would be judged a failure. Speed limits on roads, for example, do not prevent every instance of speeding, yet they are not generally regarded as entirely ineffective. A more pragmatic and analytically useful approach, therefore, is to assess the relative effectiveness of an intervention, that is, to determine whether it is having an impact compared to a baseline scenario. This leads to the widely accepted recommendation that regulatory effectiveness should be treated not as a binary, all-or-nothing quality, but rather as a matter of degree. Such a graduated

---

391 Goodwin and Brownsword (n 24).

392 ibid.

393 Roger Brownsword and Morag Goodwin, 'Regulatory Effectiveness II: Failure by Regulators' *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://doi.org/10.1017/CBO9781139047609.017> accessed 26 October 2025.

394 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://doi.org/10.1017/CBO9781139047609.018> accessed 30 May 2025.

395 Roger Brownsword and Morag Goodwin, 'Regulatory Effectiveness IV: Third-Party Interference and Disruptive Externalities' *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://doi.org/10.1017/CBO9781139047609.019> accessed 26 October 2025.

396 Goodwin and Brownsword (n 24).

perspective is particularly apposite for interventions designed to incentivise or promote certain behaviours, such as a patent regime intended to foster innovation, where a definitive ceiling for absolute effectiveness is conceptually elusive.

Having established this general account of regulatory effectiveness, it is now possible to contextualise the concept within the specific domain of artificial intelligence standardisation. In undertaking this contextualisation, it is imperative to recognise that within a co-regulatory framework such as that established by the AI Act, the regulatory objectives against which effectiveness is measured are not endogenous to the standards themselves. Rather, they are exogenously determined by the overarching legislative instrument that delegates to standards the function of operationalising its essential requirements. Consequently, the regulatory effectiveness of a given standard, both in the context of AI and more generally, is not a measure of its own intrinsic goals but of its functional adequacy in providing the technical means to achieve the objectives stipulated by the corresponding legislation[397].

For harmonised standards under the AI Act, this means their effectiveness must be assessed by their capacity to translate the high-level, and often abstract, legal principles concerning the protection of health, safety, and fundamental rights into clear, verifiable, and state-of-the-art technical specifications. The ultimate purpose of the regulation is to ensure a high level of protection for these values. Therefore, an AI standard is regulatorily effective to the extent that its faithful implementation by providers of AI systems directly and demonstrably contributes to the realisation of these legislative aims, thereby enabling a presumption of conformity with the law's essential requirements. The standard, in this view, functions as an intermediary instrument, and its effectiveness is contingent upon its success in bridging the gap between legal prescription and technical practice. Accordingly, the problem of the regulatory effectiveness of technical standards investigated in this section concerns their suitability to provide precise and verifiable technical requirements that are aligned with the goals and essential requirements of the corresponding legislation, thereby ensuring that the technical specifications effectively address the safety, health, environmental, fundamental rights, or any other risks that the legislation seeks to mitigate. This issue of regulatory effectiveness is distinct from the question of legitimacy widely discussed in the literature and analysed in Chapter 3 Regulatory effectiveness is about whether standards can achieve their intended purpose, whereas legitimacy is about whether standards and standards makers have the right to set those standards in the first place. The problem of regulatory effectiveness is thus an epistemic one, focused on the technical feasibility and efficacy of standards, rather than a normative one, concerned with the source of their legitimacy.

This issue has received limited attention in legal scholarship, and it has also been sidestepped in legal practice, as evidence in the Stichting Rookpreventie Jeugd and Others (C-160/20) case before the European Court of Justice[398]. This case is instructive because it demonstrates the difficulties in dealing with the regulatory effectiveness of technical standards.

In this case, Stichting Rookpreventie Jeugd and 14 other consumer health protection associations contested the reference to the standards ISO 4387, ISO 10315, ISO 8454, and ISO 8243 in the Article 4 of the Tobacco Products Directive[399]. These standards are used as methods for measuring tar, nicotine, and carbon monoxide emissions from cigarettes to comply with the maximum emission levels prescribed in Article 3 of the Tobacco Products Directive.

---

[397] Tartaro, 'Value-Laden Challenges for Technical Standards Supporting Regulation in the Field of AI' (n 55).

[398] *Stichting Rookpreventie Jeugd and Others v Staatssecretaris van Volksgezondheid, Welzijn en Sport* [2022] ECJ Case C-160/20.

[399] Directive 2014/40/EU of the European Parliament and of the Council of 3 April 2014 on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products and repealing Directive 2001/37/EC (Text with EEA relevance)Text with EEA relevance 2023.

The plaintiffs argued that the measurement methods in the ISO standards are inadequate to implement the requirements of the Directive, essentially questioning their regulatory effectiveness as understood in this chapter. The ISO standards mandate the use of a smoking machine to measure the emission levels of cigarettes. However, according to the plaintiffs, this measurement method does not account for how a real person smokes, where fingers and lips cover small holes in the filter. In the smoking machine, unobstructed small holes in the filter allow clean air to mix with smoke. However, when these small holes are closed by fingers and lips in real-world smoking, the concentration of harmful substances rapidly grows. According to scientific evidence presented by the plaintiffs, the "smoking machine", regulated by the ISO standards and mandated by Article 4(1) of the Tobacco Products Directive, records emission quantities up to 20 times lower than those actually produced.

When called to rule on this case, the ECJ did not examine the content of these standards or assess their suitability in achieving the Tobacco Products Directive's goals, i.e., to protect consumers' health in the EU. Instead, it framed the issue as one of transparency and legal certainty concerning references to international standards in EU law.

While these are undoubtedly relevant legal issues, Canale provides compelling reasons to argue that this was not the core of the matter[400]. Rather, the central question was whether the referenced ISO standards significantly undermined the objective pursued by the Tobacco Products Directive, as expressed in Article 1 and the preamble: the protection of consumer health. If this were the case, the regulatory effectiveness of these standards would be lost.

Evaluating this question would have necessitated delving into the technical content of the ISO standards. In other words, the ECJ would have had to assess whether the measurement methods in the ISO standards were indeed flawed and if alternative methods, such as the Canadian intense method, were not more accurate and therefore more effective in implementing the Directive. However, as Schepel argues, "judicial scrutiny of the technical merits of the standard in question would be the ultimate nightmare scenario[401]." There could be many reasons for this, both legal and non-legal. Among these, Canale seems to provide the most convincing argument: judges lack the expertise to make technical judgments and are compelled to defer to the experts who establish the technical rules. This approach, however, is problematic because it creates opacity in the law and excludes consideration of conflicting epistemic reasons, such as those presented by the plaintiffs[402].

The *Stichting Rookpreventie Jeugd* case thus serves as a powerful illustration of the inadequacy of conventional judicial review as a mechanism for assessing the regulatory effectiveness of technical standards. This inadequacy stems from the judiciary's institutional reluctance to engage in substantive scrutiny of the technical content of standards. Consequently, courts are ill-equipped to determine whether the technical specifications embedded within a standard are actually capable of furnishing the means to achieve the high-level objectives, such as the protection of public health, that the delegating legislation aims to secure. This deficit in *ex post* judicial oversight is compounded by a parallel absence of robust *ex ante* procedures.

This systemic vulnerability in the co-regulatory model is not confined to established sectors like tobacco products; indeed, there is significant reason to anticipate its emergence, perhaps in a more acute form, within the nascent regulatory landscape for artificial intelligence. The high-level requirements

[400] Damiano Canale, 'Quando gli esperti creano diritto: deferenza, opacità, legittimità' (2022) 22 Analisi e Diritto <https://doi.org/10.4454/analisiediritto.v22i2.508> accessed 9 April 2024.

[401] Harm Schepel, *The Constitution of Private Governance: Product Standards in the Regulation of Integrating Markets* (Hart Publishing 2005).

[402] Damiano Canale, 'The Opacity of Law: On the Hidden Impact of Experts' Opinion on Legal Decision-Making' (2021) 40 Law and Philosophy 1 <https://doi.org/10.1007/s10982-021-09408-8>.

articulated in the AI Act are characterised by a significant degree of conceptual vagueness and inherent contestability, as discussed in the next section. This ambiguity creates considerable friction in the translation process, presenting a formidable challenge for the ESOs tasked with operationalising these abstract legal principles into concrete, verifiable, and state-of-the-art technical specifications. The very contestability of a term like "acceptable risk" or "appropriate level of accuracy", which can be interpreted in numerous, often conflicting, ways, means that any resulting standard is inherently susceptible to the same critique levelled in the *Stichting Rookpreventie Jeugd* case: namely, that its technical implementation fails to adequately capture and realise the legislative intent.

3. Vagueness and contestability in the AI Act essential requirements

The challenge of ensuring regulatory effectiveness is acutely exacerbated in the domain of artificial intelligence by several key provisions within the AI Act. As will be explored in this section, the very formulation of these requirements is such that they raise profoundly value-laden questions, which technical standards, by their nature, are ill-equipped to resolve. The reliance on standardisation as a primary mechanism for operationalising the Act's principles thus creates a significant tension, as the translation of high-level legal norms into precise, verifiable metrics encounters the irreducible complexity of ethical and societal values.

To dissect these complexities, this analysis will adopt a metaethical lens[403], revealing how specific terminology embedded within the AI Act's requirements is inherently value-laden and, consequently, necessitates normative judgments for its implementation. The deployment of such terms confers upon these legal provisions a distinct form of vagueness. This is not merely a matter of linguistic imprecision but rather vagueness in the particular sense articulated by Jeremy Waldron, who adapts the concept from Wittgensteinian philosophy[404]. Vagueness, in this context, denotes a situation where different subjects may legitimately and reasonably disagree on whether a specific term applies to a given set of facts or a particular instantiation of a system[405].

Furthermore, an assessment of the correct implementation of these requirements is contingent upon specific, and often unstated, normative assumptions. This dependency on underlying value commitments gives rise to what Waldron, in the same theoretical framework, defines as contestability[406]. This makes them instances of "thick concepts"[407]. More fundamentally than vagueness, contestability pertains to the persistent and principled disagreements that can arise concerning the correct application of a concept, even when all relevant empirical facts are agreed upon. Such divergence is possible, and indeed probable, because the correctness of the application itself is predicated on differing foundational value assumptions held by the assessing parties. Consequently, the operationalisation of the AI Act's requirements becomes a site of potential normative conflict, where divergent interpretations are not merely a result of ambiguity but of deeply rooted disagreements about the values the regulation is meant to instantiate.

To illustrate the pervasive nature of vagueness and contestability within the AI Act, it is instructive to examine several critical examples. To begin with, Article 9 of the AI Act mandates the establishment of a comprehensive risk management system for high-risk AI systems. Notably, this article stipulates that

[403] Gianluca Verrucci, *Introduzione Alla Metaetica* (FrancoAngeli 2014).

[404] Ludwig Wittgenstein, *Philosophical Investigations* (3nd ed., repr, Blackwell 1953).

[405] Jeremy Waldron, 'Vagueness in Law and Language: Some Philosophical Issues' (1994) 82 California Law Review 509 <https://doi.org/10.2307/3480971>.

[406] ibid.

[407] Väyrynen (n 22).

risk management measures must be implemented to ensure that "the overall residual risk of the high-risk AI systems is judged to be acceptable" (Article 9(5)). However, the legal text conspicuously omits a definition of what constitutes an "acceptable risk"[408]. This raises fundamental questions regarding the process and the individuals or entities responsible for determining when a risk has been minimised to an acceptable level.

This vagueness extends to other value-laden domains throughout the AI Act. For instance, Article 15(1) imposes an obligation on high-risk AI systems to achieve an "appropriate level of accuracy, robustness, and cybersecurity." Focusing specifically on accuracy, an identical challenge to that identified in Article 9 emerges: there is no clear indication of who establishes and by what criteria the appropriateness of accuracy levels is determined. The contextual dependencies of AI systems, their specific use cases, and a myriad of other factors can significantly influence what is considered an appropriate level of accuracy.

Furthermore, the requirements for human oversight articulated in Article 14(3) further exemplify this issue. This provision mandates that oversight measures shall be "commensurate with the risks, level of autonomy, and context of use" of the high-risk AI system. It is readily apparent which term within Article 14(3) presents a critical indeterminacy. Specifically, the questions surrounding the acceptable level of risk, the appropriate level of accuracy, and the commensurate degree of human oversight serve as salient examples of "hard normative questions" left open in the AI Act[409]. These questions are fundamentally characterised by their reliance on inherently normative concepts such as "acceptable," "appropriate," and "commensurate," thereby necessitating the establishment of preferred trade-offs among competing interests. The responsibility for resolving these complex normative inquiries is, at first glance, entirely devolved to the domain of technical standards.

We contend that the terms "acceptable," "appropriate," and "commensurate" within the AI Act are inherently vague and contestable. To rigorously support this assertion, we draw upon insights from the field of metaethics. Metaethics, in essence, concerns the philosophical study of the nature, scope, and meaning of ethical concepts and judgments[410]. It investigates questions such as whether moral values are objective or subjective, whether moral language is descriptive or prescriptive, and how moral knowledge is acquired. By examining the foundational principles of metaethics, we can gain a deeper understanding of how these concepts function and why their application in legal frameworks can lead to indeterminacy.

The metaethical lens offers a potent framework for analysing the vagueness and contestability of these critical terms. The ambiguity inherent in moral discourse, as studied in metaethics, directly mirrors into regulatory language. When ethical concepts are employed, they often carry with them a spectrum of interpretations and valuations, making their application context-dependent and subject to individual or collective value judgments. This inherent indeterminacy, a core concern of metaethics, translates into legal provisions that require interpretation.

The vagueness of legal language has been a subject of extensive discussion within legal scholarship[411]. Applying insights from Wittgenstein's philosophy of language[412], Waldron argues that vagueness is an

[408] Fraser and Villarino (n 361).

[409] 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 <https://doi.org/10.1016/j.clsr.2024.105957>.

[410] Sayre-McCord (n 21).

[411] Timothy AO Endicott, 'Vagueness and Legal Theory' (1997) 3 Legal Theory 37 <https://doi.org/10.1017/S135232520000063X>.

[412] Wittgenstein (n 404).

integral component of a broader indeterminacy present in legal language, encompassing not only vagueness but also ambiguity and contestability[413].

A term is considered ambiguous when it admits of multiple distinct meanings. For example, the sentence "Flying planes can be dangerous" exhibits ambiguity as it can be interpreted to mean that the act of piloting an aircraft is perilous or that aircraft in flight pose a hazard. However, not all instances of ambiguity in legal language are inherently problematic, as they can often be resolved through the provision of supplementary information that clarifies the intended meaning. This phenomenon is also observable in everyday language, where contextual cues or additional information typically facilitate the determination of the intended meaning of a sentence such as "Flying planes can be dangerous."

