Regulation for Systemic Artificial Intelligence Safety: Interoperable Regulations and a Pragmatic Enforcement Strategy for Mitigating Unintended Consequences

Farhana Ferdousi Liza, School of Computing Sciences, University of East Anglia, UK.

Submitted 15th July 2024. Accepted 9th November 2024.

1 Introduction

Developing and managing systemically safe, advanced Artificial Intelligence (AI) technologies has become increasingly critical due to the rapid development and deployment of AI systems across various sectors, together with the evolving regulatory landscape. Ensuring AI safety, trustworthiness and fostering global cooperation are paramount for effective governance and maintaining power relations conducive to a liberal future. Achieving these goals calls for a balanced regulatory approach and being prepared for unintended consequences. A systemic and pragmatic approach to AI regulation can assure the safe, responsible use of advanced AI technologies while promoting cross-border business and technological research and innovation.

Governments and regulatory bodies around the world are developing regulations to address a range of AI-related concerns like safety, ethics, bias, fairness, security, privacy, transparency, and accountability. Notable examples of such regulations include the European Union's AI Act (EU AI Act), the United States' AI Bill of Rights, and various national AI strategies, including the United Kingdom's (UK) National AI Strategy. Stanford University has reported a huge increase in the number of countries with laws containing term "AI" – growing from 25 countries in 2022 to 127 in 2023 (Marcin, 2024). These regulations aim to protect rights and liberties by safeguarding individuals' rights and preventing discrimination. They envision compliance by establishing standards and guidelines for the development and deployment of AI systems, and promote innovation by encouraging the responsible development of AI technologies that benefit society and the economy.

While regulations are essential for guiding the safe and legal use of AI, it is crucial to analyse their unintended consequences, particularly perverse results, to make sure they do not stifle innovation or create new problems. Overly stringent, partially analysed and difficult-to-enforce regulations might hinder technological innovation, increase discrimination, and slow the advancement of beneficial AI research and innovation. Regulations can hold considerable social, economic, and power implications, affecting businesses' operating costs and competitiveness. Understanding these impacts is vital for designing policies that support economic growth while ensuring responsible AI use in a liberal society.

Small and medium-sized enterprises (SMEs) and the public sector may find it challenging to comply with complex regulations, potentially leading to less competition and innovation and making large enterprises more powerful. Assessing the compliance burden can guide the creation of more accessible and scalable SME friendly regulatory frameworks. Different regions are adopting varying regulatory approaches, bringing inconsistencies and potential disparities in AI development and deployment globally. Analysing these differences can promote harmonisation and international cooperation. Some regulatory requirements may be technologically challenging to implement. Analysing the feasibility of these requirements assures they are practical and achievable with the current knowledge of science and technologies. Further, regulations can influence the behaviour of developers, users and organisations. Understanding these incentives can help predict and mitigate any negative behavioural changes, such as gaming the system, finding loopholes, or focusing on compliance over innovation.

This chapter analyses issues in international cooperation (interoperability) and technological challenges in implementing AI technologies able to comply with current AI regulations across borders. The discussion aims

to foster responsible AI development through a systemic approach to safety, a forward-looking AI regulatory system and a pragmatic enforcement plan within the European Union (EU) and beyond.

We first discuss the unintended consequences of past regulations and technology adoptions as case studies in Section 2. We show that striking the right balance in regulation across key parameters – timing, scope and method – is indispensable for mitigating unintended consequences. Next, in Section 3 we conduct a concise evidence synthesis to gain an understanding of the global landscape of AI regulation. We then briefly study the regulatory goals of key jurisdictions – the UK, the USA and the EU – in Section 4. In Section 5, we examine cross-cutting challenges in enforcing AI regulation across borders. Section 6 explores divergent regulatory approaches to AI. In Section 7, we address the problem of interoperability between statutory and non-statutory regulatory frameworks. Section 8 presents a computational analysis of the enforceability of data requirements as an example. Finally, before concluding, in Section 9 we combine our analysis of AI regulation and unintended consequences with the framework provided by the sociologist Robert K. Merton.

2 Unintended consequence: drawing insights from historical technology adoption and regulations

In regulatory analysis and debate, anecdotal evidence and case studies offer valuable insights. Studies in tissue engineering (Faulkner, 2009), agro and pharmaceutical biotechnology (Chataway et al., 2006) and pharmaceuticals (Abraham and Davis, 2007) reveal the scientific and pragmatic value of detailed analyses often grounded in social or political theory. Given the evolving nature of AI regulations, it may be premature to draw definitive theoretical conclusions. Still, historical examples can illuminate the main parameters that guide our analysis concerning the unintended consequences of AI regulations. The emergence of new technologies and their adoption, coupled with the introduction of new regulations, is not unprecedented. Examining historical examples helps us understand the impacts of regulations and technology adoption on people, the economy, society, and businesses over time.

Example 1: Impact of strict regulation on economic growth

An illustrative case of how strict regulation can hinder economic growth is India's Licence Raj system that was in place from around 1950 to 1991. The Licence Raj comprised the elaborate system of licences, regulations and red tape required to set up and run a business in India. These restrictive policies caused sluggish economic growth, often referred to as the "Hindu rate of growth", which hovered around 3.5% annually. The economy remained largely agrarian, with slow industrial development and high levels of poverty.

Economic reforms in 1991, which included reducing government control over businesses, lowering tariffs, and encouraging foreign investment, led to a significant acceleration of the economy. India emerged as one of the fastest-growing major economies in the world. Declassified documents from the World Bank and International Monetary Fund (IMF), reported by The Indian Express in 2010, reveal how both organisations had pressed India to transition from a centrally planned to a market-driven economy. This example shows the potential of overly stringent regulations to stifle economic growth and innovation.

The EU's AI Act is considered to be a strict regulation (Euronews, 2023). Although it is challenging to foresee all the unintended consequences of this regulation, given that it only came into effect on August 1, 2024, historical examples and existing literature suggest that stringent regulations generally inhibit research and innovation (Stewart, 1981).

