
Position Paper: If Innovation in AI Systematically Violates Fundamental Rights, Is It Innovation at All?

Josu A. Eguiluz

Adevinta ServicesCo S.L.U.
Pompeu Fabra University (UPF)
Ciutat de Granada, 150, Barcelona
josu.eguiluz@adevinta.com

Axel Brando

Barcelona Supercomputing Center (BSC-CNS)
Plaça d'Eusebi Güell, 1-3, Barcelona
axel.brand@bsc.es

Migle Laukyte

Pompeu Fabra University (UPF)
Ramon Trias Fargas, 25, Barcelona
migle.laukyte@upf.edu

Marc Serra-Vidal

Kleinanzeigen.de GmbH (Adevinta)
Ciutat de Granada, 150, Barcelona
marc.serrav@adevinta.com

Abstract

Artificial intelligence (AI) now permeates critical infrastructures and decision-making systems where failures produce social, economic, and democratic harm. This position paper challenges the entrenched belief that regulation and innovation are opposites. As evidenced by analogies from aviation, pharmaceuticals, and welfare systems and recent cases of synthetic misinformation, bias and unaccountable decision-making, the absence of well-designed regulation has already created immeasurable damage. Regulation, when thoughtful and adaptive, is not a brake on innovation—it is its foundation. The present position paper examines the EU AI Act as a model of risk-based, responsibility-driven regulation that addresses the Collingridge Dilemma: acting early enough to prevent harm, yet flexibly enough to sustain innovation. Its adaptive mechanisms—regulatory sandboxes, small and medium enterprises (SMEs) support, real-world testing, fundamental rights impact assessment (FRIA)—demonstrate how regulation can accelerate responsibly, rather than delay, technological progress. The position paper summarises how governance tools transform perceived burdens into tangible advantages: legal certainty, consumer trust, and ethical competitiveness. Ultimately, the paper reframes progress: innovation and regulation advance together. By embedding transparency, impact assessments, accountability, and AI literacy into design and deployment, the EU framework defines what responsible innovation truly means—technological ambition disciplined by democratic values and fundamental rights.

1 Introduction

In recent years, Artificial Intelligence (AI) has gained remarkable traction in fields ranging from specialized research domains to becoming indispensable drivers of economic growth, social transformation, and scientific discovery. Advocates of deregulation argue that reducing legal constraints fuels rapid innovation [1, 2, 3, 4, 5]. Yet this position overlooks a crucial insight from Collingridge’s Dilemma [6]: while the need for governance remains less obvious in the early stages of a technology, once its societal consequences become apparent, the technology is often too deeply embedded to be meaningfully altered by regulation and other normative instruments. **In this position paper, we argue that regulation is not only necessary but also a fundamental enabler of innovation, particularly in critical AI-based systems.**

Drawing on historical precedents—ranging from pharmaceuticals to welfare systems or aviation—this position paper demonstrates how robust regulation has frequently fostered, rather than hindered, significant advancements. AI is not different: left unregulated, it can entrench harmful risks such as bias and discrimination, and unaccountable decision-making can have high-stake consequences. By examining the proposed EU AI Act as both an illustrative example and a case study, we explore how a risk-based approach to governance [7], combined with mechanisms like regulatory sandboxes, can unlock technical breakthroughs while safeguarding fundamental rights. In fact, neglecting sustainable governance today is not just a regulatory risk—it might be an existential threat to the future of humankind¹. Ultimately, we ask whether innovations that systematically violate human dignity and perpetuate inequities can truly be labeled “innovative,” and we urge stakeholders to view legislation not as a barrier, but as a cornerstone for AI’s long-term viability.

Paper Structure This position paper is organized into seven sections. Section 2 explains why the tension between regulation and innovation constitutes a false dichotomy, drawing on historical precedents and on the Collingridge Dilemma to show why late-stage regulation often arrives too late to prevent systemic harm. Section 3 examines the concrete risks of deregulated AI, from large-scale disinformation to bias and unaccountable decision-making in high-stakes contexts. Section 4 presents the EU AI Act as a model for risk-based, innovation-enabling governance, detailing its adaptive mechanisms—regulatory sandboxes, real-world testing, and SMEs support—and summarising their benefits in Table 1. Section 5 analyses transparency, impact assessments, accountability, and AI literacy as operational tools of responsible innovation. Section 6 engages with alternative views, addressing concerns about over-regulation and demonstrating how a well-designed framework aligns innovation with trust and legal certainty. Finally, Section 7 concludes by arguing that regulation, far from constraining progress, defines the conditions for technological ambition to endure responsibly.

2 The False Dichotomy: Regulation vs. Innovation

2.1 Historic Tensions in High-Risk Sectors

The prevailing narrative suggests that regulation hinders innovation [10, 11, 12, 13, 14, 15], creating a false dichotomy between governance and technological progress [16, 17].

Historically, the United States (USA) has led AI development, buoyed by robust free markets, eminent research institutions, and a longstanding entrepreneurial ethos [18]. Nevertheless, the new administration has revoked the Executive Order 14110 of October 30, 2023 (“Safe, Secure, and Trustworthy Development and Use of AI”) [19], emphasizing the removal of regulatory “barriers” to foster unimpeded AI innovation. This deregulatory turn is reinforced in the AI Action Plan 2025, which cautions against “smothering” AI in bureaucracy “at this early stage” and calls for “removing red tape and onerous regulation” to accelerate innovation [20]. This approach signals a commitment to accelerate AI advancements by minimizing governmental constraints², under the premise that fewer regulations will spur competitive advantage and technological breakthroughs. However, it overlooks critical lessons from other sectors where lack of regulation has led to severe societal harms. No field with major public implications has flourished without a regulatory framework.

In pharmaceuticals, for instance, the thalidomide scandal [24], where over 10,000 babies were born with severe physical abnormalities due to the drug thalidomide [25], highlighted the critical need for stringent drug regulation and clinical trial oversight. This tragedy led to the Kefauver-Harris Amendments [26] in the USA, which mandated rigorous clinical trials and oversight for new drugs, ensuring such a disaster would not occur again. These regulations have since formed the foundation of modern pharmaceutical safety protocols.

¹*Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war:*” [8] The 2025 UN “Red Lines on AI” initiative [9] echoed this call by urging global ethical boundaries to prevent catastrophic risks from high-impact AI systems.

²Despite this federal deregulatory vision, several U.S. states—most notably California—have begun to adopt AI-specific legislation in recognition of the societal importance of the technology. For instance, California’s SB 53 [21] requires certain obligations for frontier models, while SB 243 [22] regulates AI companions chatbots, including disclosure and safety protocols for interactions with minors. For a state-by-state overview, see the IAPP’s US State AI Governance Legislation Tracker [23].

Additionally, the Dutch SyRI scandal [27, 28] exemplifies the dangers of opaque AI-driven decision-making in welfare systems. In 2020, a court [29] ruled that the automated fraud detection system violated human rights due to its lack of transparency and disproportionate targeting of low-income communities, ultimately leading to the government’s resignation. This case underscores the risks of unregulated AI in public administration, where biased algorithms can entrench systemic discrimination and erode trust in institutions.