Conversely, a term is deemed vague when it introduces uncertainty regarding its applicability to a particular object or situation. This uncertainty does not stem from a deficiency in information pertaining to the object or situation itself, but rather from the inherently fuzzy boundaries of the word's semantic scope. Consider, for instance, the question of what constitutes a "hot" temperature. Whether 27°C is classified as hot or warm is contingent upon context and individual perception. The absence of a definitive cut-off point for the descriptor "hot" renders the term vague.

Finally, a word or phrase is characterised as contestable if it possesses two key attributes. Firstly, akin to vagueness, it permits multiple reasonable applications of the term to different objects or situations. Secondly, and crucially, the term inherently incorporates a value judgment or evaluative assessment. Disagreements arise regarding the precise objects or situations to which the term applies because individuals hold differing values. For example, "fair" and "fairness" are contestable terms, as they can be understood through divergent perspectives, such as equity or equality.

Based on this analytical framework, it becomes evident that terms such as "acceptable," "appropriate," and "commensurate" within the AI Act are both vague and contestable. Let us consider "acceptable" as an illustrative example. First, the conceptual boundaries of the term "acceptable" are inherently fuzzy. Consequently, there exists uncertainty regarding whether this term can be meaningfully applied to a given risk situation. Second, the decision to apply, or not apply, the term involves a value judgment. What one individual or group may deem an acceptable level of risk could be entirely unacceptable to another. This inherent subjectivity renders the notion of an "acceptable risk level" contestable.

Similar conceptual challenges are encountered with the determination of "appropriate" levels of accuracy and the formulation of "commensurate" oversight measures. The term "appropriate" is devoid of objective, universally applicable benchmarks; its meaning is entirely contingent upon the specific domain of application and the potential consequences of system failure. For instance, the threshold for appropriate accuracy in a low-stakes recommender system for music streaming is vastly different from that required for an automated diagnostic tool in clinical oncology or a predictive policing algorithm. Furthermore, the adjudication of appropriateness is itself a locus of contention, implicating a trade-off between competing axiological frameworks. A model optimized for raw predictive accuracy, which a developer might argue is "appropriate," could simultaneously perpetuate systemic biases against protected groups, a result that other stakeholders would rightly deem profoundly inappropriate. The term thus becomes a proxy for a deeper debate over whether system design should prioritise technical performance, distributive fairness, or other societal values.

Likewise, the legislative requirement for human oversight to be "commensurate" with the risks posed by an AI system introduces a comparable degree of vagueness and contestability. The principle of

[413] Waldron (n 405).

proportionality is central, yet the AI Act provides no clear calculus for how this commensurability is to be measured or achieved. This lack of a formal methodology transforms the requirement into a qualitative standard that is susceptible to widely divergent interpretations. A technology firm might argue that a system of ex-post audits and notifications constitutes oversight commensurate with a system's risks. Conversely, a human rights organisation could contend that for the very same system, nothing short of continuous, real-time human-in-the-loop intervention would be commensurate with the potential impact on fundamental rights. The determination of what is "commensurate" is therefore not a technical exercise to be encapsulated in a standard, but a political negotiation, reflecting a balance between competing desiderata such as innovation, economic competitiveness, public safety, and individual autonomy. This reveals that such terms function less as precise legal standards and more as arenas for ongoing normative deliberation, where the ultimate meaning is forged through regulatory practice, judicial interpretation, and societal debate, not technical standards.

4. Technical standards cannot provide technical means for implementing vague and contestable requirements

One of the central arguments of this chapter posits that there is a profound structural misalignment between the legislative intent of the AI Act and the actual capabilities of technical standards, which are tasked with providing the technical means to implement the Act's vague and contestable requirements. To elucidate why this is the case, it is necessary to analyse the fundamental assumption underpinning the legislative technique wherein EU legislation specifies objectives while technical standardisation specifies the technical means for achieving them. This assumption is deeply rooted in the European approach to technical regulation, which presumes that technical standards can provide a further, functional specification of the abstract requirements outlined in the primary legislation. A critical examination of this premise is essential to understanding why this approach is destined for insufficiency within the specific context of artificial intelligence governance.

It is important to acknowledge that the assumption in question is not inherently flawed; indeed, it maintains validity across numerous regulatory instances. This logic serves as the bedrock for the regulation of various products and technologies governed by the New Approach and the NLF, including toys, medical devices, fertilizing products, cosmetics, and radio equipment. In these domains, the assumption is generally effective, and the regulatory approach derived from it operates efficiently, allowing technical standards to provide the necessary specifications to achieve the objectives of the corresponding legislation.

A paradigmatic example is the Toy Directive[414], which is supported by the EN 71 series of standards[415], which establish mechanical, physical, chemical, and electrical properties to ensure toy safety. The utilisation of technical standards to specify regulatory requirements in such cases is advantageous, as it facilitates the incorporation of state-of-the-art knowledge and the maintenance of up-to-date safety practices. By decoupling the essential requirements, which remain fixed within the legislative text, from

[414] Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys (Text with EEA relevance)Text with EEA relevance 2022.

[415] Commission Implementing Decision (EU) 2023/740 of 4 April 2023 on harmonised standards for toys drafted in support of Directive 2009/48/EC of the European Parliament and of the Council 2023 (OJ L); Commission Implementing Decision (EU) 2025/1785 of 9 September 2025 amending Implementing Decision (EU) 2023/740 as regards the harmonised standards for certain ride-on toys (wave rollers) drafted in support of Directive 2009/48/EC of the European Parliament and of the Council 2025.

the technical specifications required to satisfy them, the regulatory regime achieves a necessary degree of agility. This separation allows for the rapid integration of scientific advancements and engineering innovations without necessitating the arduous and time-consuming process of legislative amendment. Consequently, standards serve as a dynamic repository of technical consensus, ensuring that compliance mechanisms evolve concurrently with the technological landscape they govern. Furthermore, this delegation to standardisation bodies leverages the specialised expertise of technical committees, thereby translating abstract legal safety obligations into measurable, verifiable metrics, such as toxicity limits or flammability thresholds, that provide legal certainty to manufacturers through the presumption of conformity.

However, the efficacy of this model is not absolute. Cases such as the standards for measuring tobacco emissions reveal distinct limitations to this assumption. Measurement uncertainty and potential flaws embedded in the design of measurement methods can compromise the ability of technical standards to accurately specify regulatory requirements, potentially undermining the objectives of the corresponding regulation as discussed in Section 2 of this Chapter. Nevertheless, in the context of material measurement standards, it remains possible to satisfy the assumption that standards can further specify regulatory requirements, provided that the law or standards are revised to address methodological shortcomings.

When applying this framework to the AI Act, however, the assumption that standards can further specify the regulation becomes entirely untenable. As argued previously in section, this failure stems from the inherent nature of the AI Act's requirements. It is arguably impossible to technically specify what constitutes an "acceptable" level of risk, an "appropriate" level of accuracy, and a "commensurate" level of human oversight solely through technical standards, exactly because they are subject of normative deliberations rather than technical standardisation.

One might posit a counter-argument: if standards in other contexts provide very specific values and thresholds, why can they not do so here? For instance, the standard "EN 71 Safety of toys - Part 7: Finger paints - Requirements and test methods,"[416] which supports the Toy Directive, stipulates that the concentration of climbazole, a preservative used to prevent microbiological spoilage, in finger paints shall not exceed 0.2%. Why, then, cannot we envision a standard that determines, for example, that the appropriate level of accuracy for an AI-based melanoma diagnosis system entails fewer than 13.6% false negatives?[417] The answer to this question is complex and necessitates a brief sketch of the epistemological differences between the fields. The fundamental reason is that, in the case of AI, there is no methodological and factual basis for establishing that a specific threshold, such as 13.6%, is the appropriate limit to minimise risks to health, safety, and fundamental rights.

The selection of such a threshold depends on complex value judgments which are not grounded in facts that can be constructed and assessed according to shared scientific methods. This threshold embodies value judgments about the importance of being accurate in this sensitive context, as well as the normative assumption that AI systems should be deployed only when they match or outperform humans. However, technical standards currently do not, and arguably cannot, make such value judgments. This is

[416] EUROPEAN COMMITTEE FOR STANDARDIZATION, 'EN 71-7:2014: Safety of Toys - Part 7: Finger Paints - Requirements and Test Methods' (April 2014) <https://law.resource.org/pub/eu/toys/en.71.7.2014.html> accessed 18 November 2025.

[417] 13,6% is the rate of diagnostic error for melanoma according to state of the art research on this subject David E Newman-Toker and others, 'Rate of Diagnostic Errors and Serious Misdiagnosis-Related Harms for Major Vascular Events, Infections, and Cancers: Toward a National Incidence Estimate Using the "Big Three"' (2021) 8 Diagnosis 67 <https://doi.org/10.1515/dx-2019-0104>.

because standardisation, being the product of multi-stakeholder process, brings varying values, interests, and perspectives to the table, making it challenging to reach consensus on complex value-laden issues.

In contrast, the determination of safe levels of climbazole in finger paints is based on a well-established scientific foundation. Regulatory sciences play a crucial role in testing, evaluating, and monitoring technologies, their effects, and their risks[418]. Institutionalised bodies, such as the Scientific Committee on Consumer Safety (SCCS), possess the capability to determine safe levels of substances like climbazole in various products according to shared metrics, accepted literature, and methods well-established in toxicology[419]. In stark contrast, the field of AI lacks a scientific consensus on the methods to identify, evaluate, and mitigate the risks arising from algorithmic systems. This scientific foundation is essential for the regulation of technologies and their risks, as this activity requires proofs and evidence in the same manner as scientific research. As Demortain notes, before any scientific method of assessing health, safety, or fundamental rights risks of a technology can be advanced, "there needs to be an agreement around the fact that this risk is a relevant experience to measure. This experience needs to be named and categorised, and it needs to be agreed-upon as real, as well as valued as something that needs to be measured and controlled"[420].

Unfortunately, the field of AI has not yet reached this level of paradigmatic maturity. This implies that attempting to regulate AI with the support of technical standards, as is done in fields with an established scientific foundation, may be inadequate, particularly when hard normative questions are involved. This problem is particularly acute in areas where AI is proliferating and creating risks for which there are no pre-existing, robust, and well-established evaluation practices to understand and manage them[421]. For example, the risks associated with automatic content moderation affecting the right to freedom of expression, the psychological effects of interaction with artificial agents, and the fairness of recruitment processes or the assignment of educational and social benefits based on AI, represent areas where it is difficult to establish shared criteria and thresholds for acceptability, commensurability, or appropriateness. Conversely, in areas where AI intersects with risks that are already recognised and understood, and for which risk management practices exist, such as health and safety risks in the automotive and healthcare sectors, it may be possible to manage these problems by adapting existing methods, criteria, and thresholds, although this involves challenges that cannot be fully explored here.

This analysis suggests that a reassessment of the scope of what technical standards can accomplish at this juncture in the field of AI is necessary, requiring a tempering of regulatory expectations. Currently, standards can at best provide a set of methods and processes for managing risk, measuring and evaluating accuracy, and implementing oversight measures. While this represents significant progress, technical standards are not equipped to make evaluative judgments about the acceptability of risk, the appropriateness of accuracy, or the adequacy of oversight. The proposed solution is therefore cautious, inviting a critical evaluation of the suitability of the tools employed for achieving the proposed objectives. Based on this analysis, it appears that technical standards are not yet capable of delivering what the AI Act requires. The assumption that technical standards can technically specify the requirements of the

[418] David Demortain, 'Expertise, Regulatory Science and the Evaluation of Technology and Risk: Introduction to the Special Issue' (2017) 55 Minerva 139 <https://doi.org/10.1007/s11024-017-9325-1>.

[419] SCCS, 'ADDENDUM to the Scientific Opinions on Climbazole (P64) - Ref. SCCS/1506/13 and SCCS/1590/17" SCCS/1600/18'
<https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_220.pdf>.

[420] Demortain (n 418).

[421] Piercosma Bisconti and Marcello Galisai, 'Standards for Trustworthy AI in the European Union: Technical Rationale, Structural Challenges, and an Implementation Path' (arXiv, 21 January 2026) <https://doi.org/10.48550/arXiv.2602.00078> accessed 20 April 2026.

regulation is unfounded in this context. As a result, alternative solutions, which will be explored in section 7 of this Chapter, may be more effective for regulating AI.

Nevertheless, the challenge of complying with the AI Act remains, and addressing its vague and contestable requirements is essential. This leaves the analysis with a critical, yet unanswered question: how can regulatee achieve compliance with the requirements of the AI Act, when the technical standards do not provide a definitive solutions for implementing the vague and contestable requirements?

5. Decision problems under the AI Act

To properly answer the previous question, we first need to reframe it into more tractable dimensions. Indeed, the strategic postures adopted by regulated entities, or regulatees, mostly AI providers, in their pursuit of alignment with regulatory mandates exhibit significant heterogeneity. This spans from highly risk-averse, conservative methodologies, wherein statutory requirements are subjected to rigorous, often literal interpretation. In such scenarios, operational protocols are executed with exactitude to satisfy these strictly construed obligations, ensuring that organisational conduct remains safely within the bounds of legal foresight. Conversely, the continuum extends to far more permissive approaches, characterised by a relaxed adherence to governance structures which may, in extreme manifestations, involve the deliberate disregard of regulatory imperatives. In such contexts, potential sanctions are frequently stripped of their normative or punitive signalling and are instead re-contextualized purely as economic inputs. Here, the decision to exhibit non-compliance is not necessarily a product of negligence, but rather the result of a calculated cost-benefit analysis. Under this utilitarian calculus, potential pecuniary penalties are internalised as variable operating costs associated with the development of an otherwise highly remunerative project. Consequently, if the projected returns on investment sufficiently eclipse the probabilistic costs of enforcement and subsequent fining, the regulatee may view the absorption of such penalties as a rational business expense, effectively a "licensing fee" for infraction, rather than a compliance failure.

Consider, for instance, the operational reality of a regulatee tasked with determining whether a high-risk artificial intelligence system currently under development presents a level of residual risk that may be deemed "acceptable," as mandated by Article 9(5) of the AI Act. Because this statutory provision constitutes an inherently vague and contestable requirement, devoid of rigid quantitative definitions, the ultimate determination of compliance is susceptible to significant epistemic fluidity. Indeed, the methodological latitude afforded to the regulatee means that granular variations in the risk calculus, including the specific algorithms employed for probability estimation, the calibration of sensitivity thresholds, and the discretionary selection of variables included in the assessment model, can all exercise a determinative influence on the judgment of whether a specific risk profile is admissible.

Consequently, this interpretative ambiguity allows for a bifurcation of compliance strategies. Regulatees characterised by a conservative institutional logic may impose stringent criteria for risk acceptability, prioritising a precautionary ethos to minimise liability exposure. Conversely, entities adopting a more permissive posture may implement significantly laxer parameters, widening the scope of permissible risk to favor innovation or speed of deployment. Crucially, in the absence of unequivocal, standardised criteria delineating the precise boundaries of acceptable risk within the algorithmic domain, both these divergent methodologies—despite their radically different substantive outcomes—could demonstrate formal adherence to the regulatory requirement of risk acceptability.