Example 2: Impact of faulty technology on people and society

In 1999, while the UK was still a member state of the EU, the Postmaster Scandal, also known as the Horizon scandal, unfolded. It entailed the wrongful prosecution of over 700 sub-postmasters and sub-postmistresses due to errors in information technology (IT) support for UK post offices – the Post Office's Horizon IT system. Despite early indications of discrepancies, the Post Office steadfastly defended the system's reliability, resulting in numerous unjust convictions for theft and fraud. The lack of effective central oversight

compounded sector-specific issues. Ideally, the system should have undergone a thorough evaluation both before and during its deployment to mitigate the technical loopholes and perceived vulnerabilities. Persistent advocacy efforts and a pivotal 2019 High Court ruling finally exposed the systemic flaws. The scandal was responsible for serious consequences, including financial ruin, mental health challenges and, tragically, suicides among those affected. In response, the UK government initiated a compensation programme and launched a public inquiry to scrutinise the shortcomings. This case points to the critical need for stringent but pragmatic and enforceable technological oversight (both centrally and institutionally), transparency, and comprehensive institutional reforms.

Example 3: Impact of a regulatory loophole or hasty regulation on people and society

It is often not fully appreciated that the Enabling Act of 1933 in Germany was not an unheard of measure within its historical context. This legislation granted Adolf Hitler's government the power to enact laws without needing to involve the Reichstag, effectively establishing a dictatorial authority. Prior to 1933, the unstable Weimar Republic had already experienced a series of enabling acts intended to circumvent constitutional challenges. This historical example highlights the importance of considering the broader implications of such hasty measures – addressing one problem may inadvertently create another, potentially more pressing issue. Moreover, loopholes in the legal system and the influence of money can undermine regulatory frameworks. For instance, in the USA, Robert Durst (Editors, 2022), despite admitting to having 'accidently' killed and dismembered a body in 'self-defence', was able to avoid jail time by being able to recruit the best and most expensive lawyers and leveraging substantial bail funds, thereby circumventing both legal consequences and social stigma associated with his actions.

Although these analogical comparisons are disturbing given their association with mass genocide and killing, they may be useful for understanding the scenario. In the contemporary context of regulating artificial intelligence (AI), one should be mindful of this lesson – the hasty making of regulations, underestimating the power of money, and the inevitability of legal loopholes. While regulation is necessary to manage the risks associated with AI, care must be taken to avoid introducing new problems that could arise from cross-boundary business needs and the imperative for research and innovation. Balancing regulation with the need to foster technological advancement requires careful and thoughtful policymaking to make sure the solution (regulation) does not become a bigger issue than the original problem.

On a hasty regulation note, some European legislation on AI had been anticipated at least since 16 July 2019. On that date, Ursula von der Leyen pledged that, within 100 days of being elected President of the European Commission, she would propose new legislation on AI (Floridi, 2021). It remains unclear, though, how this 100-day timeline was determined.

Example 4: Impact of late technology adoption on businesses

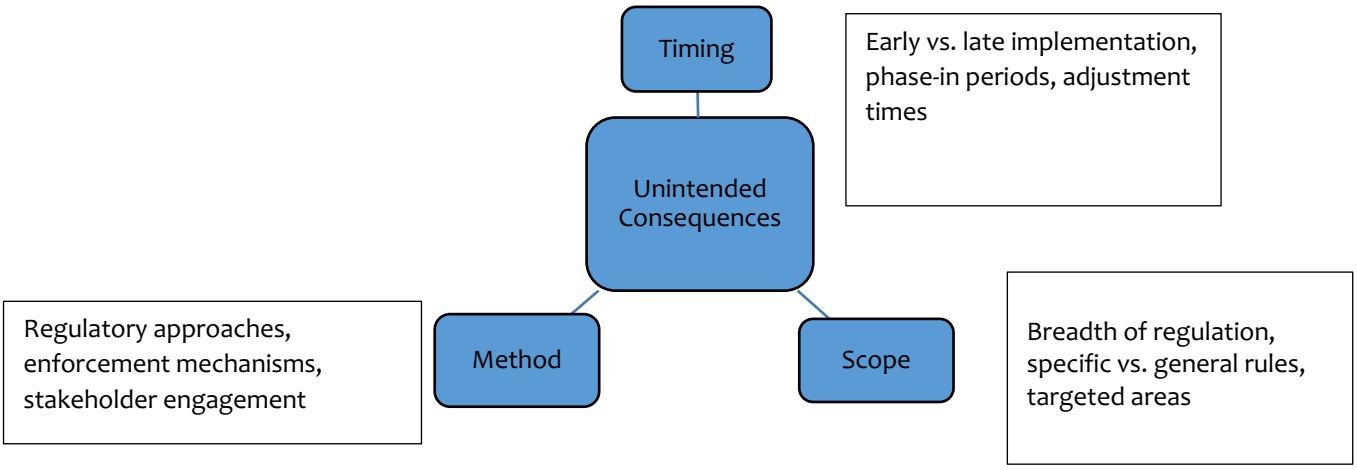
Kodak's decline following to delay in adopting digital photography serves as a poignant example of the adverse consequences of postponing technological innovation. Despite pioneering the first digital camera in 1975, Kodak remained focused on its lucrative film business, thereby missing the digital revolution led by competitors like Canon and Sony. This strategic hesitancy led to a significant loss of market share, financial deterioration and, ultimately, bankruptcy in 2012. Kodak's downfall shows the perils associated with resistance to innovative changes and misalignment with prevailing market trends. While an individual entity's adoption may impact its trajectory (e.g., Kodak), regulations can impact a whole region systemically. The systematic adoption of new technologies can be complicated by stringent regulations (e.g., the EU AI Act), added compliance burdens, and heightened liability (Wendehorst, 2020). Such regulatory challenges hold the potential to stifle innovation as they create additional obstacles for businesses attempting to adapt to and integrate emerging AI technologies.

Striking the right balance in key parameters (timing, scope, method)

From these examples, we may conclude that achieving the right balance in adopting technology and regulating those technologies is vital. Based on the above case-study discussion, to mitigate unintended consequences, three main parameters must be considered: timing, method and scope. First, appropriate

timing involves implementing regulations neither too early nor too late, with phase-in periods and adjustment times to facilitate smooth adoption. These strategic tools help stakeholders ease into compliance, reduce immediate burdens, and allow for necessary infrastructure and skill development. For instance, the GDPR's 2-year transition period and new vehicle emission standards' phase-in periods illustrate this approach. Second, the method must be stable, specific and future-proof so as to enable flexible enforcement and mitigate unintended consequences. This includes making sure of the interoperability of regulatory approaches and robust enforcement mechanisms, as well as appropriate stakeholder engagement. Finally, the scope involves determining what to regulate, the breadth of regulation, and whether to use specific or general rules, focusing on targeted areas. For AI, this is particularly complex due to cross-cutting challenges, which means it is essential to carefully balance these parameters to support long-term compliance, research and innovation. The figure below provides an illustration.