Similarly, the aviation industry once operated with far higher accident rates; in 1959, a passenger in the USA or Canada faced a 1-in-25,000 chance of a fatal crash on each departure. However, with the implementation of increasingly stringent safety regulations, aviation safety improved dramatically. The Federal Aviation Act of 1958 established the Federal Aviation Administration (FAA) [30], centralizing safety oversight and enforcement in the USA. In the 1980s, the introduction of Advisory Circular AC 25.1309-1 [31, 32] provided systematic guidelines for aircraft system safety assessments, further reinforcing risk mitigation measures. Over time, these regulatory developments, along with international safety protocols [33], contributed to an over 1,000-fold improvement in aviation safety. In the last years, the odds of a fatal accident in the USA or EU have dropped to 1 in 29 million [34, 35, 36], demonstrating how robust oversight not only mitigates catastrophic risks [37] but also fosters public confidence and technological innovation, as seen in modern aviation advancements.

A failed or non-existent regulation can stifle progress, among other things, but what these examples reflect are design flaws in the regulatory frameworks (or their absence), rather than an inherent tension between rules and advancement. Effective regulation must be thoughtfully designed to support innovation while safeguarding fundamental rights, ensuring that AI serves as a transformative force [38] for good [39] without compromising or even improving ethical standards. While the nature of harms differs—physical in aviation and pharmaceuticals, systemic in AI—the rationale for regulation remains: to prevent irreparable consequences before they scale, ensure the security of investment, create public trust, etc.

By doing so, a shared understanding of what constitutes valid innovation is essential; this is what the following Section is dedicated to.

2.2 Defining Responsible Innovation

Traditionally, innovation has been measured by economic growth and technological novelty, often with scant attention to its social or ethical ramifications. As Smith [40] notes, the conceptual foundations of innovation stem from both management and economics, where innovation has long been seen as a driver of competitiveness and productivity. Building on Schumpeter’s [41] concept of *creative destruction*—the process of innovation that replaces obsolete technologies and transforms the economic structure—, formalised by Aghion and Howitt [42], the 2025 Nobel Prize in Economics reaffirmed that sustained prosperity depends on openness to knowledge and technological renewal [43].

However, the landscape of innovation has evolved to incorporate broader considerations. The *Oslo Manual* [44] has expanded its guidelines to highlight sustainability and societal well-being as core aspects of modern innovation. This shift acknowledges that true innovation must not only advance technology and economic performance but also contribute positively to society and the environment.

In this regard, the OECD has enshrined a novel concept: responsible innovation, which means “*a trustworthy technology development guided by democratic values, responsive to social needs and accountable to society. Adopting a responsible innovation approach in the development of emerging technologies can help align research and commercialisation with societal needs*” [45].

Interestingly, the revoked Executive Order 14110 [46] also emphasized the promotion of responsible innovation. It’s Section 2. b) stated: “*Promoting responsible innovation, competition, and collaboration will allow the USA to lead in AI and unlock the technology’s potential to solve some of society’s most difficult challenges...*”. Nevertheless, its revocation has signaled a shift away from these principles, potentially deprioritizing the alignment of AI [47] innovation with ethical and societal values.

In the EU context, the EU AI Act [48] embodies the principles of responsible innovation by integrating regulatory oversight as a means to ensure that AI development aligns with ethical standards and societal values. The Recital 138 of the EU AI Act acknowledges that “*AI is a rapidly developing family of technologies that necessitates both a regulated environment and safe spaces for experimentation,*

while ensuring responsible innovation and integration of appropriate safeguards and risk mitigation measures...

This responsible innovation framework upholds that rigorous governance can coexist with, and even enhance, innovation by ensuring that technological advancements do not come at the expense of societal well-being. By treating regulation as inherent to innovation, this approach demonstrates that effective governance is not a barrier but an integral part of fostering sustainable and ethical technological progress.

Consequently, in this position paper, innovation is inherently synonymous with responsible innovation, thereby rejecting the notion of innovation devoid of accountability and ethical consideration.

2.3 Collingridge Dilemma

To address the issue of responsible innovation, the timely introduction of regulation plays a crucial role. The tension between early and late regulation has been characterized by Collingridge as a dilemma [6]: in the early stages of technology, it is difficult to foresee all risks, yet once the technology is deeply entrenched, change becomes exceedingly costly or even impossible. In AI, this is exacerbated by continuous deployment in high-stakes sectors and critical infrastructures where undesirable consequences may only be recognized after significant harm has occurred.

Collingridge’s Dilemma highlights the challenge of regulating technology at the right time: *“when change is easy, the need for it cannot be foreseen; when the need for change is apparent, change has become expensive, difficult and time consuming”* [6]. Although this captures the epistemological and societal hurdles of anticipating future harms, Collingridge does not claim the dilemma is insurmountable [49]. He advocates for proactive governance mechanisms that adapt to new evidence—a perspective highly relevant to AI, where deployment in high-stakes sectors can embed problematic issues before they are fully recognized. This dilemma is not a static, universal constant; rather, it emerges from human decisions and institutional frameworks³, suggesting that deliberate regulatory strategies can and should be employed to address the inherent risks of this technology.

3 The Risks of Deregulated AI

While some critiques frame AI regulation as a precautionary response to hypothetical risks, there is growing empirical evidence that insufficient oversight has already enabled significant and recurring harm across multiple domains [51]. During the 2023 Slovak elections, deepfake audio mimicking a candidate went viral, undermining trust in the democratic process [52]. AI enables large-scale political manipulation via synthetic media, voice cloning, and micro-targeted persuasion [53].

In parallel, generative AI has enabled the spread of synthetic non-consensual sexual content, including deepfake pornography and child sexual abuse material (CSAM) [54]. Notable cases include Spain’s “Almendralejo case” involving minors [55] or AI-generated explicit images of Taylor Swift [56]. These cases highlight the urgent need for enforceable safeguards [57] to prevent reputational, emotional, and psychological harm.

Beyond information harms, courts and regulators have sanctioned opaque AI systems in housing, credit, welfare, and education for violating due process and fundamental rights [58]. Further risks have emerged in healthcare, financial fraud, and autonomous systems—often amplified by regulatory gaps [59]. Complementing this picture, a comparative review confirms that vulnerable groups are disproportionately affected across high-stakes domains [60].

These examples show that AI-related harms are neither speculative nor isolated—they are real, ongoing, and escalating. The next sections examine two of the most urgent manifestations of these risks: bias and discrimination, and unaccountable decision-making in high-impact domains.

3.1 Bias and Discrimination

Recent discussions, particularly in USA, have emphasized deregulation as a means to accelerate AI development [61]. However, even with existing regulations, AI presents significant risks, notably the perpetuation and amplification of societal biases [62], especially when affecting high-stakes domains.

³As Carissa Véliz reminds us, “AI is not God-given, we are designing it. We can do better.”[50]

AI systems frequently learn from historical data imbued with racial [63, 64], gender [65], or socioeconomic biases [66], thereby reinforcing existing disparities [67]. The lack of stringent regulations mandating transparent reporting and fairness audits allows these biases to remain undetected and unaddressed, potentially scaling injustices across entire populations [68].

Biases are not purely algorithmic: human and machine decisions often reinforce each other [69], creating compounded discrimination [59]. Such dynamics highlight the complexity of mitigating discrimination, necessitating not only the identification and correction of algorithmic biases but also addressing human factors that interact with AI systems [69].