To avoid reducing this phenomenon to a mere exercise in strategic manoeuvring, whether opportunistic or otherwise on the part of the regulatee, it is necessary to look beyond the surface level of compliance behaviour. Regardless of the concrete strategy employed, be it a stringent adherence to the precautionary principle or a more relaxed interpretation of safety parameters, the situation facing the regulated entity can be analytically framed as a formal decision problem. In this context, drawing upon the foundational definitions provided by Fischhoff, a decision problem is understood not simply as a choice between clear alternatives, but as a complex structural arrangement characterised by three essential elements: a set of possible courses of action (options); a set of events or states of the world that are beyond the decision-maker's control (uncertainties); and the consequences that result from the interaction of these options and events, which must be evaluated against a set of values or outcomes[422].

Determinations of risk acceptability are, by their very nature, archetypal decision problems. 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.

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. When grappling with unavoidable accident scenarios, the provider is not tasked with scripting explicit moral responses to discrete dilemmas, but rather with navigating a structural decision problem defined by the architecture of trajectory planning algorithms and optimisation logic. The vehicle's behaviour in critical situations is not the product of hard-coded ethical subroutines, but rather emerges consequentially from the calibration of cost functions, wherein the engineering team must assign numerical weights to competing variables such as vehicle stability, passenger G-force impact, and the predicted trajectories of vulnerable road users. These technical design choices implicitly codify the system's normative orientation; for example, assigning a disproportionately high penalty to ego-vehicle damage or strictly limiting lateral acceleration to ensure passenger comfort effectively programs a preference for occupant safety, potentially precluding evasive manoeuvres that might save a pedestrian but imperil the vehicle's integrity. Conversely, a utilitarian configuration designed to minimise the aggregate sum of predicted trauma across all actors in the scene might dictate a trajectory that exposes the vehicle's occupants to greater mechanical risk in order to avoid a catastrophic collision with a crowd. Consequently, the determination of acceptable risk is transposed from a legal abstraction into a mathematical calibration, forcing the regulatee to determine whether the system should be optimized to protect the cabin as an inviolable constraint or to reduce the total casualty count regardless of the victim's location. This is a profound ethical decision problem with no single "correct" answer, which goes well beyond the abstractions of the "moral machine"[423].

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, they remain highly relevant due to their intersection with the Digital Services Act[424]. Is it more "acceptable" to implement highly aggressive filtering algorithms for potentially harmful content, thereby risking the erroneous removal of legitimate

[422] Fischhoff and others (n 9).

[423] Edmond Awad and others, 'The Moral Machine Experiment' (2018) 563 Nature 59 <https://doi.org/10.1038/s41586-018-0637-6>.

[424] 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.

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 straightforward, universally agreed-upon, or purely technical solutions are often elusive. A comparable tension emerges in the design of recommender systems, where algorithmic choices can subtly shape user autonomy by influencing exposure to information, preferences, and even identity formation; while such systems can help users navigate complex information environments, they may also introduce risks of manipulation, reduced critical engagement, and epistemic narrowing, thereby requiring similarly careful ethical balancing in their design and deployment[425]

In these situations, 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[426]? 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[427]. These are not purely technical questions; they involve value judgments about fairness, equity, and the potential consequences of data-driven discrimination. As recent work in AI ethics emphasizes, fairness cannot be reduced to the mere absence of statistical bias or equal distribution of outcomes, but must be understood as a broader ethical value encompassing socio-relational dimensions and respect for individuals as persons; consequently, operationalizing "representativeness" inevitably embeds normative assumptions about what it means to treat individuals and groups fairly in practice[428].

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 makes the codification of unambiguous answers to these crucial questions within harmonised standards truly improbable.

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.

---

[425] Sofia Bonicalzi, Mario De Caro and Benedetta Giovanola, 'Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems' (2023) 42 Topoi 819 <https://doi.org/10.1007/s11245-023-09922-5>.

[426] 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.

[427] 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>.

[428] Benedetta Giovanola and Simona Tiribelli, 'Beyond Bias and Discrimination: Redefining the Ai Ethics Principle of Fairness in Healthcare Machine-Learning Algorithms' (2023) 38 AI and Society 549 <https://doi.org/10.1007/s00146-022-01455-6>.

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 Medical Device Regulation (MDR). Within this legal framework, the rigor of regulatory oversight is strictly commensurate with the potential vulnerability associated with the device, operationalised through a risk-based taxonomy that categorises products into four distinct classes: Class I, Class IIa, Class IIb, and Class III. This hierarchical classification is governed by a complex set of implementing rules that assess the device based on variables such as its degree of invasiveness, duration of body contact, and the criticality of its medical function. Crucially, this taxonomy dictates the conformity assessment route; while Class I devices generally require self-certification, higher risk classes mandate the involvement of a Notified Body to verify compliance, thereby linking the technical characteristics of the device directly to the procedural burden placed upon the manufacturer.

According to MDR[429], software intended to provide information which is used to make decisions with 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[430]. 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.

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[431]. 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.

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

[429] See Annex VIII, Section 6.3, Rule 11 of the MDR.

[430] 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.

[431] Here, 'accuracy' can be broadly understood as functional correctness and reliability in generating clinically relevant and sound recommendations.

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.

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.

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[432]. 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."

The considerations delineated above appear to constitute a formidable, if not insurmountable, impediment to the efficacy of technical standards. Specifically, the capacity of such standards to prescribe precise technical specifications seems fundamentally compromised in contexts characterised by contested axiological values and profound epistemic uncertainty. Upon closer inspection, however, the situation described above 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, but not equally intractable, to those described for AI regulatees.

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)

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

faced by AI regulatees like the provider of DoctorLLM. Both involve making a judgment call about safety and risk in the face of uncertainty.

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

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.

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)[433]. 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.

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"[434]. Under the stringent criteria for evidence-based policymaking and risk assessment typically applied in established domains such as pharmaceutical regulation or environmental risk assessment[435], such an acknowledged scarcity of robust,

[433] Douglas (n 27).

[434] '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.

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

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 (justifiably or not) 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[436]. While this proactive stance may be considered as 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.

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 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.

6. Answering value-laden questions in conformity assessment procedures

The analysis developed thus far delineates a scenario in which technical standards appear fundamentally ill-equipped to furnish the granular technical specifications necessary for the implementation of the AI Act's vague and axiologically contestable requirements. Consequently, a significant epistemic and practical lacuna remains regarding the precise modalities through which compliance with these statutory provisions may be achieved and subsequently demonstrated by regulatees. In the absence of exhaustive standards capable of translating abstract legal mandates into binary technical code, the mechanism for establishing adherence shifts from reliance on pre-defined specifications to dynamic verification processes. As explored in Chapter 4, Section 5, the architecture of conformity assessment procedures is therefore poised to assume a function of paramount importance. Consequently, a crucial, if not definitive, role is assumed either by Notified Bodies, in cases necessitating third-party involvement, or by the providers themselves, in instances where the legislative text permits internal control. Irrespective of whether this assessment is externalised or conducted in-house, the entity responsible is expected to function not merely as an administrative auditor of clear-cut rules, but effectively as an interstitial arbiter charged with bridging the gap between open-textured legislative values and concrete technological applications. The intervention of these actors, be they independent auditors or internal compliance units, thus becomes the locus where the indeterminacy of the regulation is

[436] 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.

negotiated, requiring the exercise of a degree of interpretive discretion that transcends traditional technical verification.

Consequently, either the provider itself or the notified body will have the burden of making value judgments when they declare the compliance of a high-risk AI system with the requirements of the AI Act. For example, they will have the responsibility to decide that a given level of risk is acceptable, an accuracy level is appropriate or an oversight measure is commensurate. This effectively pushes the locus of real decision-making under the AI Act beyond the standardisation process and into the conformity assessment process, where "regulatory intermediaries"[437] will decide whether an AI system is compliant with the AI Act and its value-laden requirements. It is crucial to explore the implications and concerns arising from this shift before concluding. The delegation of responsibility from technical standards to the actors involved in conformity assessment mirrors the concerns raised about the delegation from legislative acts to technical standards. Issues of legitimacy, accountability, and transparency are not mitigated but rather replicated at a lower level. Furthermore, the challenge of regulatory effectiveness extends from technical standards to conformity assessment. Can these regulatory intermediaries, whether internal assessors or notified bodies, reliably ensure compliance with the AI Act's objectives, particularly regarding fundamental rights and European values?

This complex question lacks a simple answer, but several considerations can inform further analysis. First, the AI Act presents novel challenges for the conformity assessment system. Traditionally, certification bodies have evaluated product conformity against quantitative thresholds established by EU safety regulations and corresponding technical standards (e.g., chemical properties under the Toy Directive). Even in such scenarios with clear quantitative benchmarks, challenges exist. The case of ISO standards for analysing tar, nicotine, and carbon monoxide emissions in cigarettes is a clear example, as discussed in section 2 of the present chapter. More generally, measurement uncertainties or inadequate methods can complicate conformity assessment even when quantitative thresholds exist. The value-laden nature of the AI Act's requirements introduces an additional layer of complexity. Firstly, the regulation and technical standards may lack quantitative thresholds, or established ones might reflect value judgments reflecting the positionality of the subjects making these judgments. Secondly, certain requirements, like transparency or human oversight, are inherently nonquantifiable, leading to significant discretion during conformity assessment. This discretion can create uncertainty and inconsistencies in the application of the AI Act, as well as expose it to manipulation and abuse. Second, while often heralded as a success for regulating safety risks within the EU, this regulatory approach relying on conformity assessment might not be fully adaptable to novel risks like those posed by, e.g., artificial intelligence or cybersecurity[438]. For example, "risks to fundamental rights" fundamentally differs from a "safety risks". Safety risks are typically measurable and predictable, with regulations aiming to minimise their occurrence. Conversely, fundamental rights risks are complex, involving intangible concepts, unforeseen scenarios, and interactions between technological and social systems. Right-based regulations, in this context, strive to maximise the protection of fundamental rights. Since the two approaches, i.e., product safety and fundamental right protection, differ in substantial methodological ways[439], this suggests the potential need for distinct approaches to safeguard fundamental rights. Finally, even within the realm of safety, historical examples demonstrate potential shortcomings of the conformity assessment system. The

[437] Kenneth W Abbott, David Levi-Faur and Duncan Snidal, 'Introducing Regulatory Intermediaries' (2017) 670 The ANNALS of the American Academy of Political and Social Science 6 <https://doi.org/10.1177/0002716217695519>.

[438] Galland, 'Standards, Certification, and Accreditation' (n 294).

[439] Marco Almada and Nicolas Petit, 'The EU AI Act: Between Product Safety and Fundamental Rights' [2022] SSRN Electronic Journal <https://doi.org/10.2139/ssrn.4308072> accessed 16 March 2025.

PIP breast implant scandal serves as a stark reminder[440]. From 2001 to 2010, French company Poly Implant Prothèse manufactured breast implants with substandard silicone gel, causing harm to thousands of women globally. Despite using unauthorized materials, PIP's products received certification from a notified body, Tüv Rheinland, and the CE marking. In the conformity assessment process, the certification body relied solely on paperwork and technical documentation, such as design dossiers and manufacturing instructions. Notably, the certification body was not legally required to conduct physical tests on the implants themselves. Similarly, the AI Act does not mandate testing during the conformity assessment process (besides narrow exceptions in Annex VII, 4.4). While mandatory testing could be burdensome for notified bodies, this exposes potential vulnerabilities in the current system.

Since technical standards alone are constitutionally and functionally incapable of adjudicating these value-laden determinations, the substantive responsibility for addressing these intricate normative questions is effectively delegated to the often-opaque procedural machinery of conformity assessment. This displacement of decision-making authority inevitably precipitates significant concerns regarding legitimacy, reliability and accountability. By shifting the resolution of value-laden issues away from democratic deliberation and into the distinctively technocratic and confidential realm of compliance verification, the regulatory framework risks obfuscating the normative reasoning behind critical deployment decisions. Such a dynamic underscores the various theoretical and practical tensions that have been critically interrogated throughout the course of this chapter.

Notwithstanding these identified structural vulnerabilities, the ultimate capacity of the AI Act, when functioning as a composite regulatory assemblage alongside technical standards and the conformity assessment regime, to successfully realise its dual objectives of safety and cannot be determined *a priori* resolution, as the efficacy of this complex interplay will largely depend on implementation dynamics that have yet to unfold. While a more speculative and forward-looking reasoning regarding the potential success or failure of this framework is reserved for the conclusion of this dissertation, it is necessary at this juncture to pivot towards a more granular analysis. Accordingly, the subsequent section endeavours to delineate the role that technical standards might plausibly sustain within the specific institutional architecture of the AI Act, irrespective of their limitations in resolving high-level normative conflicts.

7. Which role for technical standards?

The 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[441].

As discussed in the previous sections, the first and main implication of this potential misalignment between the NLF's assumptions and the realities of AI regulation is the risk of undermining the regulatory effectiveness of the AI Act[442]. In this context, regulatory effectiveness can be understood as the capacity

[440] Rott (n 299).

[441] 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>.

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

of a regulatory regime to achieve its stated objectives[443], 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.

Secondly, without clear, widely accepted benchmarks derived from such standards, conformity assessments 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 (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.

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[444]. The promise of "presumption of conformity" via harmonised standards, a key pillar of the NLF[445], 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.

To address the 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 solely relying on standards to deliver substantive solutions, a complementary approach focusing on

[443] 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.

[444] 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.

[445] Philippe Portalier, 'Myths and Realities of the Presumption of Conformity' <https://orgalim.eu/insights/myths-and-reality-presumption-conformity>.

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.

Given the recent debate on the potential revision of the AI Act[446], 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[447]. 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.

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:

- 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.
- 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[448], and dedicated professionals can support the process[449].
- 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).
- 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

[446] '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.

[447] 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.

[448] 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.

[449] 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.

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.

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

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.

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.

8. Conclusion of the Chapter

This chapter has provided a critical interrogation of the adequacy of technical standards within the governance framework of the AI Act, challenging the prevailing legislative assumption that such instruments can effortlessly translate high-level legal principles into granular technical specifications. The analysis began by establishing a specific conceptualisation of "regulatory effectiveness", distinguishing it from questions of legitimacy. By examining the limitations exposed in the Stichting Rookpreventie Jeugd case, the discussion highlighted the inherent risks of relying on technical standards to operationalise legal objectives when the underlying methodologies are contested or fail to capture the complexity of the protected values. It was argued that this systemic vulnerability is acutely exacerbated in the context of AI, where the gap between legal prescription and technical realisation is widened by the unique characteristics of the technology.