Figure 1: Striking the right balance in key parameters (timing, scope, method)



Timing: Market disruption, innovation lag, compliance rush

Method: Overregulation, underregulation, enforcement gaps

Scope: Sector-specific issues, regulatory loopholes, comprehensive burden

Source: Own

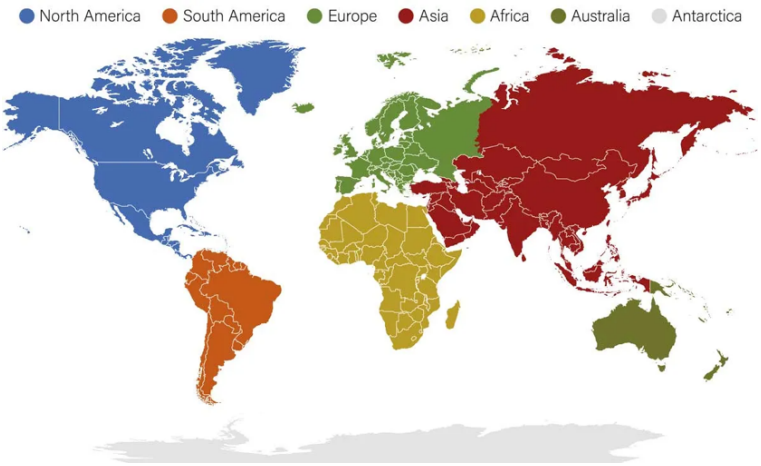
3 The evidence synthesis of global AI regulations

This section provides a concise overview of how AI regulation is developing around the world. The rapid advancement of AI is prompting countries and regions to develop regulations tailored to their unique cultural and political landscapes. According to the September 2023 Global AI Legislation Tracker, "countries worldwide are designing and implementing AI governance legislation commensurate to the velocity and variety of proliferating AI-powered technologies. Legislative efforts include the development of comprehensive legislation, focused legislation for specific use cases, and voluntary guidelines and standards" (Marcin, 2024).

Continent-wise (see Figure 2), in North America the United States is shaping its regulatory landscape through federal initiatives and ethical principles along with state-based legislations, whereas Canada is advancing its AI oversight with the Pan-Canadian AI Strategy and Algorithmic Impact Assessment (AIA). In Asia, China's National AI Development Plan and Ethical Guidelines, Japan's AI Strategy 2021 and contributions to global AI standards, and other initiatives are driving regulatory developments. Europe is considerably influenced by

the EU's AI Act, with individual countries like the UK, Germany, France and Italy also pursuing their national AI strategies. In South America, Brazil is at the forefront, while in Africa, South Africa and Kenya are leading in AI strategy and regulatory development. Australia complements its AI Ethics Framework and AI Action Plan with regulatory sandboxes to test AI applications in controlled conditions. Notably, Antarctica remains without any dedicated AI regulatory initiatives.

Figure 2: Continents of the world



Source: Toby, 2023

Despite the diversity of these regulatory efforts, common themes emerge, particularly in addressing ethical and legal considerations to ensure safety, security, transparency, fairness, accountability, and respect for human rights. Ethical frameworks often precede legal regulations, given their foundational role and slower legal development pace, especially when interoperability is critical and International Law is suggested as the principal legal framework for the regulation of AI (Carrillo, 2020). Promoting AI research, development and innovation via public-private partnerships, funding and international collaboration is another common priority, albeit the emphasis varies by region. The balance between promoting innovation and ensuring regulation also differs, with the EU generally adopting more stringent controls than the more flexible, innovation-friendly approaches of the USA and UK. Countries like the USA, the UK and Australia favour sectoral regulations, addressing specific industries such as health, transportation, finance, the public sector, smart city technology, and robotics. In contrast, the EU is aiming for a comprehensive, risk-classified cross-sectoral framework. Some regulations are statutory, as seen in the EU, whereas others are non-statutory with a possibility of statutory regulation in the future, such as in the UK, which would have long-term implications for adoption and enforcement.

Given the substantial early contributions to frontier AI developments from USA industries, it is likely that the USA will play a leading role in setting international standards for AI regulations. The preference for sectoral regulation, which is currently favoured by several countries and in line with the USA approach, contrasts with the EU's risk-based methodology. The entity that establishes the dominant and pragmatic regulatory framework for AI and possesses the capacity and influence to enforce it will significantly impact the global balance of power concerning AI.

4 Goal of the AI technology regulation by key jurisdictions

In this section, we are considering the goals and approaches of key jurisdictions such as the EU, USA and UK, which are continuously working to regulate AI. Analysing all jurisdictions lies beyond the scope of this chapter. The EU has finalised legislations to regulate AI which came into effect on 1st August 2024. In contrast, the Conservative UK government argued that it was premature to legislate effectively given the present stage of AI technology's evolution, suggesting that doing so now might be counterproductive. The Artificial Intelligence (Regulation) Bill, a private member's bill proposed by Lord Holmes of Richmond, seeks to

establish a new body, the AI Authority, tasked with addressing AI regulation in the UK. Following the UK General Election on 4 July 2024, the new Labour government may well change the current approach (BTO, 2024). In the USA, AI regulation is being explored on both federal and state levels. While some states have introduced legislation focusing on privacy and accountability, there is no federal legislation yet. Instead, the White House issued an executive order in October 2023 outlining key principles and actions for the safe development and use of AI. Other countries are also initiating AI regulations, each at different stages of development.

Given the EU's advanced stage in regulating AI, the purpose of the regulation is articulated in Article 6 of the EU's AI Act: "The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, and fundamental rights enshrined in the Charter, including democracy, the rule of law, and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation". There are 85 articles, and several governing bodies are set up for proper enforcement, including an AI Office within the Commission to enforce the common rules across the EU, a scientific panel of independent experts to support enforcement activities, an AI Board with member states' representatives to advise and assist the Commission and member states on consistent and effective application of the AI Act, and an advisory forum for stakeholders to provide technical expertise to the AI Board and the Commission.