Effectively mitigating these challenges requires robust strategies bolstered by specialized tools and frameworks. A comprehensive literature review identifies 152 concrete measures for addressing and reducing biases in AI systems, encompassing interventions such as bias detection algorithms, the establishment of diverse and inclusive datasets, and rigorous auditing processes [70]. By employing these tools, organizations can systematically identify and rectify discriminatory patterns, thereby fostering fairer and more equitable AI-driven outcomes.

3.2 Unaccountable Decision-Making in High-Risk Scenarios

When AI-based decisions lack regulatory oversight, affected individuals are unable to effectively challenge or seek redress for resulting harms [71]. The absence of humans in the loop is worrisome. But integrating human overseers without adequate training, competence or authority could be even worse, creating a false sense of legitimacy and reducing oversight to a token gesture [72, 73]. Furthermore, oversight systems frequently lack clear guidelines on when to override AI decisions, how to incorporate additional relevant information, or how to systematically audit the combined AI-human decision-making processes for cumulative biases.

This deficiency leads to “unaccountable decision-making”, wherein neither the human operators nor the algorithms bear full responsibility for the outcomes. Such unaccountability is particularly detrimental in high-stakes scenarios, including medical diagnoses and criminal proceedings, where the concealment of biases within opaque models or data pipelines results in unjust consequences for individuals, who are then deprived of fair and transparent appeals mechanisms [74]. Effective oversight therefore requires not only technical safeguards but also structured governance mechanisms built on institutionalised distrust [75] and inclusive review processes, engaging diverse stakeholders across the AI lifecycle—such as operators, auditors, end-users, and affected communities [76].

Additionally, the phenomenon of moral outsourcing [77, 78] exacerbates this issue by transferring ethical responsibility from human decision-makers to AI systems. This occurs through the attribution of agency to AI ⁴, allowing both developers and deploying institutions to deflect accountability for negative outcomes onto the algorithms themselves. As a result, technical advancements are pursued as inevitable progress, while adverse effects are reframed as neutral byproducts of complex systems. This linguistic and conceptual shift places the burden of adverse outcomes on the technology itself, enabling the continued deployment of profitable AI products while perpetuating the illusion of mathematical objectivity, neutrality and accuracy [80].

Real-world examples, such as predictive policing software that utilizes racially skewed datasets [81, 82, 83], opaque credit-scoring algorithms that disadvantage low-income applicants or under-represented minorities [84], or automated eligibility assessments in welfare and healthcare [85], illustrate the potential for unaccountable systems and irresponsible actors to inflict irreversible harm. These systems often lack auditability, transparency, or mechanisms for individual redress, leading to decisions that are difficult to contest or revise. Once embedded into institutional infrastructures, they become resistant to scrutiny or structural change, embodying Collingridge Dilemma of being “too late to fix”. However, regulations like the EU AI Act seek to address these risks proactively, ensuring that harm can be mitigated before it becomes irreversible and expands.

⁴A scenario that could be aggravated by the emergence of AI agents. AI agents are autonomous systems capable of sensing, learning and acting upon their environments [79].

4 The EU AI Act: Perceived Burdens vs. Actual Benefits

4.1 Overview of key provisions

The proposed EU AI Act introduces a risk-based classification for AI systems [86, 87], stipulating stricter requirements for 'high-risk' systems in fields like employment, healthcare, and education ⁵.

While some critics perceive these constraints as innovation-blocking [14], the EU AI Act explicitly acknowledges the need to foster responsible AI innovation, offering mechanisms such as regulatory sandboxes and support measures for SMEs and start-ups. These tools aim to accelerate compliance through a structured and supervised environment. This section aims to debunk the myth that regulation is an impediment to innovation [88]. On the contrary, it could become a competitive advantage.

4.2 Regulatory sandboxes and real-world testing

Regulatory sandboxes [89, 90, 91, 92] —a concept well established in fintech ⁶—will be mandatory in each EU Member State by August 2026, according to the Art. 57 AI Act. They provide a controlled environment where developers can train, test, and validate AI systems under regulatory supervision before the AI systems are placed on the market or otherwise put into service. Spain has already launched its first national sandbox for high-risk AI systems, offering practical insights into early implementation under real-world conditions [94].

These are not deregulated zones, but co-regulatory spaces where legal and technical innovation align to reduce uncertainty and uphold rights. This approach yields several benefits:

- **Legal certainty through structured guidance:** Competent authorities provide iterative regulatory mentorship that helps participants identify and mitigate risks, clarify compliance expectations, and generate documentation that accelerates conformity assessments.
- **Risk mitigation:** If significant health, safety, or rights violations arise, the sandbox environment allows authorities to temporarily suspend or even end experimentation.
- **No fines:** As long as companies follow the sandbox plan and act in good faith, no administrative fines apply for AI Act infringements during sandbox testing.
- **Cross-sectoral collaboration:** The sandbox fosters meaningful interaction between public institutions, private companies, and academia, reinforcing shared understanding and co-responsibility in the interpretation of AI obligations.
- **Mutual Learning:** Different actors co-construct applicable guidelines based on real evidence and case-specific interaction.

Outside these sandboxes, providers of specific high-risk AI systems—like biometrics, education, worker management—may conduct real-world testing (Art. 60 AI Act) under conditions ensuring informed consent, robust oversight, and the right to withdraw without penalty. Notably, the market surveillance authority may veto or stop tests if unaddressed risks emerge.

4.3 Support measures for SMEs and start-ups

Given the potentially steep costs of compliance for smaller developers, the Art. 62 of the AI Act prescribes four main forms of support:

- **Priority access to sandboxes:** SMEs that have either a registered office or a branch established within the EU receive preferential treatment for sandbox participation, facilitating early compliance, removing barriers and faster market. This principle is also reflected in Spain's national sandbox, where Article 8(j) of the Royal Decree explicitly favors start-ups and SMEs in the selection process [95];

⁵According to the EU AI Act, this category comprises a strictly circumscribed subset of AI systems whose deployment or use may adversely affect individuals' safety, health, or fundamental rights.

⁶Currently, 57 countries operate 73 fintech sandboxes [93].

- **Awareness and training:** Member States must provide targeted outreach, such as accessible information resources and tailored training sessions on the AI Act ⁷, to raise awareness among smaller businesses and support regulatory implementation;
- **Dedicated communication channels:** National authorities should promptly address SMEs inquiries to minimize confusion regarding compliance procedures. Where appropriate, they may also establish dedicated communication channels to enhance clarity and support. Thus, the Commission has launched the *AI Act Service Desk* and the *AI Act Single Information Platform*, which include tools such as a Compliance Checker, an AI Act Explorer, and direct channels for expert guidance [98];
- **Facilitating standardization:** SMEs should be encouraged to join standard-setting processes [99, 100, 101], ensuring that industry norms align with real-world developer challenges.

Furthermore, microenterprises (fewer than 10 employees, less than €2 million turnover) can implement a simplified version of the quality management system required for high-risk AI, minimizing administrative burdens without compromising user safety, as stated in the Art. 63 AI Act. Lastly, the Commission will provide standardized templates for the areas covered by the regulation and should work with Member States to lower red tape and compliance costs [102].

4.4 Competitive advantages of regulation

Contrary to perceptions of regulatory overreach, structured regulation can bolster consumer trust and market stability [16].