Central to this critique was a metaethical analysis of the AI Act's essential requirements, which revealed that key concepts such as "acceptable risk," "appropriate accuracy," and "commensurate oversight" are characterised by a profound degree of vagueness and contestability. Unlike NLF sectors, such as toy safety or medical devices, where consensus on safety thresholds is grounded in established scientific paradigms, the field of AI currently lacks the shared epistemic foundations necessary to resolve these normative questions through technical standardisation alone. Consequently, the chapter demonstrated that compliance with the AI Act does not present itself to regulatees as a mere technical checklist, but rather as a series of complex "decision problems." Whether in the context of autonomous vehicles or medical diagnostic tools like DoctorLLM, regulatees are forced to navigate trade-offs between competing values and uncertain evidence, making judgments that are inherently political and ethical rather than purely technical.

Finally, the analysis illuminated the significant institutional implications of this indeterminacy. It was shown that the inability of harmonised standards to fully resolve these decision problems inevitably displaces the burden of normative adjudication onto the conformity assessment process. This shift risks transforming conformity assessment bodies and internal compliance teams into de facto policy-makers, operating within a framework that may lack sufficient transparency and accountability. In response to these challenges, the chapter concluded by proposing a reorientation of the regulatory strategy from a reliance on substantive, outcome-based standards to a focus on procedural rigour. The proposal for an "AI Act Compliance Case" suggests a pathway where compliance is demonstrated not through the passive application of standard thresholds, but through the active, documented justification of the design choices, value trade-offs, and risk management strategies employed. This procedural pivot acknowledges the current limitations of technical standards while offering a pragmatic mechanism to enhance regulatory effectiveness, ensuring that the operationalisation of the AI Act remains robust, transparent, and aligned with its fundamental rights objectives.

Conclusion

Artificial Intelligence stands as one of the preeminent drivers of contemporary paradigmatic change. As such, the imperative to regulate its development and deployment is a direct function of the disruptive impact it threatens to exert upon the social fabric. In the preceding chapters, this dissertation has extensively reconstructed a specific legislative attempt to govern this phenomenon: the European Union's AI Act. We have critically analysed its heavy reliance on technical standards and underscored the significant limitations inherent in such approach. It is imperative to acknowledge, however, that the European model represents merely one distinct approach among others. Other geopolitical regions are currently cultivating alternative frameworks that stand in contrast to the European one. While a comprehensive comparative assessment lies beyond the specific scope of this inquiry, and thus we shall abstain from evaluating whether these alternatives possess greater or lesser efficacy than the European strategy, their existence highlights the lack of consensus on the best way to regulate AI.

In this concluding analysis, however, it is necessary to widen the lens and interrogate the subject matter from a higher level of abstraction. A foundational premise underlying this entire thesis, as articulated in the introduction, is that leaving AI unregulated would precipitate cumulatively deleterious outcomes, whereas regulation serves as a mechanism to channel its potential toward positive societal ends. In other words, the investigation has been predicated on the assumption that regulatory intervention leads to a trajectory distinct from, and superior to, the one that would materialise in its absence. Yet, one must pause to ask: is this assumption entirely sound?

This query evokes a profound tension regarding the relationship between technological advancement and social change, and specifically, our capacity to govern the effects of the former on the latter. While regulatory interventions rest on the assumption that such effects are controllable, a diametrically opposed viewpoint suggests that technology is the governor rather than the governed. From this vantage point, laws, technical standards, and other normative interventions may be structurally incapable of dictating the aggregate manner in which technology impacts the human condition.

Within the domain of the philosophy of technology, this perspective is codified as technological determinism[450]. At its core, this theoretical framework posits that technical forces serve as the primary movers of history and social structure, often operating independently of human will or legislative intent. Crucially, this view challenges the instrumental definition of technology, which regards tools as neutral vessels that can be used for good or evil depending on the user's intent. Instead, technological determinism argues that technologies are inherently value-laden; they impose specific forms of organisation, perception, and interaction upon the societies that adopt them. By structuring the material conditions of existence, a transformative technology dictates the values and hierarchies of the future, rendering political attempts to curb its influence reactive and often futile. The medium itself determines the nature of the social bond, irrespective of the regulatory constraints attempting to encase it.

If one accepts the validity of this deterministic perspective, though a comprehensive defense of it lies beyond the scope of this discussion, the implications for the EU's legislative efforts are severe. Beyond the fundamental structural deficiencies already highlighted in Chapter 5, which arguably render the AI

[450] Merritt Roe Smith and Leo Marx, *Does Technology Drive History?: The Dilemma of Technological Determinism* (MIT Press 1994).

Act just another chapter of "how to make bad law"[451], we are compelled to drastically recalibrate our expectations regarding the role of regulation in steering technological developments and their societal aftershocks. To be clear, this concession does not imply that technology and its evolution possess a specific autonomy or exist in a vacuum. The genesis and uptake of any transformative technology, including AI, is undeniably influenced by a complex interplay of social, cultural, and economic factors. However, the deterministic argument underscores that once set in motion, such technology tends to unfold a specific set of effects and potentialities intrinsic to its logic.

Consequently, one must inquire into the specific trajectories and potentialities that artificial intelligence is poised to actualise. Much like the transformative technologies that preceded it, and viewing the phenomenon from our current vantage point where these consequences are not yet fully unfolded, we can observe that they will likely oscillate between the antipodal poles of subjugation and liberation. On the one hand, AI promises to function as a formidable apparatus for reinforcing the dominion of man over man. Ranging from ubiquitous mass surveillance to the algorithmic manipulation of public discourse and the automated enforcement of social stratification, AI stands to become a primary instrument for malevolent actors and entrenched power structures to consolidate their hegemony. There is little doubt that, in this regard, the technology harbors profound negative potentialities when scrutinized through the lens of power dynamics and individual sovereignty.

At the same time, however, the liberating capacity of this technology cannot be dismissed. Artificial intelligence holds the potential to act as a technology of freedom by emancipating humanity from the burdens of cognitive drudgery and physical hazard. By automating routine and dangerous tasks, AI offers the possibility of redirecting human labor toward more creative and fulfilling endeavors. Furthermore, the democratisation of access to sophisticated analytical tools and vast repositories of knowledge, facilitated by intelligent systems, could dismantle traditional barriers to entry in fields such as medicine, law, and engineering, thereby enhancing individual agency and fostering a more equitable distribution of cognitive resources. This potential for cognitive augmentation suggests a future where human capability is not replaced, but rather exponentially expanded.

Notwithstanding the pervasive fervor and hype currently surrounding these developments, we stand merely at the threshold of the AI transformation. Its profound effects remain largely latent, waiting to be fully actualised. Whether the positive or negative manifestations will eventually prevail is not a foregone conclusion, but rather a question that belongs to the indefinite dialectic of the history of humans and their technical artifacts.

[451] Chris Reed, 'How to Make Bad Law: Lessons from Cyberspace' (2010) 73 The Modern Law Review 903.

References

Abbot C, 'Bridging the Gap – Non-State Actors and the Challenges of Regulating New Technology' (2012) 39 Journal of Law and Society 329 <https://doi.org/10.1111/j.1467-6478.2012.00588.x>

Abbott KW, Levi-Faur D and Snidal D, 'Introducing Regulatory Intermediaries' (2017) 670 The ANNALS of the American Academy of Political and Social Science 6 <https://doi.org/10.1177/0002716217695519>

Acemoglu D and Restrepo P, 'Artificial Intelligence, Automation, and Work' *The economics of artificial intelligence: An agenda* (University of Chicago Press 2018)

'Agreement on Technical Co-Operation between ISO and CEN (Vienna Agreement)' <https://boss.cen.eu/media/CEN/ref/vienna_agreement.pdf>

Ahern DM, 'Regulators Nurturing FinTech Innovation: Global Evolution of the Regulatory Sandbox as Opportunity Based Regulation' (Social Science Research Network, 9 March 2020) <https://doi.org/10.2139/ssrn.3552015> accessed 23 December 2025

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

'AI Recruitment Statistics: What Is the Future of Hiring?' (*Tidio*, 5 March 2025) <https://www.tidio.com/blog/ai-recruitment/> accessed 5 July 2025

Albaroudi E, Mansouri T and Alameer A, 'A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring' (2024) 5 AI 383 <https://doi.org/10.3390/ai5010019>

Algorithm Watch, 'Automating Society Report 2020' (*Automating Society Report 2020*, 2020) <https://automatingsociety.algorithmwatch.org> accessed 3 October 2023

Almada M and Petit N, 'The EU AI Act: Between Product Safety and Fundamental Rights' [2022] SSRN Electronic Journal <https://doi.org/10.2139/ssrn.4308072> accessed 16 March 2025

Ambec S and others, 'The Porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness?' (2013) 7 Review of Environmental Economics and Policy 2 <https://doi.org/10.1093/reep/res016>

Appleton AE, 'Conformity Assessment Procedures' *Research Handbook on the WTO and Technical Barriers to Trade* (Edward Elgar Publishing 2013) <https://www.elgaronline.com/edcollchap/edcoll/9780857936714/9780857936714.00008.xml> accessed 30 December 2025

'Artificial Intelligence - The Consequences of Artificial Intelligence on the (Digital) Single Market, Production, Consumption, Employment and Society (Own-Initiative Opinion) | EESC' (11 October 2016) <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and> accessed 19 August 2025

Awad E and others, 'The Moral Machine Experiment' (2018) 563 Nature 59 <https://doi.org/10.1038/s41586-018-0637-6>

Ballor G, 'The Origins of CE Marking: Standards, Business, and the European Market in the 1980s–1990s' Harvard Business School - Working Paper 21-142 <https://www.hbs.edu/faculty/Pages/item.aspx?num=60460> accessed 20 June 2024

Balzarova MA and Castka P, 'Stakeholders' Influence and Contribution to Social Standards Development: The Case of Multiple Stakeholder Approach to ISO 26000 Development' (2012) 111 Journal of Business Ethics 265 <https://doi.org/10.1007/s10551-012-1206-9>

Besen SM and Farrell J, 'Choosing How to Compete: Strategies and Tactics in Standardization' (1994) 8 Journal of Economic Perspectives 117 <https://doi.org/10.1257/jep.8.2.117>

Bianchini S, Müller M and Pelletier P, 'Artificial Intelligence in Science: An Emerging General Method of Invention' (2022) 51 Research Policy 104604 <https://doi.org/10.1016/j.respol.2022.104604>

Biddle B, White A and Woods S, 'How Many Standards in a Laptop? (And Other Empirical Questions)' (Social Science Research Network, 10 September 2010) <https://doi.org/10.2139/ssrn.1619440> accessed 12 August 2025

Bijlmakers S, 'No ISO Fix for Human Rights: A Critical Perspective on ISO 26000 Guidance on Social Responsibility' *Research Handbook on Global Governance, Business and Human Rights* (Edward Elgar Publishing 2022) <https://www.elgaronline.com/edcollchap/edcoll/9781788979825/9781788979825.00017.xml> accessed 23 May 2024

Binns R and others, 'Like Trainer, like Bot? Inheritance of Bias in Algorithmic Content Moderation' (2017) 10540 LNCS Social Informatics (SocInfo 2017) <https://ora.ox.ac.uk/objects/uuid:b0723053-f564-4a49-be8f-85e1b0780ea7> accessed 5 July 2025

Bisconti Lucidi P and Nardi D, 'Companion Robots: The Hallucinatory Danger of Human-Robot Interactions' *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (Association for Computing Machinery 2018) <https://doi.org/10.1145/3278721.3278741> accessed 20 April 2026

Bisconti P, 'Will Sexual Robots Modify Human Relationships? A Psychological Approach to Reframe the Symbolic Argument' (2021) 35 Advanced Robotics 561 <https://doi.org/10.1080/01691864.2021.1886167>

——, 'A Formal Account of Trustworthiness: Connecting Intrinsic and Perceived Trustworthiness' *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (2024) <https://ojs.aaai.org/index.php/AIES/article/view/31624> accessed 20 April 2026

——, 'Adversarial Poetry as a Universal Single-Turn Jailbreak Mechanism in Large Language Models' (arXiv, 16 January 2026) <https://doi.org/10.48550/arXiv.2511.15304> accessed 20 April 2026

Bisconti P and Galisai M, 'Standards for Trustworthy AI in the European Union: Technical Rationale, Structural Challenges, and an Implementation Path' (arXiv, 21 January 2026) <https://doi.org/10.48550/arXiv.2602.00078> accessed 20 April 2026

Black J, 'Decentring Regulation: Understanding the Role of Regulation and Self-Regulation in a "Post-Regulatory" World' (2001) 54 Current Legal Problems 103 <https://doi.org/10.1093/clp/54.1.103>

——, 'Critical Reflection on Regulation' <https://eprints.lse.ac.uk/35985/>

Blind K, 'A Taxonomy of Standards in the Service Sector: Theoretical Discussion and Empirical Test' [2006] The Service Industries Journal <https://doi.org/10.1080/02642060600621597> accessed 12 August 2025

——, 'Standardisation as a Catalyst for Innovation' [2009] The Review of Economic Studies 1

——, 'The Overall Impact of Economic, Social and Institutional Regulation on Innovation: An Update' *Handbook of Innovation and Regulation* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781800884472/b-9781800884472.00020.xml> accessed 3 July 2025

Blind K, Petersen SS and Riillo CAF, 'The Impact of Standards and Regulation on Innovation in Uncertain Markets' (2017) 46 Research Policy 249 <https://doi.org/10.1016/j.respol.2016.11.003>

'Bluetooth Technology Website | The Official Website of Bluetooth Technology.' (*Bluetooth® Technology Website*, 19 December 2025) <https://www.bluetooth.com/> accessed 28 December 2025

Bommasani R and others, 'A Path for Science- and Evidence-Based AI Policy' [2024] https://understanding-ai-safety.org/

Bonicalzi S, Caro MD and Giovanola B, 'Artificial Intelligence and Autonomy: On the Ethical Dimension of Recommender Systems' (2023) 42 Topoi 819 <https://doi.org/10.1007/s11245-023-09922-5>

Borz G and others, 'The EU Soft Regulation of Digital Campaigning: Regulatory Effectiveness through Platform Compliance to the Code of Practice on Disinformation' (2024) 45 Policy Studies 709 <https://doi.org/10.1080/01442872.2024.2302448>

Braunerhjelm P and others (eds), *Handbook of Innovation and Regulation* (Edward Elgar Publishing 2023) <https://doi.org/10.4337/9781800884472> accessed 3 July 2025

Brem A and Nylund P, 'The Inertia of Dominant Designs in Technological Innovation: An Ecosystem View of Standardization' (2024) 71 IEEE Transactions on Engineering Management 2640 <https://doi.org/10.1109/TEM.2022.3192094>