In contrast, the UK government outlined its approach to AI regulation in the March 2023 White Paper "A Pro-Innovation Approach to AI Regulation". The White Paper does not propose creating a new AI regulator or introducing primary legislation to establish AI regulatory principles or structures. Instead, the government intends to implement a new framework to bring "clarity and coherence" to the AI regulatory landscape, underpinned by five principles: safety, security and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress. These principles will be addressed on a non-statutory basis and implemented by existing sectoral regulators. The framework relies on the UK's existing sectoral regulators, supplemented by government-led "central functions" to provide support, coordination and coherence. Therefore, the AI regulation approach is still in progress but remains pro-innovation and the UK government has allocated GBP 10 million in funding for regulators to develop skills and enhance capabilities.

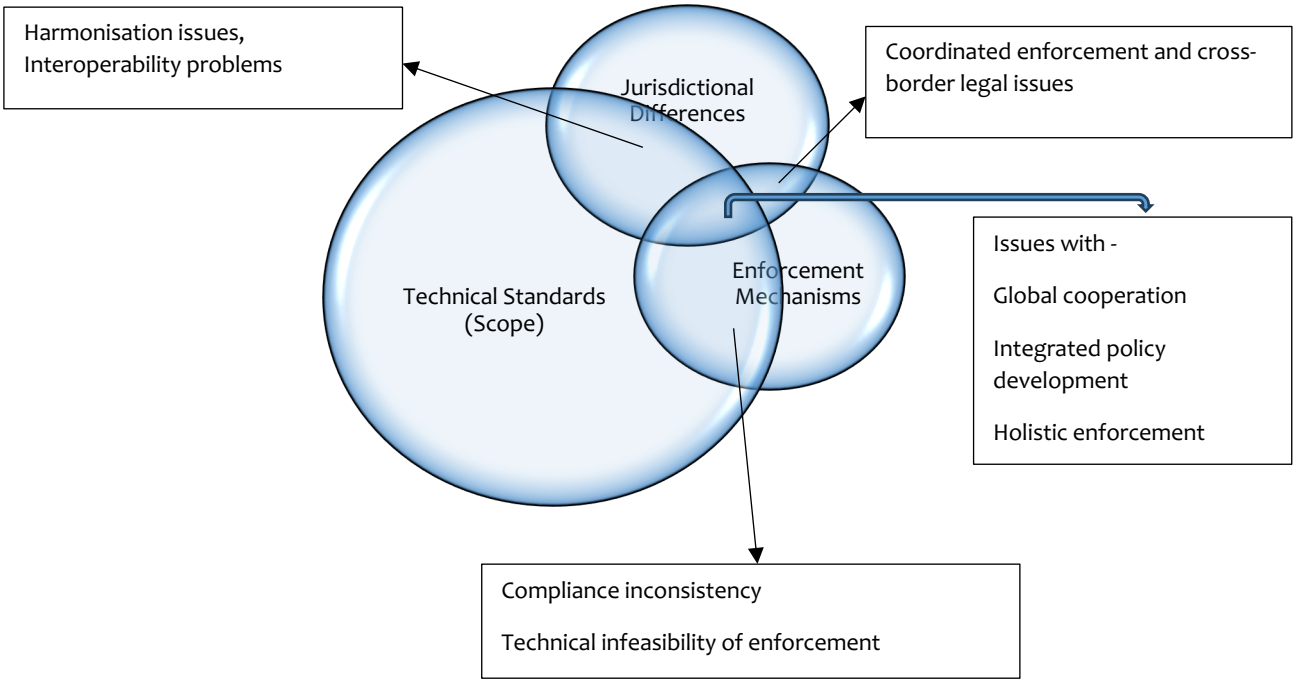
In the United States, AI regulation is being examined on both state and federal levels. On the federal one, the White House issued an executive order in October 2023 directing new standards for AI safety and security and several measures on privacy. The order also contained measures on advancing equity and protecting civil rights; the impact of AI on consumers, patients and students; supporting workers; promoting innovation and competition; assuring responsible and effective government use of AI; and "advancing American leadership abroad". Under the last heading, the order refers to the expansion of bilateral, multilateral and multi-stakeholder engagement, and the acceleration, development and implementation of "vital AI standards with international partners and in standards organisations, ensuring that the technology is safe, secure, trustworthy, and interoperable". Since 2019, 17 states have enacted 29 bills focused on regulating the design, development and use of AI. These bills primarily address two regulatory concerns: data privacy and accountability. Legislatures in California, Colorado and Virginia have established regulatory and compliance frameworks for AI systems. In the USA budget announced in March 2024, President Biden also noted there would be significant new funding (over USD 3 billion) aimed at developing responsible AI.

While key jurisdictions take distinct approaches to AI regulation, the USA aims to promote interoperability with international regulatory frameworks. The chief goal of regulating AI across different jurisdictions is to make sure that AI technologies are developed and utilised in a way that maximises their benefits for innovation and the economy, while mitigating potential risks and harms to individuals, society and the environment. The implementation of AI regulation goals varies in several respects: timing (e.g., the EU has already implemented regulations, whereas the UK and USA are still in progress), method (e.g., the EU adopts a strict but innovation-supportive approach, while the UK and USA are more pro-innovation, with the USA focusing on cross-border interoperability and the EU on regional interoperability), and scope (e.g., the EU employs a risk-based approach, while the USA and UK use a sectoral approach). Further, budget allocations for enabling responsible AI development and deployment vary, with the UK allocating GBP 10 million and the USA USD 3 billion.

5 Cross-cutting challenges with enforcing AI regulation across borders

Now that we understand the goals of different jurisdictions, we shall turn to how to technically support compliance and enforcement. Shetty (2024) noted that "The EU needs the technical standards supporting its AI Act to be restrictive enough to protect consumers, but flexible enough to enable innovation. Given society's current understanding of AI, there are serious doubts as to whether such standards are technically feasible". Cross-cutting challenges (CC) impact all aspects of regulation development and enforcement plans, influencing each other. We will discuss the major challenges and their relationship to unintended consequences within the scope of this chapter, as illustrated in the figure below.