Historical evidence from data protection (e.g., General Data Protection Regulation (GDPR) [103]) shows that companies adhering to clear legal frameworks often reap reputational benefits and avoid hefty compliance [104]. While initial compliance may disrupts operations and increases costs, it ultimately enhances privacy awareness, modernizes IT infrastructure, and improves risk management processes [104]. Notably, GDPR catalyzed the development of privacy-enhancing technologies (PETs), including differential privacy and federated learning [105, 106], both embedding fundamental rights into system design and reinforcing accountability and transparency [107, 63].

A similar dynamic is now unfolding under AI regulation. Legal obligations are accelerating the adoption of watermarking technologies to identify AI-generated content—including synthetic deepfakes—[108, 109, 110]. In parallel, developers are being required to implement state-of-the-art technical measures to safeguard copyright—particularly during the training of general-purpose AI models—[111] ⁸. These compliance-driven innovations not only meet regulatory thresholds but also function as trust-building mechanisms that differentiate responsible actors.

Beyond technical innovation, a distinct yet complementary strategy lies in embedding ethical principles at the core of a product’s identity. A pertinent example is Apple Inc., where CEO Tim Cook strategically positioned user data privacy as both a business strategy and an ethical imperative, thereby transforming privacy into a market differentiator [113, 16].

In light of these examples, by proactively ensuring AI meets ethical and legal standards, organizations could promote responsible innovation, while reducing the risk of reputation damage, costly litigation, and opposition from civil society. As Lucilla Sioli, head of the European Commission’s AI Office, aptly states, “*You need the regulation to create trust, and that trust will stimulate innovation*” [114].

The strategic benefits of early compliance are well illustrated by two complementary insights: the Porter hypothesis, which holds that regulation can stimulate firms to innovate in ways that produce both societal and market benefits [115, 116]; and the already discussed Collingridge Dilemma, which highlights the tension between limited foresight at early stages and reduced flexibility once technologies become entrenched [6]. While the dilemma does not prescribe a solution, it underscores the cost of inaction. In this context, proactive regulatory strategies—anchored in early engagement—can help mitigate long-term risks, reinforce institutional trust, and sustain the long-term viability of the AI

⁷This includes support through initiatives like the EU’s Living Repository on AI Literacy and its accompanying FAQ guidance [96, 97]

⁸Complying with Intellectual Property (IP) law has become one of the most pressing legal and operational challenges for the AI industry, as illustrated by the growing wave of litigation concerning training data and copyright infringement. See: Updated Map of US Copyright Suits v. AI Companies [112].

Perceived Burdens vs. Actual Benefits of the EU AI Act	
Sandboxes (Regulatory Testing & Experimentation)	
<i>Regulation blocks innovation</i>	Regulatory Sandboxes (Art. 57) – Safe environments for AI testing.
<i>Limited flexibility for developers</i>	Real-world testing (Art. 60) – Controlled testing under oversight.
Support for SMEs & Startups	
<i>High compliance costs</i>	SME support (Art. 62) – Priority access, training, and cost reduction.
<i>Complex compliance for SMEs</i>	Standardized templates (Art. 62) – Simplifies compliance process.
<i>Excessive bureaucracy</i>	Dedicated support (Art. 62) – Direct communication channels.
General Measures & Competitive Advantages	
<i>Slows AI adoption</i>	Consumer trust – Clear rules prevent legal and market risks.
<i>Limits technological progress</i>	Compliance-driven innovation – Rights-based technologies by design.
<i>Stifles market growth</i>	Legal certainty – Ensures stability and reduces compliance costs.
<i>Competitive disadvantage</i>	Ethical leadership – EU compliance signals responsible AI.
<i>Short-term burden</i>	Long-term advantage – Reduces legal, reputational, and market risks.

Table 1: Perceived Burdens vs. Actual Benefits of the EU AI Act: Categorized by Key Mechanisms

sector, which ultimately depends on its ability to align with ethical and regulatory imperatives by design [117]. Firms that anticipate regulatory expectations may also capture a first-mover advantage, shape emerging markets, and convert compliance into strategic differentiation [16].

History reinforces this logic. Industries that internalized strong safety regulations—such as medical devices or chemical manufacturing—have not only avoided crises but also secured sustained public trust and long-term leadership. Like those sectors, AI presents distinct risks when applied in sensitive domains such as healthcare, public services, and education. It must therefore follow a comparable regulatory trajectory: short-term gains from unregulated deployment may appear attractive, but risk triggering systemic failures, undermining trust, and eroding long-term viability.

Far from impeding commercial viability, the EU AI Act introduces targeted incentives—such as regulatory sandboxes, real-world testing, and SMEs support—that foster responsible experimentation under supervised conditions. These mechanisms not only facilitate compliance but also reduce deployment risks and accelerate market readiness.

Table 1 provides a structured comparison of common perceived burdens and their corresponding advantages under the EU AI Act. While not exhaustive, it highlights key instruments outlined above that exemplify how regulation can transform constraints into drivers of responsible AI innovation.

5 Transparency, Impact Assessments, Accountability and AI Literacy

Beyond incentives and support structures, regulation also defines how innovation is practiced. By translating legal obligations into actionable requirements and measurable design routines, governance tools such as transparency, impact assessments, accountability mechanisms, and AI literacy operationalize regulation into continuous feedback and institutional learning.

5.1 Transparency

In a democratic context, the legitimacy of any decision-making process depends on transparency⁹. The EU AI Act embodies this principle through structured documentation duties (Arts. 11, 13, and 50), ensuring that deployers, operators, and affected individuals can scrutinize AI systems, contest outcomes, and seek effective redress while aligning with fundamental rights [119, 120, 67]. Building on this approach, the *Code of Practice on General-Purpose AI*—endorsed by most major technology

⁹ “With AI decision-making and its perceived legitimacy, an important question to ask is not whether we should have transparency, but rather which kind of transparency should be applied.” [118]

companies except Meta—extends these principles by promoting voluntary yet structured transparency commitments for general purpose AI models (Arts. 53 and 55) [121]. Transparency—understood not merely as the provision of well-structured documentation but as the capacity to comprehend, interpret, and scrutinize the functioning of AI models—constitutes a cornerstone of trustworthy AI [122, 123, 124].

This principle extends beyond EU borders, as reaffirmed by the Council of Europe’s Framework Convention on AI, Human Rights, Democracy and the Rule of Law [125], which mandates context-sensitive transparency and oversight across the AI lifecycle. Recent proposals advance this toward meaningful transparency—information accessible and interpretable by citizens and regulators alike [126]. In this sense, transparency reduces the opacity¹⁰, that breeds distrust, transforming compliance into a shared basis for accountability and innovation.

5.2 Impact Assessments

The integration of impact assessments, such as the Fundamental Rights Impact Assessment (FRIA), represents a regulatory innovation that shifts AI governance from reactive compliance to anticipatory design. Under Article 27 of the EU AI Act, the FRIA operationalizes the protection of fundamental rights [128] through ex ante evaluation of potential interferences. Rather than a procedural checklist, it constitutes a substantive, context-sensitive process grounded in legal and ethical expertise [129, 130].

Emerging scholarship proposes hybrid frameworks that assess the severity, duration, and reversibility of rights interferences [131]. Through supervisory oversight, the FRIA embeds proportionality, stakeholder participation, and institutional transparency, transforming abstract legal rights into measurable criteria and accountability mechanisms. Practical implementations [132, 133, 134, 135, 136], show how iterative, context-aware evaluation and continuous feedback loops sustain regulatory learning—evidence that adaptive governance itself can drive responsible innovation.