Brownsword R, 'So What Does the World Need Now? Reflections on Regulating Technologies' in Roger Brownsword and Karen Yeung (eds), *Regulating technologies* (Hart 2008)

Brownsword R and Brownsword R, *Rights, Regulation, and the Technological Revolution* (Oxford University Press 2008)

Brownsword R and Goodwin M, 'Regulatory Effectiveness II: Failure by Regulators' *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://doi.org/10.1017/CBO9781139047609.017> accessed 26 October 2025

——, 'Regulatory Effectiveness IV: Third-Party Interference and Disruptive Externalities' *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://doi.org/10.1017/CBO9781139047609.019> accessed 26 October 2025

Buocz T, Pfotenhauer S and Eisenberger I, 'Regulatory Sandboxes in the AI Act: Reconciling Innovation and Safety?' (2023) 15 Law, Innovation and Technology 357 <https://doi.org/10.1080/17579961.2023.2245678>

Burden H and Stenberg S, *Fit for Purpose – Data Quality for Artificial Intelligence* (RISE Research Institutes of Sweden 2024) <https://urn.kb.se/resolve?urn=urn:nbn:se:ri:diva-74929> accessed 19 August 2025

Burns T, Cosgrove J and Doyle F, 'A Review of Interoperability Standards for Industry 4.0.' (2019) 38 Procedia Manufacturing 646 <https://doi.org/10.1016/j.promfg.2020.01.083>

Busch L, *Standards: Recipes for Reality* (The MIT Press 2011) <https://doi.org/10.7551/mitpress/8962.001.0001> accessed 16 November 2023

——, 'Standards and Their Problems: From Technical Specifications to World-Making' in Mara Miele and others (eds), *Research in Rural Sociology and Development*, vol 24 (Emerald Publishing Limited 2017) <https://doi.org/10.1108/S1057-192220170000024005> accessed 8 February 2024

Butenko A and Larouche P, 'Regulation for Innovativeness or Regulation of Innovation?' (2015) 7 Law, Innovation and Technology 52 <https://doi.org/10.1080/17579961.2015.1052643>

Cabral TS, 'Rethinking the List-Based Approach to High-Risk Systems under the AI Act' (Social Science Research Network, 3 April 2025) <https://doi.org/10.2139/ssrn.5206860> accessed 19 August 2025

Canale D, 'The Opacity of Law: On the Hidden Impact of Experts' Opinion on Legal Decision-Making' (2021) 40 Law and Philosophy 1 <https://doi.org/10.1007/s10982-021-09408-8>

——, 'Quando gli esperti creano diritto: deferenza, opacità, legittimità' (2022) 22 Analisi e Diritto <https://doi.org/10.4454/analisiediritto.v22i2.508> accessed 9 April 2024

Cantero Gamito M, 'The Role of ETSI in the EU's Regulation and Governance of Artificial Intelligence' (2024) 37 Innovation: The European Journal of Social Science Research 1425 <https://doi.org/10.1080/13511610.2024.2349627>

Castán CT, 'The Legal Concept of Artificial Intelligence: The Debate Surrounding the Definition of AI System in the AI Act' [2024] BioLaw Journal - Rivista di BioDiritto 305 <https://doi.org/10.15168/2284-4503-3000>

Castka P and Corbett CJ, 'Management Systems Standards: Diffusion, Impact and Governance of ISO 9000, ISO 14000, and Other Management Standards' (2015) 7 Foundations and Trends® in Technology, Information and Operations Management 161 <https://doi.org/10.1561/0200000042>

Castro D and McLaughlin M, 'Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence | ITIF' [2019] Information Technology and Innovation Foundation (ITIF) <https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence/> accessed 3 August 2023

'CENELEC Guide 13. IEC-CENELEC Agreement on Common Planning of New Work and Parallel Voting (the Frankfurt Agreement)' <https://boss.cenelec.eu/media/Guides/CLC/13_cenelecguide13.pdf>

'CEO of Anthropic Warns against AI Deregulation' (6 June 2025) <https://www.semafor.com/article/06/06/2025/the-ceo-of-ai-company-anthropic-warns-against-deregulation> accessed 3 July 2025

Chen RJ and others, 'Algorithmic Fairness in Artificial Intelligence for Medicine and Healthcare' (2023) 7 Nature Biomedical Engineering 719 <https://doi.org/10.1038/s41551-023-01056-8>

Chen W and others, 'Can the Establishment of a Personal Data Protection System Promote Corporate Innovation?' (2024) 53 Research Policy 105080 <https://doi.org/10.1016/j.respol.2024.105080>

Cocchiaro MZ 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

Collingridge D, *The Social Control of Technology* (London : Frances Pinter 1980) <http://archive.org/details/socialcontroloft0000coll> accessed 5 July 2025

Colombo C and Eliantonio M, 'Harmonized Technical Standards as Part of EU Law: Juridification with a Number of Unresolved Legitimacy Concerns?: Case C-613/14 James Elliot Construction Limited v. Irish Asphalt Limited, EU:C:2016:821' (2017) 24 Maastricht Journal of European and Comparative Law 323 <https://doi.org/10.1177/1023263X17709753>

'Commission Guidelines on AI System Definition' <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application> accessed 19 August 2025

'Communication: Building Trust in Human Centric Artificial Intelligence | Shaping Europe's Digital Future' <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence> accessed 19 August 2025

Corbett CJ, 'Global Diffusion of ISO 9000 Certification Through Supply Chains' in Christopher S Tang, Chung-Piaw Teo and Kwok-Kee Wei (eds), *Supply Chain Analysis: A Handbook on the Interaction of Information, System and Optimization* (Springer US 2008) <https://doi.org/10.1007/978-0-387-75240-2_7> accessed 11 August 2025

Cordeiro F and Ramsey MH, 'Sampling in the Context of Conformity Assessment' (2023) 28 Accreditation and Quality Assurance 181 <https://doi.org/10.1007/s00769-023-01542-1>

Corrêa NK and others, 'Worldwide AI Ethics: A Review of 200 Guidelines and Recommendations for AI Governance' (2023) 4 Patterns 100857 <https://doi.org/10.1016/j.patter.2023.100857>

Council NR, Policy B on S Technology, and Economic and Committee IS Conformity Assessment, and US Trade Policy Project, *Standards, Conformity Assessment, and Trade: Into the 21st Century* (National Academies Press 1995)

Crane RJ, *The Politics of International Standards: France and the Color TV War* (Bloomsbury Academic 1979)

Cuccuru P, 'Regulating by Request: On the Role and Status of the Standardisation Mandate under the New Approach' in Mariolina Eliantonio and Caroline Cauffman (eds), *The Legitimacy of Standardisation as a Regulatory Technique* (Edward Elgar Publishing 2020) <https://doi.org/10.4337/9781789902952.00008> accessed 5 June 2024

Cusumano MA, Mylonadis Y and Rosenbloom RS, 'Strategic Maneuvering and Mass-Market Dynamics: The Triumph of VHS over Beta' (1992) 66 Business History Review 51

De Oliveira EC and Lourenço FR, 'Risk of False Conformity Assessment Applied to Automotive Fuel Analysis: A Multiparameter Approach' (2021) 263 Chemosphere 128265 <https://doi.org/10.1016/j.chemosphere.2020.128265>

de Vries S, Kanevskaia O and de Jager R, 'Internal Market 3.0: The Old "New Approach" for Harmonising AI Regulation' (2023) 2023 8 European Papers - A Journal on Law and Integration 583610 <https://doi.org/10.15166/2499-8249/677>

Delaney H, Delaney H and Van De Zande R, 'A Guide to EU Standards and Conformity Assessment' (0 edn, National Institute of Standards and Technology 2000) NIST SP 951 <https://doi.org/10.6028/NIST.SP.951> accessed 21 February 2025

Delimatsis P, 'Global Standard-Setting 2.0: How the WTO Spotlights ISO and Impacts The Transnational Standard-Setting Process' [2018] SSRN Electronic Journal <https://doi.org/10.2139/ssrn.3184824> accessed 4 April 2024

Demortain D, 'Expertise, Regulatory Science and the Evaluation of Technology and Risk: Introduction to the Special Issue' (2017) 55 Minerva 139 <https://doi.org/10.1007/s11024-017-9325-1>

'Directive - 83/189 - EN - EUR-Lex' <https://eur-lex.europa.eu/eli/dir/1983/189/oj/eng> accessed 14 August 2025

Donaire JAC, 'Notifying Authorities and Notified Bodies (Chapter III, Section 4). Standards, Conformity Assessment, Certificates, Registration (Chapter III, Section 5)' in Alejandro Huergo Lora and Gustavo Manuel Díaz González (eds), *The EU regulation on artificial intelligence: a commentary*

Douglas M, 'The Depoliticization of Risk' *Culture Matters* (Routledge 1997)

'DVD Forum's Mission' <https://web.archive.org/web/20241128004023/http://www.dvdforum.org/about-mission.htm> accessed 28 December 2025

Dwivedi APD, 'Laboratory Accreditation: Evolution and Development' in Anuj Bhatnagar and others (eds), *Handbook of Quality System, Accreditation and Conformity Assessment* (Springer Nature 2025) <https://doi.org/10.1007/978-981-99-4637-2_26-1> accessed 20 November 2024

Easterbrook F, 'Cyberspace and the Law of the Horse' (1996) 1996 University of Chicago Legal Forum <https://chicagounbound.uchicago.edu/uclf/vol1996/iss1/7>

Egyedi T, 'The Standardized Container: Gateway Technologies in Cargo Transportation' in MJ Holler and E Niskanen (eds), *EURAS Yearbook of Standardization* (EURAS 2000)

Eliantonio M, 'Judicial Control of the EU Harmonized Standards: Entering a Black Hole?' (2017) 44 Legal Issues of Economic Integration <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\LEIE\LEIE2017022.pdf>

Eliantonio M and Cauffman C, *The Legitimacy of Standardisation as a Regulatory Technique: A Cross-Disciplinary and Multi-Level Analysis* (Edward Elgar Publishing 2020) <https://www.elgaronline.com/display/edcoll/9781789902945/9781789902945.xml>

Endicott TAO, 'Vagueness and Legal Theory' (1997) 3 Legal Theory 37 <https://doi.org/10.1017/S135232520000063X>

Enqvist L, '"Human Oversight" in the EU Artificial Intelligence Act: What, When and by Whom?' [2023] Law, Innovation and Technology <https://www.tandfonline.com/doi/abs/10.1080/17579961.2023.2245683> accessed 21 August 2025

Ernst D, Lee H and Kwak J, 'Standards, Innovation, and Latecomer Economic Development: Conceptual Issues and Policy Challenges' (2014) 38 Telecommunications Policy 853 <https://doi.org/10.1016/j.telpol.2014.09.009>

ETSI, 'Study into the Challenges of Developing Harmonised Standards in the Context of Future Changes to the Environment in Which Products Are Being Developed and Operated'

'EUROPA – European Commission – Growth – Regulatory Policy - SMCS' <https://webgate.ec.europa.eu/single-market-compliance-space/notified-bodies> accessed 3 January 2026

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.' <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

——, 'An EU Strategy on Standardisation - Setting Global Standards in Support of a Resilient, Green and Digital EU Single Market' <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0031>

European Commission. Joint Research Centre. and Organisation for Economic Co operation and Development., *AI Watch, National Strategies on Artificial Intelligence: A European Perspective.* (Publications Office 2021) <https://data.europa.eu/doi/10.2760/069178> accessed 23 June 2025

'European Standardisation – Evaluation' <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13446-European-standardisation-evaluation_en> accessed 16 August 2025

Fahey DM and Gliozzo CA, 'The Metric System and the French Revolution' (1969) 68 South Atlantic Quarterly 74 <https://doi.org/10.1215/00382876-68-1-74>

Falke J and Joerges C, 'Traditional Law Approximation Policy Approaches to Removing Technical Barriers to Trade and Efforts at a Horizontal European Product Safety Policy' (2010) 6 Hanse Law Review 237

Farrell J, 'Economic Issues in Standardization'

Farrell J and Saloner G, 'Economic Issues in Standardization' (Massachusetts Institute of Technology (MIT), Department of Economics 1985) Working paper 393 <https://EconPapers.repec.org/RePEc:mit:worpap:393> accessed 21 April 2026

——, 'Coordination Through Committees and Markets' (1988) 19 The RAND Journal of Economics 235 <https://doi.org/10.2307/2555702>

Fatima S, Desouza KC and Dawson GS, 'National Strategic Artificial Intelligence Plans: A Multi-Dimensional Analysis' (2020) 67 Economic Analysis and Policy 178 <https://doi.org/10.1016/j.eap.2020.07.008>

Faure M and Philipsen N, 'Standardisation from a Law and Economics Perspective' in Mariolina Eliantonio and Caroline Cauffman (eds), *The Legitimacy of Standardisation as a Regulatory Technique* (Edward Elgar Publishing 2020) <https://doi.org/10.4337/9781789902952.00013> accessed 4 April 2024

Fennell A, 'AI in Recruitment Statistics UK 2025 | Latest Reports & Data' (12 June 2024) <https://standout-cv.com/stats/ai-in-recruitment-statistics-uk> accessed 5 July 2025

Fenwick M, Kaal W and Vermeulen E, 'Regulation Tomorrow: What Happens When Technology Is Faster than the Law?' (2017) 6 American University Business Law Review <https://digitalcommons.wcl.american.edu/aublr/vol6/iss3/1>

Ferraris M, *Documentality: Why It Is Necessary to Leave Traces* (Fordham University Press 2013)

Filippo Bagni and Seferi Fabio (eds), *White Paper on Regulatory Sandboxes for AI and Cybersecurity* (Cybersecurity National Lab 2025)

Finck M, 'Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy' (Social Science Research Network, 20 June 2017) <https://doi.org/10.2139/ssrn.2990043> accessed 30 June 2025

——, 'In Search of the Lost Research Exemption: Reflections on the AI Act' (2025) 74 GRUR International 903 <https://doi.org/10.1093/grurint/ikaf100>

Floridi L, 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical' (2019) 32 Philosophy & Technology 185 <https://doi.org/10.1007/s13347-019-00354-x>

——, 'The End of an Era: From Self-Regulation to Hard Law for the Digital Industry' (2021) 34 Philosophy & Technology 619 <https://doi.org/10.1007/s13347-021-00493-0>

——, 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28 Minds and Machines 689 <https://doi.org/10.1007/s11023-018-9482-5>

Fortes PRB, Baquero PM and Amariles DR, 'Artificial Intelligence Risks and Algorithmic Regulation' (2022) 13 European Journal of Risk Regulation 357 <https://doi.org/10.1017/err.2022.14>

Frankel C and Galland J-P, 'Market, Regulation, Market, Regulation' (Wissenschaftsverlag Mainz Aachen, Germany 2015) <https://enpc.hal.science/hal-01275604> accessed 24 May 2024