Figure 3: Cross-Cutting Challenges (in circles) and Unintended Consequences (in boxes) in AI Regulation



Source: Own

Cross-cutting challenges 1 (CC1) :

Part 1: Technical standard – lack of a standardised definition (related to the Scope parameter)

The lack of public awareness about AI largely stems from insufficient knowledge. Mythology, culture, religion, literature and science fiction have contributed to an anthropomorphic view of AI (Ramírez, 2018; Muehlhauser & Helm, 2012). Technologically compliable effective regulation of artificial intelligence in contrast requires precise and universally accepted terminologies. The fact that currently there is no cross-border consensus on the legal definition of AI (Samoili et al., 2020; Begishev et al., 2020; Schuett, 2023) poses significant challenges for the regulatory clarity and consistency, policy development, trusted AI innovation, and establishment of uniform ethical and safety standards. This fundamental issue complicates the sociotechnical landscape, notably for advanced AI technologies like Large Language Models and innovations such as ChatGPT. Within existing regulatory frameworks, these technologies can be categorised as either high- or low-risk depending on their application – high-risk in healthcare, for example, and low-risk in the music industry.

Autonomy is another key term used in regulations to describe AI characteristics. However, what constitutes autonomy remains ambiguous. Advanced AI technologies capable of learning and producing contextualised outcomes present greater legal and enforcement challenges than purely technical uncontextualised automation, such as vending machines. For example, an AI system is not akin to an automated coffee vending machine where pressing a button brings about a predictable outcome – a cup of coffee. Each AI system varies

significantly in its learning capabilities and types of autonomy. For simplicity, autonomy can be conceptualised as the degree of variance with respect to a context inherent in an AI system.

Moreover, regulations often call for the explanation of decision-making processes. But what exactly is meant by explanation? In the sciences, rationalism's focus on reason, logic, and clear deductive processes enhances the explainability of knowledge. Rationalism in social science creates logical, clear and systematic interpretive explanations of social phenomena. Scientific explanations aim to discover universal laws and principles that can predict and explain natural phenomena. In contrast, social science seeks to understand and interpret social phenomena within specific contexts rather than discovering universal laws.

For advanced AI models, providing explanations involves offering understandable reasons for the model's outputs or decisions. While interacting with humans, these explanations need to adapt to specific contexts. AI models, which identify patterns and generate responses based on their training data, require clarity concerning what type of explanation (deductive vs. interpretive) is mandated by AI regulations. This clarity is essential for ensuring that AI systems are transparent and accountable.

Part 2: Technical standard – tackling challenges in logical reasoning, explanation and uncertainty quantification in AI systems (related to the Method parameter)

Artificial Intelligence began to emerge as a field in the mid-20th century, influenced by seminal events such as the proposal of the Turing Test to assess machine intelligence and coining of the term "artificial intelligence" at the 1956 Dartmouth Conference where researchers gathered to discuss how machines could simulate human intelligence. A few decades later, advanced AI technologies like Large Language Models (LLMs) and ChatGPT reached a development stage with the potential to disrupt the socio-legal-technological landscape. These technologies rely on information theory, statistics, and probability theory, utilising evaluation measures such as perplexity and accuracy during training. However, systematic reasoning strategies like logic, decision trees or uncertainty quantification methods such as Bayesian modelling were not technically feasible in the training objectives of these models, which often use algorithms like backpropagation for the neural networks. From the beginning to the present, two main philosophies – rationalism and empiricism – have been explored in the AI domain, specifically in computational linguistics, within which Large Language Models (LLMs) fall. These philosophies can be divided into four eras, as shown in Table 1.

Table 1: Historical stages (Church & Liberman, 2021) of the philosophical adoption of research approaches in AI (specifically Computational Linguistics)

Era	Philosophy	Approach
1950s	Empiricism I	Information theory
1970s	Rationalism I	Formal language theory and logic
1990s	Empiricism II	Stochastic grammars
2010s	Empiricism III	Deep neural networks

In Table 1, logic played a larger role during the rationalism emphasis in the 1970s, while probability was more prominent in the empiricism phases during the 1950s and 1990s. In the 2010s, both logic and probability faded into the background as deep neural networks introduced a procedural, associationist flavour of empiricism. One might ask why rationalism has not been revived until now. Despite numerous attempts to incorporate reasoning into state-of-the-art advanced AI technologies, no significant breakthroughs have been made. Achieving reasoning in advanced AI technologies remains a considerable challenge within the current scope

of known science. The integration of logical reasoning and uncertainty quantification into AI systems is essential for advancing their capabilities and assuring robust and trustworthy outcomes, although that requires significant technical hurdles to be overcome. Imposing laws on requirements that are technically infeasible seems to amount to unrealistic criteria in the current landscape of legislations.

Part 3: Technical standard – challenges in establishing a social AI agent align with the EU AI Act (related to the Scope, Timing and Method parameters)

What these regulations are calling for is a Social AI Agent – one capable of understanding ethics, social customs such as fairness, able to adapt the explanation of an activity of decision based on the context, and essentially be part of society. Creating a socially adept AI agent necessitates significant technical and methodological advancements in AI, natural language processing, computer vision, machine learning, and ethical AI design. For example, transforming ChatGPT into a social agent capable of engaging in human-like conversations involves overcoming several key challenges. First, ChatGPT must adapt its responses to diverse cultural contexts and continuously learn from user interactions to handle complex social scenarios effectively. This requires a nuanced understanding of the context and intent behind user queries. Developing emotional intelligence within ChatGPT is also crucial in such scenarios; the AI needs to recognise and respond empathetically to user emotions to facilitate meaningful and safe interactions. In addition, integrating multimodal capabilities to handle inputs such as text, images and videos can enrich the conversational experience. Consistency in dialogue, respect for privacy, and adherence to ethical standards are essential for building and maintaining user trust. Nonetheless, known science and technology has not invented these mechanisms yet and we do not know when that will be possible.

Cross-cutting challenges 2 (CC2):

Jurisdictional differences – innovation across borders and adapting cross-border contracts (related to the Method parameter)

The vagueness of definitions along with technically infeasible criteria in regulations complicates the harmonisation of international regulations on business and sales across various systems and contexts. While one can find existing legal mechanisms like the United Nations Convention on Contracts for the International Sale of Goods (CISG), questions have been raised about whether changes in contract law are needed to accommodate issues specific to AI (Janssen, 2022). The CISG aims to provide a uniform and equitable framework for international trade by harmonising the laws governing cross-border sales transactions.