5.3 Accountability Mechanisms

Accountability mechanisms address the structural gaps that render AI decisions unaccountable. Under Article 14 of the EU AI Act, human oversight must be exercised by competent, authorised individuals within institutional architectures that ensure traceability and reversibility. Moving beyond symbolic supervision, effective oversight relies on structured distrust—a governance principle that embeds verification and challenge throughout the AI lifecycle [137, 106, 75].

Human oversight entails trained discretion, the ability to interrupt or override automated operations, and procedures for detecting and responding to adverse outcomes [138]. Yet, the deeper risk is not full automation but induced irreflexivity [139]: automated systems press humans to decide without questioning authorship or consequence, eroding the reflective judgement. Effective governance must therefore preserve human spaces for reflexivity, to interrogate, contest and revise algorithmic outputs. When implemented through layered responsibility—linking providers, deployers, auditors and affected stakeholders—these mechanisms ensure that accountability remains both operational and reflective, sustaining trustworthy and rights-aligned innovation.

Additionally, the AI Act also foresees enforceable remedies: affected persons may file complaints with market-surveillance authorities and obtain clear, meaningful explanations for decisions that significantly impact their rights [140, 141]. Whistleblower protections further strengthen this ecosystem. Together, these measures ensure that accountability is not only procedural but actionable—closing the loop from oversight to redress.

5.4 AI literacy

AI literacy functions as the connective tissue of AI governance. Beyond technical fluency, it enables institutions and individuals to develop and use AI systems, while navigating complex regulatory landscapes and ethical dilemmas [142]. Defined in Article 3(56) of the EU AI Act, literacy encompasses the skills, knowledge, and understanding required to deploy and scrutinize AI responsibly.

¹⁰AI models can identify complex correlations and make predictions based on vast numbers of interacting parameters, often rendering their decision-making processes opaque, even to experts. This “black box” effect can obscure the rationale behind AI-generated decisions, potentially leading to misplaced trust or over-reliance, with adverse consequences for individuals. [127]

Regular, practice-oriented training [143]—particularly in high-risk domains—ensures that oversight remains meaningful rather than symbolic or ineffective. Through initiatives such as the EU Living Repository on AI Literacy [96], these programs transform awareness into operational capacity, allowing teams to conduct risk assessments and engage with feedback mechanisms.

Ultimately, literacy sustains the reflexive dimension of governance: it turns transparency into understanding, impact assessment into learning, and accountability into institutional memory. In this sense, AI literacy is not peripheral but constitutive of responsible innovation.

6 Alternative Views

Some scholars argue that regulation constrains innovation, especially in fast-moving fields like AI. The Draghi report [10] warns that excessive regulation may hinder EU digital competitiveness. Ridley [11] and Braben [12] contend that scientific breakthroughs require freedom from bureaucratic and ethical constraints, and caution against premature intervention when risks are still unclear. Castro et al. [13] argue that anxiety-driven regulation risks delaying adoption, constraining beneficial applications of AI, and undermining innovation itself, calling instead for optimism-driven and targeted governance frameworks. Specific critiques of the EU AI Act include Gikay’s argument that its rigid risk classification fails to balance innovation and risk [14], and Wachter’s [144] identification of legal loopholes that may undermine enforcement and clarity.

While we acknowledge these concerns, we argue that, as documented in Section 3, the absence of well-designed regulation has already triggered severe harms: the viral spread of synthetic misinformation, bias and discrimination embedded in data-driven systems, and unaccountable decision-making in high-stakes contexts such as healthcare and welfare. These are not speculative threats but concrete evidence of governance failures.

The risk-based approach adopted by the EU AI Act is therefore the appropriate regulatory model: it differentiates obligations according to risk categorization while preserving the conditions for innovation. Far from stifling innovation, it enables it through adaptive, risk-proportionate mechanisms that convert perceived burdens into competitive advantages (see table 1). These instruments foster consumer trust, legal certainty, and ethical leadership, transforming the deregulation race into a driver of competitive trust.

The governance instruments analysed in Section 4 and Section 5 operationalise this balance, showing that regulation and innovation advance hand in hand. By providing the structure and accountability that make progress sustainable, the EU’s framework turns regulation into the very condition of responsible innovation.

7 Conclusion

This position paper has challenged the assumption that every technological advancement in AI automatically constitutes innovation. That belief ignores the extensive empirical and historical societal costs that certain AI systems have already imposed—from biased models and synthetic misinformation to unaccountable decision-making. Calling such systems “innovative” is, at best, questionable and, at worst, contradictory.

As AI becomes embedded in essential infrastructures and social decision-making, neglecting regulation in the name of “progress” invites the very irreversible harms that Collingridge foresaw in his Dilemma. Well-designed governance is not a brake on innovation but its catalyst—the mechanism through which technological ambition becomes sustainable, accountable, and aligned with human values. The EU AI Act exemplifies this principle: its risk-based, adaptive framework couples legal certainty with flexibility, through mechanisms such as regulatory sandboxes, small and medium enterprises (SME) support, real-world testing, fundamental rights impact assessment (FRIA), which operationalises responsible innovation.

Ultimately, innovation and regulation advance together. The future of AI will not be defined by the speed of invention but by the integrity of its governance. Regulation—when well-designed, proportionate, and rooted in fundamental rights—is the discipline that allows innovation to endure. Indeed, if AI systematically violates rights, can we truly call it innovation?

Acknowledgements

We gratefully acknowledge Adevinta for sponsoring the industrial PhD. We also thank the Government of Catalonia's Industrial PhDs Plan for funding part of this research. The research leading to these results has received funding from the Horizon Europe Programme under the Horizon Europe Programme under the AI4DEBUNK Project (<https://www.ai4debunk.eu>), grant agreement num. 101135757. Additionally, this work has been partially supported by JDC2022-050313-I funded by MCIN/AEI/10.13039/501100011033al by European Union NextGenerationEU/PRTR.