Fraser H and Villarino J-MB y, '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 <https://doi.org/10.1017/err.2023.57>

Freiberg A, 'Re-Stocking the Regulatory Tool-Kit' <http://regulation.huji.ac.il/dp.php>

——, *The Tools of Regulation* (Federation Press 2010) <https://research.monash.edu/en/publications/the-tools-of-regulation> accessed 10 April 2024

'Front Page | USB-IF' <https://www.usb.org/> accessed 28 December 2025

Galland J-P, 'The Difficulties of Regulating Markets and Risks in Europe through Notified Bodies' (2013) 4 European Journal of Risk Regulation 365 <https://doi.org/10.1017/S1867299X00002634>

——, 'Big Third-Party Certifiers and the Construction of Transnational Regulation' (2017) 670 ANNALS of the American Academy of Political and Social Science 263

——, 'Standards, Certification, and Accreditation: Indispensable Tools for European Safety Regulations?' in Jean-Christophe Le Coze and Benoît Journé (eds), *The Regulator–Regulatee Relationship in High-Hazard Industry Sectors: New Actors and New Viewpoints in a Conservative Landscape* (Springer Nature Switzerland 2024) <https://doi.org/10.1007/978-3-031-49570-0_8> accessed 30 April 2024

Galvagna C, 'Discussion Paper: Inclusive AI Governance' <https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/>

Gamito MC, 'The Role of ETSI in the EU's Regulation and Governance of Artificial Intelligence' (2024) 0 Innovation: The European Journal of Social Science Research 1 <https://doi.org/10.1080/13511610.2024.2349627>

Gervais D, 'The Regulation of Inchoate Technologies' (2010) 47 Houston Law Review 665

Gibson CS, 'Globalization and the Technology Standards Game: Balancing Concerns of Protectionism and Intellectual Property in International Standards' (2007) 22 Berkeley Technology Law Journal 1403

Gikay AA and others, 'High-Risk Artificial Intelligence Systems under the European Union's Artificial Intelligence Act: Systemic Flaws and Practical Challenges' (Social Science Research Network, 2 November 2023) <https://doi.org/10.2139/ssrn.4621605> accessed 4 January 2026

Giovanola B and Tiribelli S, 'Beyond Bias and Discrimination: Redefining the Ai Ethics Principle of Fairness in Healthcare Machine-Learning Algorithms' (2023) 38 AI and Society 549 <https://doi.org/10.1007/s00146-022-01455-6>

Goodwin M and Brownsword R (eds), 'Regulatory Effectiveness I' *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://doi.org/10.1017/CBO9781139047609.016> accessed 30 May 2025

—— (eds), 'Regulatory Effectiveness III: Resistance by Regulatees' *Law and the Technologies of the Twenty-First Century: Text and Materials* (Cambridge University Press 2012) <https://doi.org/10.1017/CBO9781139047609.018> accessed 30 May 2025

Gorwa R, Binns R and Katzenbach C, 'Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance' (2020) 7 Big Data & Society 2053951719897945 <https://doi.org/10.1177/2053951719897945>

Grady P, 'The AI Act Should Be Technology-Neutral' (2023) <https://itif.org/publications/2023/02/01/the-ai-act-should-be-technology-neutral/> accessed 4 January 2026

Graz J-C, 'Global Corporations and the Governance of Standards' in Andreas Nölke and Christian May (eds), *Handbook of the International Political Economy of the Corporation* (Edward Elgar Publishing 2018) <https://doi.org/10.4337/9781785362538.00037> accessed 8 February 2024

Grindley P and Grindley P, *Standards, Strategy, and Policy: Cases and Stories* (Oxford University Press 1995)

Grossman JH, 'Standardization (Standardisation)' (2018) 44 Critical Inquiry 447 <https://doi.org/10.1086/696912>

Hagendorff T, 'The Ethics of AI Ethics: An Evaluation of Guidelines' (2020) 30 Minds and Machines 99 <https://doi.org/10.1007/s11023-020-09517-8>

——, 'Blind Spots in AI Ethics' [2021] AI and Ethics <https://doi.org/10.1007/s43681-021-00122-8> accessed 7 June 2022

Halevy A, Norvig P and Pereira F, 'The Unreasonable Effectiveness of Data' (2009) 24 IEEE Intelligent Systems 8 <https://doi.org/10.1109/MIS.2009.36>

Hassan KT and Atwan RS, 'The Civil Liability of Artificial Intelligence Applications: Between the Limitations of Traditional Liability and the Evolution of the Product Concept' (2025) 15 Indian Journal of Information Sources and Services 276 <https://doi.org/10.51983/ijiss-2025.IJISS.15.3.31>

Héder M, 'A Criticism of AI Ethics Guidelines' (2020) 20 Információs Társadalom: Társadalomtudományi Folyóirat 57

——, 'AI and the Resurrection of Technological Determinism' (2021) 21 Információs Társadalom: Társadalomtudományi Folyóirat 119

Hempel CG, 'Vagueness and Logic' (1939) 6 Philosophy of Science 163

Hendrickx V and Ooms W, 'Commentary to Article 48 of the EU AI Act. CE Marking' (Social Science Research Network, 16 April 2024) <https://doi.org/10.2139/ssrn.4890939> accessed 21 February 2025

Hesser W, 'The Need for Interdisciplinary Research on Standardization' [1997] Presentation for the SCANCOR/SCORE Seminar on Standardization September 18-20 1997 Lund, Sweden

——, *Standardisation in Companies and Markets* (Helmut-Schmidt-Univ 2010)

'Historic Documents | EU Artificial Intelligence Act' <https://artificialintelligenceact.eu/documents/> accessed 19 August 2025

'History of PDF', *Wikipedia* (2025) <https://en.wikipedia.org/w/index.php?title=History_of_PDF&oldid=1316915536> accessed 28 December 2025

Ho-Dac M, Pellegrini C and Long B, 'The Academic Guide to AI Act Compliance' (Social Science Research Network, 30 September 2025) <https://doi.org/10.2139/ssrn.5775382> accessed 23 December 2025

Holzinger A, Zatloukal K and Müller H, 'Is Human Oversight to AI Systems Still Possible?' (2025) 85 New Biotechnology 59 <https://doi.org/10.1016/j.nbt.2024.12.003>

'Home' (*JTC 1*) <https://jtc1info.org/> accessed 28 December 2025

'How Many Companies Use AI In Hiring & Recruiting? (2025)' (7 February 2025) <https://joingenius.com/statistics/how-many-companies-use-ai-in-hiring/> accessed 5 July 2025

Huang C and others, 'An Overview of Artificial Intelligence Ethics' (2023) 4 IEEE Transactions on Artificial Intelligence 799 <https://doi.org/10.1109/TAI.2022.3194503>

Hwang T, 'Computational Power and the Social Impact of Artificial Intelligence' (Social Science Research Network, 23 March 2018) <https://doi.org/10.2139/ssrn.3147971> accessed 23 June 2025

IEC, 'Inclusion and Diversity' <https://iec.ch/basecamp/inclusion-and-diversity> accessed 8 February 2024

'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

Institute AN, 'Artificial Power: 2025 Landscape Report' (*AI Now Institute*, 3 June 2025) <https://ainowinstitute.org/publications/research/ai-now-2025-landscape-report> accessed 3 July 2025

'International Organization for Standardization (ISO)' (*Oxford Public International Law*) <https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e1887?d=%2F10.1093%2Flaw%3Aepil%2F9780199231690%2Flaw-9780199231690-e1887&p=emailA%2Fb.zRQKUG9aU> accessed 28 December 2025

ISO, 'ISO Strategy 2030' <https://www.iso.org/publication/PUB100364.html>

——, 'Conformity Assessment - Basic Concepts' (*ISO*) <https://casco.iso.org/cms/render/live/en/sites/cascoregulators/home/conformity-assessment/basic-concepts.html> accessed 5 August 2024

——, 'Guidance on Use of Artificial Intelligence (AI) for ISO Committees' <https://www.iso.org/files/live/sites/isoorg/files/developing_standards/who_develops_standards/docs/use%20of%20AI.pdf> accessed 28 December 2025

ISO, IEC and UNIDO, 'Building Trust. The Conformity Assessment Toolbox' <https://www.iso.org/publication/PUB100230.html>

'ISO - A Committee to Support Developing Countries: DEVCO' (*ISO*) <https://www.iso.org/devco.html> accessed 28 December 2025

'ISO - Conformity Assessment' <https://www.iso.org/conformity-assessment.html> accessed 30 December 2025

'ISO - ISO/IEC 17025 — Testing and Calibration Laboratories' <https://www.iso.org/ISO-IEC-17025-testing-and-calibration-laboratories.html> accessed 30 December 2025

'ISO/CASCO - Committee on Conformity Assessment' <https://www.iso.org/committee/54998/x/catalogue/> accessed 30 December 2025

'ISO/IEC 17000:2020(En), Conformity Assessment — Vocabulary and General Principles' <https://www.iso.org/obp/ui/#iso:std:iso-iec:17000:ed-2:v2:en> accessed 30 December 2025

'ISO/IEC 17011:2017 - Conformity Assessment — Requirements for Accreditation Bodies Accrediting Conformity Assessment Bodies' <https://www.iso.org/standard/67198.html> accessed 30 December 2025

'ISO/IEC 17020:2012' (*ISO*) <https://www.iso.org/standard/52994.html> accessed 30 December 2025

'ISO/IEC 17029:2019' (*ISO*) <https://www.iso.org/standard/29352.html> accessed 30 December 2025

'ISO/IEC 17067:2013' (*ISO*) <https://www.iso.org/standard/55087.html> accessed 30 December 2025

'ISO/IEC Directives Part 2' <https://www.iec.ch/standards-development/isoiec-directives-part-2> accessed 12 August 2025

'ISO/IEC Guide 2:2004 - Standardization and Related Activities — General Vocabulary' <https://www.iso.org/standard/39976.html> accessed 12 August 2025

'ISO/IEC Guide 59:2019 - ISO and IEC Recommended Practices for Standardization by National Bodies' <https://www.iso.org/standard/71917.html> accessed 28 December 2025

'ISO/IEC Guide 98-4:2012' (*ISO*) <https://www.iso.org/standard/50465.html> accessed 4 August 2025

'Iso.Org/Sites/Directives/Current/Consolidated/Index.Html' <https://www.iso.org/sites/directives/current/consolidated/index.html> accessed 28 December 2025

Jakobs K, 'Responsibility by Design?! – On the Standardisation of "Smart" Systems' *Smart Technologies and Fundamental Rights* (Brill 2020) <https://doi.org/10.1163/9789004437876_014> accessed 22 January 2024

——, 'Today's Situation and Some Future Problems' (2023) 2 Standardization - Journal of Research and Innovation 16 <https://doi.org/10.58831/2831-7920-2.4.16>

'JCGM106-2012' (*BIPM*) <https://www.bipm.org/en/doi/10.59161/jcgm106-2012> accessed 30 December 2025

Jiang F and others, 'Artificial Intelligence in Healthcare: Past, Present and Future' (2017) 2 Stroke and Vascular Neurology <https://doi.org/10.1136/svn-2017-000101> accessed 23 June 2025

Jobin A, Ienca M and Vayena E, 'The Global Landscape of AI Ethics Guidelines' (2019) 1 Nature Machine Intelligence 389 <https://doi.org/10.1038/s42256-019-0088-2>

Johnson WG, 'Caught in Quicksand? Compliance and Legitimacy Challenges in Using Regulatory Sandboxes to Manage Emerging Technologies' (2023) 17 Regulation & Governance 709 <https://doi.org/10.1111/rego.12487>

Jones LB, Thornton J and Silva DD, 'Limitations of Risk Based Artificial Intelligence Regulation: A Structuration Theory Approach'

Juhl K, 'Standardization of Risk versus the Risk of Standardization: A Conceptual Analysis' *Standardization and Risk Governance* (Routledge 2019)

Kamara I, *Legal Aspects of Standardisation: Relationship of Standards and Law in the EU* (European Commission 2023)

Kaminski ME, 'Regulating the Risks of AI' [2022] SSRN Electronic Journal <https://doi.org/10.2139/ssrn.4195066> accessed 15 October 2024

Kanevskaia O, 'The Saga of Copyrighted Standards: A Perspective on Access to Regulation' *Research Handbook on Law and Technology* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781803921327/chapter20.xml> accessed 23 April 2024

——, 'Is It Really All about the Money? The Future of European Standardization after PublicResourceOrg' (2025) 16 European Journal of Risk Regulation 344 <https://doi.org/10.1017/err.2024.64>

——, 'Is It Really All about the Money? The Future of European Standardization after PublicResourceOrg' (2025) 16 European Journal of Risk Regulation 344 <https://doi.org/10.1017/err.2024.64>

Kang C, 'OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing' *The New York Times* (16 May 2023) <https://www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html> accessed 3 July 2025

Kazim E and Koshiyama AS, 'A High-Level Overview of AI Ethics' (2021) 2 Patterns <https://doi.org/10.1016/j.patter.2021.100314> accessed 22 November 2024

'Key Players in European Standardisation - Internal Market, Industry, Entrepreneurship and SMEs' <https://single-market-economy.ec.europa.eu/single-market/goods/european-standards/key-players-european-standardisation_en> accessed 28 December 2025

Khan AA and others, 'Ethics of AI: A Systematic Literature Review of Principles and Challenges' *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering* (Association for Computing Machinery 2022) <https://doi.org/10.1145/3530019.3531329> accessed 23 June 2025

Kleinsteuber HJ, 'Self-Regulation, Co-Regulation, State Regulation' <https://cdn.osce.org/sites/default/files/f/documents/7/d/36117.pdf>

Kosmyna N and others, 'Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Task' (arXiv, 10 June 2025) <https://doi.org/10.48550/arXiv.2506.08872> accessed 4 July 2025

Krislov V, *How Nations Choose Product Standards and Standards Change Nations* <https://upittpress.org/books/9780822956228/> accessed 16 November 2023

Larouche P and Baron J, 'The European Standardisation System at a Crossroads' (4 May 2023) <https://doi.org/10.2139/ssrn.4466316> accessed 9 April 2024

Larsson S and Hildén J, 'Flip or Flop? : The Pacing Problems of European Systemic Risk GPAI Regulation' <http://lup.lub.lu.se/record/98428617-b972-4e15-90e7-cc3de05f83dc> accessed 19 August 2025

——, 'Flip or Flop? The Pacing Problems of European Systemic Risk GPAI Regulation' (Social Science Research Network, 15 August 2025) <https://doi.org/10.2139/ssrn.5527918> accessed 23 December 2025

Laux J, Wachter S and Mittelstadt B, 'Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and Acceptability of Risk' (2024) 18 Regulation & Governance 3 <https://doi.org/10.1111/rego.12512>