However, emerging concerns with AI are that existing frameworks like the CISG might not adequately address the unique challenges posed by AI technologies. These challenges include the autonomy of AI systems, which can affect contract performance and liability. Therefore, a critical examination of how current international contract laws, such as the CISG, can be adapted to better address the complexities introduced by AI is essential. This includes considering new definitions and legal standards that can harmonise international AI regulations, thereby promoting businesses and innovation while ensuring legal clarity and fairness across borders.

Cross-cutting challenges 3 (CC3):

Technical standard, jurisdictional differences and enforcement mechanisms – developing fit-for-purpose transdisciplinary research (related to the Scope, Method and Timing parameters)

A pragmatic and enforceable AI regulation can be achieved by integrating sophisticated techniques across formal, natural and empirical sciences (such as logic, physics, mathematics, statistics, and computer science), social sciences (including sociology, psychology, political science, law, and education), and engineering disciplines (e.g., GPU technology). Effective AI regulation calls for a comprehensive understanding that transcends individual scientific and engineering domains. AI systems, designed to operate within a human–computer interaction framework, embody principles from these diverse fields, combining both hardware and software components. Human involvement by way of designers, developers, policymakers and users introduces crucial social science elements into the equation. The call for secure, trustworthy and ethical AI,

alongside requirements for explainability, transparency, data quality, and data quantity (as stipulated in Article 10 of the EU AI Act), presents significant challenges that current scientific capabilities may find difficult, if not impossible, to meet. Often, these requirements stem from a lack of interdisciplinary comprehension among formal, natural, social sciences, and engineering domains. Technocrats who adopt a transdisciplinary approach – seeing the whole picture – offer a promising path forward to resolving these issues. The interdisciplinary collaboration needed to create and enforce effective regulations adds another layer of complexity, underscoring the fundamentally transdisciplinary nature of the problem of AI regulation.

6 Divergent regulatory approaches to AI

"More law, less justice" is a quote by Marcus Tullius Cicero that conveys the idea that too many laws can lead to injustice. We now elaborate on jurisdictional differences, inconsistencies and potential scope for injustice. The regulatory strategies employed by the European Union (EU), the United States (USA) and the United Kingdom (UK) reflect fundamentally divergent methodologies. The EU's risk-based framework operates on a macro level, utilising top-down regulation, whereas the USA and UK have adopted sectorial approaches that function on a micro level with bottom-up regulation. While these approaches hold the potential to complement each other through rigorous international collaboration, they also bring risks of unintended consequences, particularly during transitions between the sectorial and risk-based frameworks essential for achieving cross-border interoperability.

Challenges in interoperability and consistent enforcement

Interoperability and consistent enforcement pose considerable challenges not only among the EU, UK and USA, but also within the EU itself if any member state wants to adopt a bottom-up approach in the future. Concerns have been already expressed regarding the alignment between the EU AI Act and the regulatory and policy implementations of member states. For instance, Gilbert (2024) notes that "the wording of many aspects of the Act is ambiguous, with high-level objectives stated and details expected in subsequent guidance, standards, and member state laws". The prevalence of vague objectives and lack of specificity are recurring themes in the literature (Liza, 2022).

Moreover, the EU AI Act imposes new responsibilities on developers, deployers, notified bodies, regulators, and the newly established EU AI Office and European Commission. Its impact extends to AI-driven technologies (e.g., Digital Health Technologies) developed and deployed both within and outside the EU (Gilbert, 2024). Innovation in public sectors like healthcare is consequently likely to decelerate due to inadequate funding for routine operations compounded by vague legislation and heightened compliance burdens. This lack of clarity, combined with regulatory burdens and interoperability issues, is anticipated to hinder innovation.

Exclusions and their Implications

While the legislation imposes stringent regulations on AI-based technologies, its exclusion of military and defence sectors complicates interoperability and introduces the risk of discriminatory practices within risk-based regulations. European Digital Rights (2024) observes that "the legislation establishes a separate legal framework for AI use by law enforcement, migration control, and national security authorities, potentially creating loopholes and endorsing the use of intrusive systems for discriminatory surveillance on marginalised communities".

As a result, even relatively low-risk AI innovations in public sectors such as healthcare may stall due to concerns with compliance costs, while more controversial AI applications in military and defence could evade scrutiny, potentially exacerbating discrimination against marginalised groups, including migrants. The term "migrant" lacks a universally accepted legal definition in international law and generally refers to individuals who have left their homes, whether within their own country or across borders.

Regulatory coherence and potential self-contradictions

This situation raises significant concerns regarding the coherence of regulations and potential self-contradictions. If migrants from EU member states are subjected to discriminatory AI applications by military and defence sectors of other member states, that would violate their fundamental human rights and contradict the EU AI Act's objective of safeguarding the rights of all EU citizens.

7 Statutory Vs. non-statutory regulations

Such jurisdictional differences reveal variations in enforcement plans. Regulatory approaches to AI can be broadly categorised into statutory and non-statutory frameworks. The EU is adopting a statutory regulatory approach; the USA has not decided yet – either as standalone legislation or as AI-related provisions and clauses inserted into broader acts – so at the federal level, the guidelines are non-statutory, and the UK is adopting non-statutory regulation and when the time is right the UK might take a statutory approach. As we can see, the 'timing' parameter is playing its role here. Statutory regulations are formal laws enacted by legislative bodies, whereas non-statutory regulations consist of guidelines, principles, and codes of practice that lack legal enforceability. This section explores the interaction between these two regulatory types and examines the unintended consequences that may arise when 'timing' is assessed differently.

Multiple entities will be subject to AI regulations (both statutory and non-statutory), including small and medium-sized businesses, large companies like Facebook, Google and Twitter, public sector organisations, and the financial sector. If a UK-based entity sells or distributes its AI tools within the EU, those tools must comply with the EU's AI Act. Similarly, if EU-based businesses use AI tools developed in the UK or any other country, those tools must also comply with that Act. Entities willing to grow and adopt AI innovation but struggling with compliance, vagueness and the burden of regulation might consider relocating operations to areas with non-statutory regulations or exploiting regulatory loopholes resulting in compliance on paper but not in spirit, and potentially creating new risks. This could lead to superficial compliance, adding to potential risks. More serious consequences include evasive tactics, such as creating off-the-books operations or using covert methods to avoid strict regulations. Entities might also reduce their transparency by openly sharing less information to avoid regulatory repercussions.