References

- [1] AP News. JD Vance warns that overregulation may hurt AI progress, 2025. Available at: <https://apnews.com/article/1d7826affdcdb76c580c0558af8d68d2>. Accessed October 2025.
- [2] Times of India. Bosch CEO to Europe: You are unnecessarily delaying AI's future with overregulation, 2025. Available at: <https://timesofindia.indiatimes.com/technology/tech-news/bosch-ceo-to-europe-you-are-unnecessarily-delaying-its-ai-future-with/articleshow/122072769.cms>. Accessed October 2025.
- [3] MIT Sloan. Does regulation hurt innovation? A study says yes, 2023. Available at: <https://mitsloan.mit.edu/ideas-made-to-matter/does-regulation-hurt-innovation-study-says-yes>. Accessed October 2025.
- [4] Andrew McAfee. EU Proposals to Regulate AI Are Only Going to Hinder Innovation. *Financial Times*, 2021. Available at: <https://www.ft.com/content/a5970b6c-e731-45a7-b75b-721e90e32e1c>. Accessed October 2025.
- [5] Angus Loten. Corporate Tech Leaders Are Mixed on EU Artificial Intelligence Bill. *Wall Street Journal*, 2021. Available at: <https://www.wsj.com/articles/corporate-tech-leaders-are-mixed-on-eu-artificial-intelligence-bill-11619049736>. Accessed October 2025.
- [6] David Collingridge. *The Social Control of Technology*. Palgrave Macmillan, 1980.
- [7] Giovanni De Gregorio and Peter Dunn. The european risk-based approaches: Connecting constitutional dots in the digital age. *Common Market Law Review*, 59(2):473–500, 2022.
- [8] Center for AI Safety. Mitigating the risk of extinction from ai should be a global priority alongside other societal-scale risks such as pandemics and nuclear war, 2023.
- [9] Red Lines on AI. Red lines on ai: A global call for ethical boundaries in artificial intelligence, 2025. United Nations Initiative on AI Governance.
- [10] M. Draghi. Report on the future of european competitiveness, 2024.
- [11] M. Ridley. *How Innovation Works: And Why It Flourishes in Freedom*. Harper, 2020.
- [12] D. Braben. *Scientific Freedom: The Elixir of Civilization*. Wiley, 2004.
- [13] Daniel Castro and Michael McLaughlin. Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence, 2019. Information Technology and Innovation Foundation. Available at: <https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence>. Accessed October 2025.
- [14] Asress Adimi Gikay. Risks, innovation, and adaptability in the uk's incrementalism versus the european union's comprehensive artificial intelligence regulation. *International Journal of Law and Information Technology*, 32:eaae013, 2024.
- [15] Josh Withrow. Don't stifle u.s. tech innovation with europe's rules. R Street Institute, October 9 2022.

- [16] Anu Bradford. The false choice between digital regulation and innovation. *Northwestern University Law Review*, 118(2), October 6 2024. Available at: <https://dx.doi.org/10.2139/ssrn.4753107>. Accessed October 2025.
- [17] Jacques Pelkmans and Andrea Renda. Does eu regulation hinder or stimulate innovation? CEPS Special Report 96, Centre for European Policy Studies (CEPS), November 2014. Available at SSRN: <https://ssrn.com/abstract=2528409>.
- [18] Daron Acemoğlu and Simon Johnson. *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity*. Basic Books, 2023.
- [19] President_of_USA. Removing barriers to american leadership in artificial intelligence. Executive Order, January 23 2025.
- [20] The White House. America’s ai action plan, 2025. Office of Science and Technology Policy, Executive Office of the President.
- [21] California Legislature. Senate bill no. 53, artificial intelligence models: Large developers. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=20250260SB53, 2025. Chaptered on September 29, 2025.
- [22] California Legislature. Senate bill no. 243, companion chatbots. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=20250260SB243, 2025. Chaptered on October 13, 2025.
- [23] International Association of Privacy Professionals. U.s. state ai governance legislation tracker, 2025. Accessed October 2025.
- [24] James Kingsland. How the thalidomide scandal led to safer drugs. *Medical News Today*, 2020.
- [25] Neil Vargesson. Thalidomide-induced teratogenesis: History and mechanisms. *Journal of Developmental Biology*, 2015.
- [26] United States Congress. Public law 87-781 - drug amendments of 1962, October 10 1962.
- [27] Aristi Rachovitsa and Nathalie Johann. The human rights implications of the use of ai in the digital welfare state: Lessons learned from the dutch syri case. *Human Rights Law Review*, 22(2), 2022.
- [28] Marvin van Bekkum and Frederik Zuiderveen Borgesius. Digital welfare fraud detection and the dutch syri judgment. *European Journal of Law and Technology*, 23(4), 2021. First published online August 2, 2021.
- [29] NJCM et al. v The Dutch State. The hague district court ecli:nl:rbdha:2020:1878 (syri). The Hague District Court, 2020. English translation available at: <http://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:1878>. Accesed October 2025.
- [30] Federal Aviation Administration. The federal aviation act of 1958, 1958.
- [31] Federal Aviation Administration. Advisory circular 25.1309-1a - system design and analysis, 1988.
- [32] Federal Aviation Administration. Advisory circular 25.1309-1b - system design and analysis, 2024.
- [33] International Civil Aviation Organization (ICAO). Convention on international civil aviation, December 7, 1944.
- [34] Allianz. Global aviation safety study, 2014.
- [35] International Civil Aviation Organization (ICAO). Safety report, 2024.
- [36] Esteban Ortiz-Ospina. Commercial flights have become significantly safer in recent decades. *Our World in Data*, 2024.

- [37] D.K. Yadav and H. Nikraz. Implications of evolving civil aviation safety regulations on the safety outcomes of air transport industry and airports. *Aviation*, 18:94–103, 2014.
- [38] Ross Gruetzemacher and Jess Whittlestone. The transformative potential of artificial intelligence. *Futures*, 135, 2022.
- [39] AI for Good Foundation. Ai for good, 2025. Available at: <https://ai4good.org/>. Accessed October 2025.
- [40] Keith Smith. Measuring innovation. In Jan Fagerberg, David C. Mowery, and Richard R. Nelson, editors, *The Oxford Handbook of Innovation*, pages 148–178. Oxford University Press, Oxford, 2006.
- [41] Joseph A. Schumpeter. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Harvard University Press, Cambridge, MA, 1934.
- [42] Philippe Aghion and Peter Howitt. A model of growth through creative destruction. *Econometrica*, 60(2):323–351, 1992.
- [43] Royal Swedish Academy of Sciences. Scientific background: The sveriges riksbank prize in economic sciences in memory of alfred nobel 2025. <https://www.nobelprize.org/prizes/economic-sciences/2025/press-release/>. Accessed October 2025.
- [44] OECD/Eurostat. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition*. OECD Publishing, Paris/Eurostat, Luxembourg, 2018.
- [45] OECD Science, Technology and Industry Working Papers. Responsible innovation in neurotechnology enterprises. October 11 2019.
- [46] Biden-Harris Administration. Fact sheet: Biden-harris administration announces key ai actions following president biden’s landmark executive order, 2024. Available at: www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/. Accessed October 2025.
- [47] Brian Christian. *The Alignment Problem: Machine Learning and Human Values*. W. W. Norton & Company, 2020.
- [48] European Parliament and Council. Regulation (eu) 2024/1689 on artificial intelligence (artificial intelligence act). June 13 2024.
- [49] Wolfgang Liebert and Jan C. Schmidt. Collingridge’s dilemma and technoscience. *Poiesis and Praxis*, 7(1-2):55–71, 2010.
- [50] Carissa Véliz. How privacy can save your life. YouTube. Available at: www.youtube.com/watch?v=xSPRouBvgFE. Accessed October 2025.
- [51] Partnership on AI. Ai incident database. <https://incidentdatabase.ai/>, 2025. Accessed October 2025.
- [52] Morgan Meaker. Slovakia’s election deepfakes show ai is a danger to democracy. <https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/>, 2023. Accessed October 2025.
- [53] European Parliamentary Research Service. Artificial intelligence, democracy and elections. Technical Report EPRS_BRI(2023)751478, European Parliament, 2023. Briefing Paper.
- [54] Bracket Foundation and Value for Good GmbH in cooperation with the Centre for Artificial Intelligence and Robotics at the United Nations Interregional Crime and Justice Research Institute (UNICRI). Generative ai: A new threat for online child sexual exploitation and abuse. Technical report, UNICRI, 2024. Accessed October 2025.
- [55] BBC News. Ai-generated naked child images shock spanish town of alمندralejo. <https://www.bbc.com/news/world-europe-66877718>, 2023. Accessed October 2025.