——, 'Three Pathways for Standardisation and Ethical Disclosure by Default under the European Union Artificial Intelligence Act' (2024) 53 Computer Law & Security Review 105957 <https://doi.org/10.1016/j.clsr.2024.105957>

Leeuwen BV, 'PIP Breast Implants, the EU's New Approach for Goods and Market Surveillance by Notified Bodies' (2014) 5 European Journal of Risk Regulation 338 <https://doi.org/10.1017/S1867299X0000386X>

Leufer D, 'Why Human Rights Must Be at the Core of AI Governance' (*Access Now*, 24 September 2024) <https://www.accessnow.org/human-rights-and-ai-governance/> accessed 3 July 2025

Li Q and others, 'Industry Quality Infrastructure: A Review' *2019 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE)* (2019) <https://doi.org/10.1109/QR2MSE46217.2019.9021243> accessed 30 December 2025

Light R and Panai E, 'The Self-Synchronisation of AI Ethical Principles' (2022) 1 Digital Society 24 <https://doi.org/10.1007/s44206-022-00023-1>

Liza FF, '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>

Loconto A, Stone J and Busch L, *Tripartite Standards Regime* (2012)

Lupo G, 'Risky Artificial Intelligence: The Role of Incidents in the Path to AI Regulation' (2023) 5 Law, Technology and Humans 133 <https://doi.org/10.3316/informit.139102409659905>

Mahler T, 'Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal' (Social Science Research Network, 30 September 2021) <https://papers.ssrn.com/abstract=4001444> accessed 4 January 2026

Malgieri G and Pasquale F, 'Licensing High-Risk Artificial Intelligence: Toward Ex Ante Justification for a Disruptive Technology' (2024) 52 Computer Law & Security Review 105899 <https://doi.org/10.1016/j.clsr.2023.105899>

Manheim K and Kaplan L, 'Artificial Intelligence: Risks to Privacy and Democracy' (2019) 21 Yale Journal of Law and Technology 106

Marchant GE, 'Addressing the Pacing Problem' in Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7_13> accessed 5 July 2025

——, 'The Growing Gap Between Emerging Technologies and the Law' in Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7_2> accessed 5 July 2025

Marchant GE, Allenby BR and Herkert JR (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, vol 7 (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7> accessed 22 October 2024

Martin-Bariteau F, 'Regulating Autonomous Systems in Canada' (Social Science Research Network, 30 January 2023) <https://doi.org/10.2139/ssrn.5104939> accessed 27 June 2025

Maslej N and others, 'Artificial Intelligence Index Report 2025' (arXiv, 8 April 2025) <https://doi.org/10.48550/arXiv.2504.07139> accessed 23 June 2025

Mauritzon I, Holmén M and Nordgren L, 'Paradoxical by Design: Notified Bodies (Nbs) and Artificial Intelligence (Ai) Adoption in Conformity Assessment' (Social Science Research Network, 18 July 2025) <https://doi.org/10.2139/ssrn.5356928> accessed 7 August 2025

McDaniels D and Karttunen M, 'Trade, Testing and Toasters: Bringing Conformity Assessment Procedures into the Spotlight' (2016) 50 Journal of World Trade <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\TRAD\TRAD2016031.pdf> accessed 4 September 2024

McFadden M 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/>

Meijer A and others, 'Towards Responsible Standardisation: Investigating the Importance of Responsible Innovation for Standards Development' (2023) 0 Technology Analysis & Strategic Management 1 <https://doi.org/10.1080/09537325.2023.2225108>

Micklitz H-W, 'Soft Law, Technical Standards and European Private Law' *Research Handbook on Soft Law* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781839101939/book-part-9781839101939-19.xml> accessed 4 January 2026

Mihm SJ, 'Visible Hands and Invisible Standards: The Nuts and Bolts of the Second Industrial Revolution' <https://www.hbs.edu/businesshistory/Documents/Mihm%20HBS%202014.pdf>

Mittelstadt B, 'Principles Alone Cannot Guarantee Ethical AI' (20 May 2019) <https://doi.org/10.2139/ssrn.3391293> accessed 27 November 2023

Moon S and Lee H, 'The Primary Actors of Technology Standardization in the Manufacturing Industry' (2021) 9 IEEE Access 101886 <https://doi.org/10.1109/ACCESS.2021.3097800>

Morgan B and Yeung K (eds), 'Regulatory Enforcement and Compliance' *An Introduction to Law and Regulation: Text and Materials* (Cambridge University Press 2007) <https://doi.org/10.1017/CBO9780511801112.005> accessed 9 July 2025

Moses LB, 'How to Think about Law, Regulation and Technology: Problems with "Technology" as a Regulatory Target' (2013) 5 Law, Innovation and Technology 1 <https://doi.org/10.5235/17579961.5.1.1>

Mündges S and Park K, 'But Did They Really? Platforms' Compliance with the Code of Practice on Disinformation in Review' (2024) 13 Internet Policy Review 1 <https://doi.org/10.14763/2024.3.1786>

Munn L, 'The Uselessness of AI Ethics' (2023) 3 AI and Ethics 869 <https://doi.org/10.1007/s43681-022-00209-w>

Nannini L and others, 'AI Agents Under EU Law' (arXiv, 6 April 2026) <https://doi.org/10.48550/arXiv.2604.04604> accessed 20 April 2026

Neri H, *The Risk Perception of Artificial Intelligence* (Bloomsbury Publishing PLC 2020)

Neri H and Cozman F, 'The Role of Experts in the Public Perception of Risk of Artificial Intelligence' (2020) 35 AI & SOCIETY 663 <https://doi.org/10.1007/s00146-019-00924-9>

Newell WH, 'A Theory of Interdisciplinary Studies' (2021) 19 Issues in Integrative Studies 1

Newman-Toker DE and others, 'Rate of Diagnostic Errors and Serious Misdiagnosis-Related Harms for Major Vascular Events, Infections, and Cancers: Toward a National Incidence Estimate Using the "Big Three"' (2021) 8 Diagnosis 67 <https://doi.org/10.1515/dx-2019-0104>

Nishant R, Kennedy M and Corbett J, 'Artificial Intelligence for Sustainability: Challenges, Opportunities, and a Research Agenda' (2020) 53 International Journal of Information Management 102104 <https://doi.org/10.1016/j.ijinfomgt.2020.102104>

NIST, 'AI Risk Management Framework' <https://www.nist.gov/itl/ai-risk-management-framework>

OECD, 'The OECD Artificial Intelligence (AI) Principles' <https://oecd.ai/en/ai-principles>

Olsen OE, 'The Standardization of Risk Governance' *Standardization and Risk Governance* (Routledge 2019)

Orsay GL, Medeiros KAR and De Oliveira EC, 'Use of Uncertainty Information in Conformity Assessment in thePharmaceutical Industry' (2023) 19 Current Pharmaceutical Analysis 673 <https://doi.org/10.2174/0115734129262343231020105935>

'"Overwhelming Consensus" on AI Regulation - Musk' (13 September 2023) <https://www.bbc.com/news/technology-66804996> accessed 3 July 2025

Pamment J, 'EU Code of Practice on Disinformation: Briefing Note for the New European Commission |' <https://carnegieendowment.org/research/2020/03/eu-code-of-practice-on-disinformation-briefing-note-for-the-new-european-commission?lang=en> accessed 23 December 2025

Pelkmans J, 'The New Approach to Technical Harmonization and Standardization' (1987) 25 JCMS: Journal of Common Market Studies 249 <https://doi.org/10.1111/j.1468-5965.1987.tb00294.x>

Pendrill LR, 'Using Measurement Uncertainty in Decision-Making and Conformity Assessment' (2014) 51 Metrologia S206 <https://doi.org/10.1088/0026-1394/51/4/S206>

Penmetsa P and others, 'Effects of the Autonomous Vehicle Crashes on Public Perception of the Technology' (2021) 45 IATSS Research 485 <https://doi.org/10.1016/j.iatssr.2021.04.003>

Peukert A, 'A New Kid on the EU's Co-Regulatory Block: Codes of Practice under the AI Act' (Social Science Research Network, 15 December 2025) <https://papers.ssrn.com/abstract=5923162> accessed 23 December 2025

Ping W, 'A Brief History of Standards and Standardization Organizations: A Chinese Perspective' [2011] East-West Center Working Papers

Portalier P, 'Myths and Realities of the Presumption of Conformity' <https://orgalim.eu/insights/myths-and-reality-presumption-conformity>

'Principles for the Development of International Standards, Guides and Recommendations' <https://www.wto.org/english/news_e/news19_e/ddgaw_05dec19_e.htm> accessed 28 December 2025

'Proposed EU AI Liability Rules Withdrawn - Bird & Bird' <https://www.twobirds.com/en/insights/2025/proposed-eu-ai-liability-rules-withdrawn> accessed 4 January 2026

Puydarrieux S and others, 'Role of Measurement Uncertainty in Conformity Assessment' in Sandrine Gazal (ed), *19th International Congress of Metrology (CIM2019)* (EDP Sciences 2019) <https://doi.org/10.1051/metrology/201916003> accessed 23 February 2025

Reed C, 'How to Make Bad Law: Lessons from Cyberspace' (2010) 73 The Modern Law Review 903

——, 'How Should We Regulate Artificial Intelligence?' (2018) 376 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20170360 <https://doi.org/10.1098/rsta.2017.0360>

'Regulation on Artificial Intelligence | EESC' (25 March 2021) <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/regulation-artificial-intelligence> accessed 19 August 2025

'Regulatory Sandboxes and Experimentation Clauses as Tools for Better Regulation: Council Adopts Conclusions' (*Consilium*) <https://www.consilium.europa.eu/en/press/press-releases/2020/11/16/regulatory-sandboxes-and-experimentation-clauses-as-tools-for-better-regulation-council-adopts-conclusions/> accessed 23 December 2025

Rejeski D, 'Public Policy on the Technological Frontier' in Gary E Marchant, Braden R Allenby and Joseph R Herkert (eds), *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (Springer Netherlands 2011) <https://doi.org/10.1007/978-94-007-1356-7_4> accessed 5 July 2025

Rességuier A and Rodrigues R, 'AI Ethics Should Not Remain Toothless! A Call to Bring Back the Teeth of Ethics' (2020) 7 Big Data & Society 2053951720942541 <https://doi.org/10.1177/2053951720942541>

Riccio GM, 'Tort Liability and Regulatory Sandboxes' (Social Science Research Network, 4 February 2025) <https://papers.ssrn.com/abstract=5165196> accessed 23 December 2025

Ropeik D, *How Risky Is It, Really?: Why Our Fears Don't Always Match the Facts* (McGraw Hill Professional 2010)

——, 'The Perception Gap: Recognizing and Managing the Risks That Arise When We Get Risk Wrong' (2012) 50 Food and Chemical Toxicology 1222 <https://doi.org/10.1016/j.fct.2012.02.015>

Rosenblatt F, 'The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain' (1958) 65 Psychological Review 386 <https://doi.org/10.1037/h0042519>

Rott P, 'Certification of Medical Devices: Lessons from the PIP Scandal' in Peter Rott (ed), *Certification – Trust, Accountability, Liability*, vol 16 (Springer International Publishing 2019) <https://doi.org/10.1007/978-3-030-02499-4_9> accessed 29 April 2024

Ruschemeier H, 'AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal' (2023) 23 ERA Forum 361 <https://doi.org/10.1007/s12027-022-00725-6>

——, 'AI as a Challenge for Legal Regulation – the Scope of Application of the Artificial Intelligence Act Proposal' (2023) 23 ERA Forum 361 <https://doi.org/10.1007/s12027-022-00725-6>

Russell AL, '"Industrial Legislatures": Consensus Standardization in the Second and Third Industrial Revolutions' (2009) 10 Enterprise & Society 661 <https://doi.org/10.1093/es/khp038>

Saltelli A and others, 'Science, the Endless Frontier of Regulatory Capture' (2022) 135 Futures 102860 <https://doi.org/10.1016/j.futures.2021.102860>

Sayre-McCord G, 'Metaethics' in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Spring 2023, Metaphysics Research Lab, Stanford University 2023) <https://plato.stanford.edu/archives/spr2023/entries/metaethics/> accessed 21 August 2025

SCCS, 'ADDENDUM to the Scientific Opinions on Climbazole (P64) - Ref. SCCS/1506/13 and SCCS/1590/17" SCCS/1600/18' <https://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/sccs_o_220.pdf>

Schepel H, *The Constitution of Private Governance: Product Standards in the Regulation of Integrating Markets* (Hart Publishing 2005)

Scherer MU, 'Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies' (Social Science Research Network, 30 May 2015) <https://doi.org/10.2139/ssrn.2609777> accessed 29 June 2025

Schmidt VA, 'Democracy and Legitimacy in the European Union Revisited: Input, Output and "Throughput"' (2013) 61 Political Studies 2 <https://doi.org/10.1111/j.1467-9248.2012.00962.x>

Schoechle T, 'Digital Enclosure: The Privatization of Standards and Standardization' *ESSDERC 2003. Proceedings of the 33rd European Solid-State Device Research - ESSDERC '03 (IEEE Cat. No. 03EX704)* (IEEE 2003) <https://doi.org/10.1109/SIIT.2003.1251210> accessed 28 December 2025

Schuett J, 'Risk Management in the Artificial Intelligence Act' [2023] European Journal of Risk Regulation 1 <https://doi.org/10.1017/err.2023.1>

Secretariat TB of C, 'Directive on Automated Decision-Making' (2 July 2024) <https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592> accessed 3 July 2025

Separovic L and others, 'Measurement Uncertainty and Conformity Assessment Applied to Drug and Medicine Analyses – A Review' (2023) 53 Critical Reviews in Analytical Chemistry 123 <https://doi.org/10.1080/10408347.2021.1940086>

Shetty DK and others, 'Analyzing AI Regulation through Literature and Current Trends' (2025) 11 Journal of Open Innovation: Technology, Market, and Complexity 100508 <https://doi.org/10.1016/j.joitmc.2025.100508>

Slattery P and others, 'The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence' (arXiv, 10 April 2025) <https://doi.org/10.48550/arXiv.2408.12622> accessed 14 July 2025

Smith B, 'How to Do Things with Documents' [2012] Rivista di estetica 179 <https://doi.org/10.4000/estetica.1480>

——, 'Document Acts' in Anita Konzelmann Ziv and Hans Bernhard Schmid (eds) (Springer Netherlands 2014) <https://doi.org/10.1007/978-94-007-6934-2_2> accessed 6 December 2023

Smith MR and Marx L, *Does Technology Drive History?: The Dilemma of Technological Determinism* (MIT Press 1994)

Smuha NA and others, 'How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act' (Social Science Research Network, 5 August 2021) <https://doi.org/10.2139/ssrn.3899991> accessed 4 January 2026