The timing of adopting statutory regulations is crucial. Non-statutory regions can learn from the consequences of these regulations and adapt their approach for more pragmatic implementation. Established statutory regulations might also influence non-statutory regions through processes of isomorphism, potentially leading to homogenisation or divergence (Beckert, 2010).

8 A computational analysis of enforceability of the data requirement

This section examines the enforcement mechanism, especially the computational complexity entailed in enforcing the data governance requirements stipulated in the EU's AI Act (focusing mainly on Article 10 of the EU AI Act). By drawing analogies with computational NP problems, we explore the inherent challenges in meeting these requirements and the implications for compliance. The analysis highlights the evolution of the wording of the AI Act and its impact on the practical enforceability of data standards.

In computing, the enforceability of a requirement can be analogous to analysing computational complexities of nondeterministic polynomial (NP) problems. Existing technical analyses, such as those by Liza (2022), reveal the difficulties in enforcing the requirements of the 2021 draft version of the AI Act. We further explore these challenges in the context of the 2024 version of Article 10 that is currently in force.

The 2021 draft version of Article 10(2(e)) of the AI Act required datasets to be: "relevant, representative, free of errors and complete". Liza (2022) analysed the enforcement difficulties of this requirement, noting the inherent impossibilities in making sure that datasets meet strict criteria. The 2024 version of Article 10(3)

presents a modified narrative: "Training, validation and testing data sets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. Those characteristics of the data sets may be met at the level of individual data sets or at the level of a combination thereof". Given the limited scope of this chapter, let us assess the compliance complexity with the requirement that a dataset be "sufficiently representative" by formulating as follows:

Problem: Ensure that the dataset is 'sufficiently representative' in view of the intended purpose (e.g., of the target population).

Complexity: This involves selecting a subset of the data that accurately represents the distribution of the target population, which can be seen as a variant of the Set Cover problem (Lund & Yannakakis, 1994). The Set Cover problem, especially when aiming for specific statistical properties, is known to be NP-hard.

Entities claiming to be in compliance must demonstrate they have solved an NP-hard problem, which is very difficult, if not impossible. In practice, there are approximate algorithms to solve such a problem which does not guarantee an exact or correct solution. The elaborate illustration and proof of this lie beyond the scope of this chapter.

One argument is that legal rules and principles can be ambiguous, open to multiple interpretations, or not clearly applicable to every situation (Kress, 1989, Endicott, 1996). This concept suggests that laws are not always precise or definitive, leading to uncertainty in how they should be applied or interpreted in given cases. The reasonableness (MacCormick, 1998; Bongiovanni et al., 2009) standard is a valuable tool for mitigating the effects of legal indeterminacy.

Both approximate algorithms for NP-hard problems and reasonableness standards in law serve to manage complexity and provide practical solutions where exact answers are impractical or impossible. While approximate algorithms provide quantitative bounds on performance, reasonableness standards rely on qualitative judgments to achieve fair outcomes. This comparison highlights how both fields use pragmatic approaches to handle intractable issues, emphasising the importance of flexibility and practicality in decision-making processes. It will be interesting to see how transdisciplinary research progresses considering the new AI era to find a meaningful solution for a safe and secured future.

9 AI regulation, unintended consequences with Merton's framework

The intricate implications of cross-cutting challenges (CC1, CC2, CC3), interoperability between divergent regulatory frameworks, human rights implications, and both statutory and non-statutory regulations, as well as technical enforceability challenges are becoming increasingly evident. The EU AI Act aims to regulate AI use in a manner that is responsible and does not stifle innovation or infringe upon human rights. However, Robert K. Merton's concept of unintended consequences is particularly relevant for understanding the potential impacts of global regulatory approaches including the EU AI Act. Merton defined unintended consequences as outcomes that are not anticipated or intended by purposeful action (Merton, 1936). He identified several factors (e.g., Ignorance, Error, **Imperative of Immediate Interests,** Basic Values, Self-defeating Predictions) leading to these outcomes, which are highly relevant to the key parameters (timing, scope, method) of AI regulations:

A. **Ignorance**: A lack of comprehensive knowledge about the full range of disciplines involved in AI regulations (see section 5, CC3) can lead to significant unintended consequences. These include challenges in global cooperation, as well as divergent policy development and enforcement strategies. A transdisciplinary approach is critical to avoid gaps in understanding, conceptualising and implementation.
B. **Error**: Incorrect assumptions about the relationships between regulatory actions (such as the feasibility of technical standards outlined in section 5, CC1) and their outcomes can result in

compliance inconsistencies (see section 6). Further, the technical and legal infeasibility of enforcement mechanisms (see section 8) can undermine regulatory efforts.

C. **Imperative of immediate interests**: The pursuit of short-term goals, such as gaining political popularity or asserting supremacy, can overshadow the long-term effects of regulations. This can lead to human rights violations, particularly affecting marginalised groups (see section 6). It is crucial to balance immediate interests with long-term consequences to ensure ethical governance.
D. **Basic values**: Actions influenced by fundamental values, such as autonomy and ethics (see section 5, CC2), can sometimes neglect the broader consequences. For instance, while ethical considerations are vital, they must be integrated with practical technological feasibilities to avoid unrealistic regulatory expectations and to avoid encouraging loophole pursuit.
E. **Self-defeating predictions**: Actions taken to avoid predicted negative outcomes can paradoxically bring them about. For example, anthropomorphic views of AI might lead to regulatory measures that inadvertently result in human rights violations for migrants (see section 6) or induce undesired behavioural changes (see section 7). Recognising and mitigating these counterproductive outcomes is essential for effective regulation.

By initiating discussions based on these factors and expanding to include additional relevant considerations, we can better anticipate and address the unintended consequences of AI regulations. This proactive approach will help in crafting regulations that not only mitigate risks but also foster innovation and protect human rights.

10 Conclusion

This chapter aims to promote discussion and inspire for inclusive transdisciplinary collaboration to manage artificial intelligence (AI) technology to effectively mitigate unavoidable consequences – specifically the perverse result. The regulation of AI is a multifaceted challenge that demands a comprehensive, transdisciplinary approach to ensure systemic safety, fairness, and innovation for economic and social progress. As AI systems become ever more integrated into various aspects of society, the need for robust, interoperable and systemically safety promoting regulations becomes more pressing. Effective AI regulation calls for a deep understanding that spans formal, natural, and empirical sciences, social sciences, and engineering disciplines, assuring that the technology is developed and deployed responsibly.