- [56] Emine Saner. Inside the taylor swift deepfake scandal: 'it's men telling a powerful woman to get back in her box'. <https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box>, 2024. Accessed October 2025.
- [57] Beatriz Kira. Deepfakes, the weaponisation of ai against women and possible solutions, 2024. Published on Verfassungsblog, available at: <https://verfassungsblog.de/deepfakes-ncid-ai-regulation/>. Accessed October 2025.
- [58] Sebastião Barros Vale and Gabriela Zanfir-Fortuna. Automated decision-making under the gdpr: Practical cases from courts and data protection authorities. Report, Future of Privacy Forum, May 2022. Accessed on January 31, 2024.
- [59] UK Government and International Expert Panel. International ai safety report 2025. Available at: <https://www.gov.uk/government/publications/international-ai-safety-report-2025/international-ai-safety-report-2025>. Accessed October 2025.
- [60] OdiseIA. Inventory of comparative judgments and rulings: Ai and vulnerable groups, 2024.
- [61] David Leslie and Andrea M. Perini. Future shock: Generative ai and the international ai policy and governance crisis. *Harvard Data Science Review*, Special Issue 5, 2024.
- [62] Sandra Mayson. Bias in, bias out. *Yale Law Journal*, 128:2218–2300, 2019.
- [63] B.C. Cheong. Transparency and accountability in ai systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6:1421273, 2024.
- [64] The Guardian. The racism of technology - and why driverless cars could be the most dangerous example yet. The Guardian (Shortcuts blog), March 13 2019.
- [65] European Council. Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination, 2023.
- [66] G. Franklin, R. Stephens, M. Piracha, S. Tiosano, F. Lehouillier, R. Koppel, and P. L. Elkin. The sociodemographic biases in machine learning algorithms: A biomedical informatics perspective. *Life (Basel)*, 14(6):652, May 21 2024.
- [67] European Union Agency for Fundamental Rights (FRA). Bias in algorithms - artificial intelligence and discrimination, December 8 2022.
- [68] European Parliamentary Research Service (EPRS). Auditing the quality of datasets used in algorithmic decision-making systems, 2022.
- [69] A. Gaudeul, O. Arrigoni, V. Charisi, M. Escobar Planas, and I. Hupont Torres. The impact of human oversight on discrimination in ai-supported decision-making: A large case study on human oversight of ai-based decision support systems in lending and hiring scenarios, 2024.
- [70] Pablo Cerezo-Martínez, Alejandro Nicolás-Sánchez, and Francisco J. Castro-Toledo. Analyzing the european institutional response to ethical and regulatory challenges of artificial intelligence in addressing discriminatory bias. *Frontiers in Artificial Intelligence*, 7, June 25 2024.
- [71] European Union Agency for Fundamental Rights (FRA). Getting the future right – artificial intelligence and fundamental rights, December 14 2020.
- [72] Rebecca Crootof, Margot E. Kaminski, and W. Nicholson Price II. Humans in the loop. *Vanderbilt Law Review*, 76:429–494, 2023. Also available as U of Colorado Law Legal Studies Research Paper No. 22-10 and U of Michigan Public Law Research Paper No. 22-011.
- [73] Article 29 Data Protection Working Party. Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679 (wp251rev.01), 2018. Adopted on 3 October 2017, as last revised and adopted on 6 February 2018.

- [74] Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.
- [75] Jens Laux. Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of ai governance under the european union ai act. *AI & Society*, 39:2853–2866, 2024.
- [76] European Data Protection Supervisor. Techdispatch: Human oversight of automated decision-making. Technical report, European Data Protection Supervisor (EDPS), 2025. Issue Authors: Vítor Bernardo, Laura Hernández. EDPS Technology and Privacy Unit.
- [77] Rumman Chowdhury. Moral outsourcing: Finding the humanity in artificial intelligence. *Forbes*, 2017.
- [78] Paula Aceve. “i do not think ethical surveillance can exist”: Rumman chowdhury on accountability in ai. *The Guardian*, 2023.
- [79] World Economic Forum. Navigating the ai frontier: A primer on the evolution and impact of ai agents, 2024.
- [80] Cathy O’Neil. *Weapons Of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 2017.
- [81] Jeff Larson, Julia Angwin, Lauren Kirchner, and Surya Mattu. How we analyzed the compas recidivism algorithm. *ProPublica*, May 2016. Accessed October 2025.
- [82] F. J. Zuiderveen Borgesius. Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, 24(10):1572–1593, 2020.
- [83] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81:77–91, 2018.
- [84] Michelle Hurley and Julius Adebayo. Credit scoring in the era of big data. *Yale Journal of Law and Technology*, 18(1):148, 2016.
- [85] Virginia Eubanks. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin’s Press, 2018.
- [86] Gustavo Gil Gasiola. Rebuilding the pyramid: The ai act’s risk-based approach using a binary decision diagram. *Computer Law & Security Review*, 58:106189, 2025.
- [87] Martin Ebers. Truly risk-based regulation of artificial intelligence: How to implement the eu’s ai act. *European Journal of Risk Regulation*, 2024.
- [88] A. Tartaro, A. L. Smith, and P. Shaw. Assessing the impact of regulations and standards on innovation in the field of ai. *arXiv preprint arXiv:2302.04110*, 2023.
- [89] OECD. Regulatory sandboxes in artificial intelligence, July 13 2023.
- [90] OECD. Regulatory sandboxes can facilitate experimentation in artificial intelligence, May 31 2023.
- [91] European Parliament. Artificial intelligence act and regulatory sandboxes, June 2022.
- [92] Claudio Novelli, Philipp Hacker, Simon McDougall, Jessica Morley, Antonino Rotolo, and Luciano Floridi. Getting regulatory sandboxes right: Design and governance under the ai act, 2025. Available at SSRN.
- [93] The World Bank Group. Global experiences from regulatory sandboxes, 2020.
- [94] Josu A. Eguiluz Castañeira and Anna Via. First european ai high-risk sandbox lessons: How smes can implement high-risk ai in a real-world environment, 2025. SSRN Working Paper.