Smuha NA and Yeung K, 'The European Union's AI Act: Beyond Motherhood and Apple Pie?' in Nathalie A Smuha (ed), *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence* (Cambridge University Press 2025) <https://doi.org/10.1017/9781009367783.015> accessed 23 June 2025

Söderlund K, 'High-Risk AI Transparency? On Qualified Transparency Mandates for Oversight Bodies under the EU AI Act' (2025) 2025 Technology and Regulation 97 <https://doi.org/10.71265/6bedar76>

Soler Garrido J 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

'SSO List | ConsortiumInfo.Org' <https://consortiuminfo.org/sso-list/> accessed 28 December 2025

'Standardisation Regulation – Revision' (*European Commission - Have your say*) <https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/14511-Standardisation-Regulation-revision_en> accessed 16 August 2025

STANDARDIZATION ECF, 'EN 71-7:2014: Safety of Toys - Part 7: Finger Paints - Requirements and Test Methods' (April 2014) <https://law.resource.org/pub/eu/toys/en.71.7.2014.html> accessed 18 November 2025

'Standards, Conformity Assessment, and Trade into the 21st Century'

Stefan OA and others, 'EU Soft Law in the EU Legal Order: A Literature Review' (Social Science Research Network, 4 March 2019) <https://doi.org/10.2139/ssrn.3346629> accessed 4 January 2026

——, 'EU Soft Law in the EU Legal Order: A Literature Review' (Social Science Research Network, 4 March 2019) <https://doi.org/10.2139/ssrn.3346629> accessed 17 August 2025

Steffek J and Wegmann P, 'The Standardization of "Good Governance" in the Age of Reflexive Modernity' (2021) 1 Global Studies Quarterly ksab029 <https://doi.org/10.1093/isagsq/ksab029>

Stelling L and others, 'Existing Industry Practice for the EU AI Act's General-Purpose AI Code of Practice Safety and Security Measures' (arXiv, 21 April 2025) <https://doi.org/10.48550/arXiv.2504.15181> accessed 30 June 2025

Stephenson SM, 'Standards and Conformity Assessment as Nontariff Barriers to Trade' (Social Science Research Network, 1 May 1997) <https://papers.ssrn.com/abstract=597242> accessed 21 February 2025

Stigler GJ, 'The Theory of Economic Regulation' (1971) 2 The Bell Journal of Economics and Management Science 3 <https://doi.org/10.2307/3003160>

Taeihagh A, 'Governance of Artificial Intelligence' (2021) 40 Policy and Society 137 <https://doi.org/10.1080/14494035.2021.1928377>

Taeihagh A, Ramesh M and Howlett M, 'Assessing the Regulatory Challenges of Emerging Disruptive Technologies' (2021) 15 Regulation & Governance 1009 <https://doi.org/10.1111/rego.12392>

Tartaro A, '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 <https://doi.org/10.57230/EJPLT222AT> accessed 27 November 2023

——, ‘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>

——, ‘Alessio Tartaro on the “State of the European Union Address”’ (*Critical AI*, 10 October 2023) <https://criticalai.org/2023/10/10/alessio-tartaro-on-the-state-of-the-european-union-address-9-13-2023/> accessed 3 July 2025

——, ‘Value-Laden Challenges for Technical Standards Supporting Regulation in the Field of AI’ (2024) 26 Ethics and Information Technology 72 <https://doi.org/10.1007/s10676-024-09809-y>

——, ‘The Fourth Wave of Standardisation: A New Kind of Standard-Setting Is Emerging and Standards for Responsible and Trustworthy AI Are Part of It.’ *Proceedings of the 28th EURAS Annual Standardisation Conference - Comprehensive Standardisation for Societal Challenges, 19-21 June 2024, Delft, The Netherlands* (Forthcoming)

Tartaro A and Héder M, ‘Ethics of AI’ *Elgar Encyclopedia of Political Communication* (Edward Elgar Publishing 2025) <https://www.elgaronline.com/display/book/9781035301447/vol1.chapter124.xml> accessed 4 January 2026

Tartaro A and Panai E, ‘Leveraging Technical Standards within AI Regulatory Sandboxes: Challenges and Opportunities’ (Social Science Research Network, 4 February 2025) <https://doi.org/10.2139/ssrn.5235132> accessed 21 August 2025

Tartaro A, Smith AL and Shaw P, ‘Assessing the Impact of Regulations and Standards on Innovation in the Field of AI’ (arXiv, 8 February 2023) <https://doi.org/10.48550/arXiv.2302.04110> accessed 4 May 2023

Tassey G, ‘The Roles and Impacts of Technical Standards on Economic Growth and Implications for Innovation Policy’ (2017) 1 Annals of Science and Technology Policy 215 <https://doi.org/10.1561/110.00000003>

Terpan F, ‘Soft Law in the European Union—The Changing Nature of EU Law’ (2015) 21 European Law Journal 68 <https://doi.org/10.1111/eulj.12090>

‘The General-Purpose AI Code of Practice | Shaping Europe’s Digital Future’ <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai> accessed 23 December 2025

Tilmes N, ‘Disability, Fairness, and Algorithmic Bias in AI Recruitment’ (2022) 24 Ethics and Information Technology 21 <https://doi.org/10.1007/s10676-022-09633-2>

Timmermans S and Epstein S, ‘A World of Standards but Not a Standard World: Toward a Sociology of Standards and Standardization’ (2010) 36 Annual Review of Sociology 69 <https://doi.org/10.1146/annurev.soc.012809.102629>

Tiribelli S and Giovanola B, ‘Weapons of Moral Construction? On the Value of Fairness in Algorithmic Decision-Making’ (2022) 24 Ethics and Information Technology 1 <https://doi.org/10.1007/s10676-022-09622-5>

Trabelsi MA, ‘The Impact of Artificial Intelligence on Economic Development’ (2024) 3 Journal of Electronic Business & Digital Economics 142 <https://doi.org/10.1108/JEBDE-10-2023-0022>

Truby J and others, ‘A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications’ (2022) 13 European Journal of Risk Regulation 270 <https://doi.org/10.1017/err.2021.52>

'Types of Conformity Assessment' <https://www.iec.ch/conformity-assessment/types-conformity-assessment> accessed 30 December 2025

UNECE, 'Standards for Sustainable Development' (2017) <https://unece.org/fileadmin/DAM/trade/wp6/documents/2017/Conference_Report.pdf>

Union PO of the E, 'General Programme of 28 May 1969 for the Elimination of Technical Barriers to Trade Which Result from Disparities between the Provisions Laid down by Law, Regulation or Administrative Action in Member States, CELEX1' (*Publications Office of the EU*, 28 May 1969) <https://op.europa.eu/en/publication-detail/-/publication/74d47cdf-6604-417a-83fc-9a0cb323a617/language-en> accessed 14 August 2025

'Vademecum on European Standardisation in Support of Union Legislation and Policies. PART I - Role of the Commission's Standardisation Requests to the European Standardisation Organisations' <https://ec.europa.eu/docsroom/documents/13507/attachments/1/translations> accessed 4 May 2023

Vallejo R, 'The Private Administrative Law of Technical Standardization' (2022) 40 Yearbook of European Law 172 <https://doi.org/10.1093/yel/yeab011>

van Bekkum M, 'Using Sensitive Data to De-Bias AI Systems: Article 10(5) of the EU AI Act' (2025) 56 Computer Law & Security Review 106115 <https://doi.org/10.1016/j.clsr.2025.106115>

Väyrynen P, 'Thick Ethical Concepts' in Edward N Zalta and Uri Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (Spring 2025, Metaphysics Research Lab, Stanford University 2025) <https://plato.stanford.edu/archives/spr2025/entrieshick-ethical-concepts/> accessed 8 September 2025

Veale M, 'Value-Laden Areas for Standardisation in the AI Act' (*michael veale*, 5 September 2022) <https://michae.lv/value-laden-areas-in-the-ai-act/> accessed 25 April 2024

Veale M, Matus K and Gorwa R, 'AI and Global Governance: Modalities, Rationales, Tensions' (2023) 19 Annual Review of Law and Social Science 255 <https://doi.org/10.1146/annurev-lawsocsci-020223-040749>

Veale M and Quintais JP, 'The Obligations of Providers of General-Purpose AI Models' (Social Science Research Network, 13 November 2025) <https://doi.org/10.2139/ssrn.5744602> accessed 23 December 2025

Verrucci G, *Introduzione Alla Metaetica* (FrancoAngeli 2014)

Voigt P and Hullen N, 'Which Requirements Apply to GPAI Models?' in Paul Voigt and Nils Hullen (eds), *The EU AI Act: Answers to Frequently Asked Questions* (Springer 2024) <https://doi.org/10.1007/978-3-662-70201-7_5> accessed 23 December 2025

Volpato A, 'The Legal Effects of Harmonised Standards in EU Law: From Hard to Soft Law, and Back?' *The Legal Effects of EU Soft Law* (Edward Elgar Publishing 2023) <https://www.elgaronline.com/edcollchap/book/9781802208917/book-part-9781802208917-16.xml> accessed 4 January 2026

——, 'The Legal Effects of Harmonised Standards in EU Law: From Hard to Soft Law, and Back?' *The Legal Effects of EU Soft Law* (Edward Elgar Publishing 2023)

<https://www.elgaronline.com/edcollchap/book/9781802208917/book-part-9781802208917-16.xml> accessed 5 November 2023

Wachter S, 'Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond' (2024) 26 Yale Journal of Law and Technology <https://ora.ox.ac.uk/objects/uuid:0525099f-88c6-4690-abfa-741a8c057e00> accessed 4 September 2024

Waldron J, 'Vagueness in Law and Language: Some Philosophical Issues' (1994) 82 California Law Review 509 <https://doi.org/10.2307/3480971>

Wei K and others, 'How Do AI Companies "Fine-Tune" Policy? Examining Regulatory Capture in AI Governance' (Social Science Research Network, 20 August 2024) <https://doi.org/10.2139/ssrn.4931927> accessed 9 July 2025

Weimer M and Marin L, 'The Role of Law in Managing the Tension between Risk and Innovation: Introduction to the Special Issue on Regulating New and Emerging Technologies' (2016) 7 European Journal of Risk Regulation 469 <https://doi.org/10.1017/S1867299X00006012>

Wenzlhuemer R, 'The History of Standardisation in Europe' [2010] Ego. European History Online <https://ieg-ego.eu/en/threads/transnational-movements-and-organisations/internationalism/roland-wenzlhuemer-the-history-of-standardisation-in-europe>

Werle R and Iversen EJ, 'Promoting Legitimacy in Technical Standardization' (2006) 2 Science, Technology & Innovation Studies 19

'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust | European Commission' <https://commission.europa.eu/document/d2ec4039-c5be-423a-81ef-b9e44e79825b_en> accessed 19 January 2023

'Who We Are – WSC' <https://www.worldstandardscooperation.org/who-we-are/> accessed 28 December 2025

Wiarda M and others, 'Responsible Innovation and De Jure Standardisation: An In-Depth Exploration of Moral Motives, Barriers, and Facilitators' (2022) 28 Science and Engineering Ethics 65 <https://doi.org/10.1007/s11948-022-00415-z>

Wiegmann PM, de Vries HJ and Blind K, 'Multi-Mode Standardisation: A Critical Review and a Research Agenda' (2017) 46 Research Policy 1370 <https://doi.org/10.1016/j.respol.2017.06.002>

Wihbey J, 'AI and Epistemic Risk for Democracy: A Coming Crisis of Public Knowledge?' (Social Science Research Network, 20 April 2024) <https://doi.org/10.2139/ssrn.4805026> accessed 5 July 2025

Wittgenstein L, *Philosophical Investigations* (3nd ed., repr, Blackwell 1953)

Wooldridge M, *A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going* (Flatiron Books 2021) <https://scholar.google.com/scholar?cluster=16348732458466591922&hl=en&oi=scholarr> accessed 23 June 2025

World Trade Organization, *Technical Barriers to Trade: The WTO Agreement Series* (WTO 2014) <https://doi.org/10.30875/7c1b8b12-en> accessed 3 August 2023

Yates J and Murphy CN, 'Coordinating International Standards: The Formation of the ISO' (2007) Working Paper <https://dspace.mit.edu/handle/1721.1/37156> accessed 20 July 2025

——, *Engineering Rules* (Johns Hopkins University Press 2019) <https://doi.org/10.1353/book.66187> accessed 16 November 2023

Yefremov A, 'Regulatory Sandboxes and Experimental Legislation as the Main Instruments of Regulation in the Digital Transformation' in Daniel A Alexandrov and others (eds), *Digital Transformation and Global Society* (Springer International Publishing 2019) <https://doi.org/10.1007/978-3-030-37858-5_7>

Yeung K and Ranchordás S (eds), 'The Rise of Risk in Regulatory Governance' *An Introduction to Law and Regulation: Text and Materials* (2nd edn, Cambridge University Press 2024) <https://doi.org/10.1017/9781009379007.005> accessed 14 June 2025

Zhou P, Song FM and Huang X, 'Environmental Regulations and Firms' Green Innovations: Transforming Pressure into Incentives' (2023) 86 International Review of Financial Analysis 102504 <https://doi.org/10.1016/j.irfa.2023.102504>

Commission implementing ecision 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 - C(2023)3215

Commission notice The 'Blue Guide' on the implementation of EU product rules 2022 (Text with EEA relevance) 2022/C 247/01 2022

Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Artificial Intelligence for Europe 2018

Council Resolution of 7 May 1985 on a new approach to technical harmonization and standards 1985

Decision No 768/2008/EC of the European Parliament and of the Council of 9 July 2008 on a common framework for the marketing of products, and repealing Council Decision 93/465/EEC (Text with EEA relevance) 2008 (OJ L)

Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council Text with EEA relevance 2012 (OJ L)

Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011 (Text with EEA relevance.) 2019 (OJ L)

Stichting Rookpreventie Jeugd and Others v Staatssecretaris van Volksgezondheid, Welzijn en Sport [2022] ECJ Case C-160/20

Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys (Text with EEA relevance)Text with EEA relevance 2022

Regulation (EU) 2022/2480 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 1025/2012 as regards decisions of European standardisation organisations concerning European standards and European standardisation deliverables (Text with EEA relevance) 2022 (OJ L)

Commission Implementing Decision (EU) 2023/740 of 4 April 2023 on harmonised standards for toys drafted in support of Directive 2009/48/EC of the European Parliament and of the Council 2023 (OJ L)

Directive 2014/40/EU of the European Parliament and of the Council of 3 April 2014 on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products and repealing Directive 2001/37/EC (Text with EEA relevance)Text with EEA relevance 2023

Commission Implementing Decision (EU) 2025/1785 of 9 September 2025 amending Implementing Decision (EU) 2023/740 as regards the harmonised standards for certain ride-on toys (wave rollers) drafted in support of Directive 2009/48/EC of the European Parliament and of the Council 2025

Coordinated Plan on Artificial Intelligence 2018