The EU AI Act highlights the need for safe, secure, trustworthy and ethical AI, stressing requirements for explainability, transparency, accountability, and appropriate data governance. However, meeting these requirements is challenging with the current scientific capabilities. Satisfying such requirements is often made difficult due to the lack a lack of interdisciplinary comprehension among various domains. Addressing this gap through transdisciplinary approaches and collaboration is crucial for developing regulations that can effectively mitigate potential risks and unintended consequences. To achieve pragmatic and enforceable AI regulation, technocrats and policymakers must embrace a holistic view that incorporates diverse perspectives and expertise. By fostering inclusive collaboration, we can create a regulatory framework that not only addresses the technical aspects of AI but also considers the societal implications, ultimately leading to safer and more reliable AI systems.

In conclusion, in this chapter we have reflected on past unintended consequences of regulations and technology adoption. We have outlined the need for striking a balance in key parameters – timing, scope and methods – while regulating evolving AI technologies. We have categorised some key cross-cutting challenges and the unintended consequences that might have a long-term negative impact on our society and economy. The path to systemic AI safety regulations involve integrating sophisticated techniques across multiple disciplines, ensuring interoperability across borders, and implementing a pragmatic enforcement plan. Doing this will allow us to mitigate the potential risks of AI and unintended consequences of regulations, fostering an environment where AI technologies can thrive safely, responsibly, legally and ethically.

References

Abraham, J., & Davis, C. (2007). Interpellative sociology of pharmaceuticals: problems and challenges for innovation and regulation in the 21st century. *Technology Analysis & Strategic Management*, *19*(3), 387-402.

Beckert, J. (2010). Institutional isomorphism revisited: Convergence and divergence in institutional change. *Sociological theory*, *28*(2), 150-166.

Begishev, I. R., Latypova, E. Y., & Kirpichnikov, D. V. (2020). Artificial Intelligence as a Legal Category: Doctrinal Approach to Formulating a Definition. *Actual Probs. Econ. & L.*, 79.

Bongiovanni, G., Sartor, G., & Valentini, C. (Eds.). (2009). *Reasonableness and law* (Vol. 86). Springer Science & Business Media.

BTO. (2024, July 4). What does the UK general election mean for AI regulation? Retrieved July 14, 2024, from https://www.bto.co.uk/blog/what-does-the-uk-general-election-mean-for-ai-regulation.aspx

Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. *Telecommunications policy*, *44*(6), 101937.

Church, K., & Liberman, M. (2021). The future of computational linguistics: On beyond alchemy. *Frontiers in Artificial Intelligence*, *4*, 625341.

Wendehorst, C. (2020). Strict Liability for AI and other Emerging Technologies. *Journal of European Tort Law*, *11*(2), 150-180. https://doi.org/10.1515/jetl-2020-0140

Chataway, J., Tait, J., & Wield, D. (2006). The governance of agro-and pharmaceutical biotechnology innovation: public policy and industrial strategy. *Technology Analysis & Strategic Management*, *18*(2), 169-185.

Editors, TheFamousPeople.com. (2022, October 03). Robert Durst Biography. TheFamousPeople.com. Retrieved July 11, 2024, from https://www.thefamouspeople.com/profiles/robert-durst-52922.php

Endicott, T. A. (1996). Linguistic indeterminacy. *Oxford J. Legal Stud.*, *16*, 667.

Euronews. (2023, December 15). *'Potentially disastrous for innovation': Tech sector says EU AI Act goes too far*. Euronews. Retrieved July 11, 2024, from https://www.euronews.com/next/2023/12/15/potentially-disastrous-for-innovation-tech-sector-says-eu-ai-act-goes-too-far

Faulkner, A. (2009). Regulatory policy as innovation: Constructing rules of engagement for a technological zone of tissue engineering in the European Union. *Research policy*, *38*(4), 637-646.

Floridi, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. *Philosophy & Technology*, *34*(2), 215-222.

Janssen, A. (2022). AI and Contract Performance. In L. A. DiMatteo, C. Poncibò, & M. Cannarsa (Eds.), *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics* (pp. 59–73). chapter, Cambridge: Cambridge University Press.

Kress, K. (1989). Legal indeterminacy. *Calif. L. Rev.*, *77*, 283.

Liza, F. F. (2022). Challenges of Enforcing Regulations in Artificial Intelligence Act – Analyzing Quantity Requirement in Data and Data Governance.

Lund, C., & Yannakakis, M. (1994). On the hardness of approximating minimization problems. *Journal of the ACM (JACM)*, *41*(5), 960-981.

MacCormick, N. (1998). Reasonableness and objectivity. *Notre Dame L. Rev.*, *74*, 1575.

Marcin, S. (2024). *United States approach to artificial intelligence*, EPRS: European Parliamentary Research Service. Belgium. Retrieved 13th July 2024 from https://policycommons.net/artifacts/11303354/united-states-approach-to-artificial-intelligence/12188724/. CID: 20.500.12592/ksn07nm.

Muehlhauser, L., & Helm, L. (2012). Intelligence Explosion and Machine Ethics. Amnon Eden, James H. Moor, Jhonny H. Soraker, Eric Steinhart. *Singularity Hypotheses: A scientific and Philosophical Assessment*.

Ramírez, N. M. O. (2018). *Inteligencia artificial: ficción, realidad y... sueños*. Real Academia de Ingeniería.

Samoili, S., Cobo, M. L., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence.

Schuett, J. (2023). Defining the scope of AI regulations. *Law, Innovation and Technology*, *15*(1), 60-82.

Shetty, S. (2023, July 14). The EU's AI Act is barreling toward AI standards that do not exist. Lawfare. Retrieved from https://www.lawfaremedia.org/article/eus-ai-act-barreling-toward-ai-standards-do-not-exist

Stewart, R. B. (1981). Regulation, innovation, and administrative law: conceptual framework. *California Law Review, 69(5),* 1256-1377.

Toby Saunders (2023, July 9) How many continents are there in the world? It depends who you ask, BBC Science Focus. https://www.sciencefocus.com/planet-earth/how-many-continents-are-there-in-the-world

Wendehorst, C. (2020). Strict liability for AI and other emerging technologies. *Journal of European Tort Law*, *11*(2), 150-180.