- [95] Ministry of Economic Affairs and Digital Transformation. Royal decree 817/2023 of 8 november establishing a controlled testing environment for the trial of compliance with the proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence, 2024. Published in the Official State Gazette (BOE) No. 268, 9 November 2023, Article 8(j).
- [96] European Commission. Living repository to foster learning and exchange on ai literacy, 2025. Accessed October 2025.
- [97] European Commission. Ai literacy – questions and answers, 2025. Accessed October 2025.
- [98] European Commission. Commission launches ai act service desk and single information platform to support ai act, 2025. Accessed October 2025.
- [99] Robert Kilian, Linda Jäck, and Dominik Ebel. European ai standards – technical standardisation and implementation challenges under the eu ai act. *European Journal of Risk Regulation*, pages 1–25, 2025.
- [100] J. Soler Garrido, D. Fano Yela, C. Panigutti, H. Junklewitz, R. Hamon, T. Evas, A. André, and S. Scalzo. Analysis of the preliminary ai standardisation work plan in support of the ai act, 2023.
- [101] J. Soler Garrido, S. De Nigris, E. Bassani, I. Sanchez, T. Evas, A. André, and T. Boulangé. Harmonised standards for the european ai act. Technical Report JRC139430, European Commission, Seville, 2024.
- [102] European Commission. Commission collects feedback to simplify rules on data, cybersecurity and artificial intelligence in upcoming initiatives, 2025. Accessed October 2025.
- [103] European Parliament. Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), December 2016.
- [104] Gerard Buckley, Tristan Caulfield, and Ingolf Becker. “it may be a pain in the backside but. . .” insights into the resilience of business after gdpr. *Proceedings of the 2022 New Security Paradigms Workshop (NSPW ’22)*, pages 21–34, 2023.
- [105] Organisation for Economic Co-operation and Development. Privacy-enhancing technologies (pets), 2025. Accessed October 2025.
- [106] European Data Protection Board. Training curriculum on ai and data protection fundamentals of secure ai systems with personal data, 2025. See Section 5: Privacy Enhancing Technologies in AI Systems.
- [107] L. Bertolaccini, P.E. Falcoz, A. Brunelli, J. Furák H.F. Batirel, S. Passani, and Z. Szanto. The significance of general data protection regulation in the compliant data contribution to the european society of thoracic surgeons database. *European Journal of Cardio-Thoracic Surgery*, 2023.
- [108] Konrad Kollnig Bram Rijsbosch, Gijs van Dijck. Adoption of watermarking measures for ai-generated content and implications under the eu ai act. arXiv preprint, 2025.
- [109] European Parliamentary Research Service. Generative ai and watermarking. Briefing Document, European Parliament, 2023. Accessed October 2025.
- [110] National Institute of Standards and Technology (NIST). Reducing risks posed by synthetic content: An overview of technical approaches to digital content transparency, 2024. NIST AI 100-4, Trustworthy and Responsible AI Series.
- [111] Alexander Peukert and Céline Castets-Renard. Code of practice for general-purpose ai models: Copyright chapter. European Commission, 2025. Working Group 1, Copyright Chapter, includes obligations under Article 53(1)(c) AI Act.
- [112] ChatGPT is Eating the World. Updated map of us copyright suits v. ai companies (oct. 19, 2025), 2025. Accessed October 2025.

- [113] Angela Au-Yeung. Apple forces competitors to play by its rules with a new operating system. *Forbes*, April 20 2021.
- [114] Martin Greenacre. Eu is ‘losing the narrative battle’ over ai act, says un adviser. *Science|Business*, December 5 2024.
- [115] Michael E. Porter. America’s green strategy. *Scientific American*, pages 168–176, April 1991.
- [116] Michael E. Porter and Claas van der Linde. Toward a new conception of the environment–competitiveness relationship. *Journal of Economic Perspectives*, 9(4):97–118, 1995.
- [117] Philip Brey and Benjamin Dainow. Ethics by design for artificial intelligence. *AI Ethics*, 4:1265–1277, 2024.
- [118] Katrin de Fine Licht and Johannes de Fine Licht. Artificial intelligence, transparency, and public decision-making. *AI & Society*, 35:917–926, 2020.
- [119] European Union Agency for Fundamental Rights. Getting the future right – artificial intelligence and fundamental rights. Technical report, European Union Agency for Fundamental Rights, December 2020.
- [120] European Data Protection Board (EDPB). Edpb opinion on ai models: Gdpr principles support responsible ai, December 18 2024.
- [121] European Commission. The general-purpose ai code of practice, July 2025. Accessed: 2025-10-22.
- [122] High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy ai. European Commission, April 2019.
- [123] Cecilia Panigutti, Ronan Hamon, Isabelle Hupont, David Fernandez Llorca, Delia Fano Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Soler Garrido, and Emilia Gomez. The role of explainable ai in the context of the ai act. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’23)*, pages 1–12, New York, NY, USA, 2023. ACM.
- [124] Isabelle Hupont, David Fernandez-Llorca, Silvia Baldassarri, et al. Use case cards: A use case reporting framework inspired by the european ai act. *Ethics and Information Technology*, 26:19, 2024.
- [125] Council_of_Europe. Framework convention on artificial intelligence and human rights, democracy and the rule of law, September 5 2024.
- [126] Gemma Galdon Clavell. Ai auditing: Proposal for ai leaflets. European Data Protection Board, SPE Programme, 2024. Accessed October 2025.
- [127] U. Peters. Explainable ai lacks regulative reasons: Why ai and human decision-making are not equally opaque. *AI and Ethics*, 3(3):963–974, 2023.
- [128] Francesco Palmiotto. The ai act roller coaster: The evolution of fundamental rights protection in the legislative process and the future of the regulation. *European Journal of Risk Regulation*, pages 1–24, 2025.
- [129] Alessandro Mantelero. The fundamental rights impact assessment (fria) in the ai act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review*, 54:106020, 2024.
- [130] Confederación Española de Consumidores y Usuarios (CECU). Towards a meaningful implementation of article 27 under the ai act: Participation in fundamental rights impact assessments (frias) & effective transparency mechanisms for their oversight and enforcement. Technical report, CECU, 2025. Accessed October 2025.
- [131] Gianclaudio Malgieri and Cristiana Santos. Assessing the (severity of) impacts on fundamental rights. *Computer Law & Security Review*, 2025. Forthcoming.

- [132] Catalan Data Protection Authority. Fria model: Guide and use cases. fria methodology for ai design and development, 2025.
- [133] Government of the Netherlands. Impact assessment fundamental rights and algorithms. Technical report, Ministry of the Interior and Kingdom Relations, 2021. Accessed October 2025.
- [134] Danish Institute for Human Rights. Guidance on human rights impact assessment of digital activities. Technical report, Danish Institute for Human Rights, 2020. Accessed October 2025.
- [135] David Leslie, Christopher Burr, et al. Human rights, democracy, and the rule of law assurance framework for ai systems: A proposal (huderia). Technical report, The Alan Turing Institute, 2022. Accessed October 2025.
- [136] Algorithm Audit. A comparative review of 10 fundamental rights impact assessments (fria) for ai systems. Technical report, Algorithm Audit, 2024. Accessed October 2025.
- [137] Spanish Data Protection Agency (AEPD). Evaluating human intervention in automated decisions, 2024.
- [138] Independent High-Level Expert Group on Artificial Intelligence. The assessment list for trustworthy artificial intelligence (altai) for self-assessment, 2020.
- [139] Daniel Innerarity. El impacto de la inteligencia artificial en la democracia. *Revista de las Cortes Generales*, (109):87–103, 2020.
- [140] Ljubiša Metikoš and Jef Ausloos. The right to an explanation in practice: Insights from case law for the gdpr and the ai act. *Law, Innovation and Technology*, pages 1–36, March 2025.
- [141] Research Institute – Digital Human Rights Center. The right to explanation in the ai act business manual. Technical report, Research Institute – Digital Human Rights Center, Vienna, Austria, 2025. Project: Enlightenment 4.0 – Human Understanding of AI Decision-Making.
- [142] R. Medaglia, P. Mikalef, and L. Tangi. Competences and governance practices for artificial intelligence in the public sector, 2024.
- [143] E. Werneman Root and M. Mahay. Assessing ai literacy needs. *IAPP*, 2025.
- [144] Sandra 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. *Yale Journal of Law and Technology*, 26(3), 2024.
- [145] P. Langley. Crafting papers on machine learning. In Pat Langley, editor, *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pages 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.