

Who Foots the Bill? State-Backed No-Fault Compensation for Experimental Harms in Artificial Intelligence Regulatory Sandboxes

Ha-Chi Tran

Independent Researcher *
chih.tran1@gmail.com

Abstract

As emerging technologies such as artificial intelligence (AI) outpace existing legal frameworks, regulatory sandboxes have become a key mechanism to reconcile innovation with risk governance by allowing firms to test novel technologies under regulatory oversight with certain temporary legal exemptions. While effective in fostering innovation, these frameworks leave unresolved liability concerns when harm results from compliant experimentation. Existing models place the burden of compensation primarily on participating firms, neglecting the potential role of state-backed liability mechanisms. This paper critically examines the accountability gap in regulatory sandboxes for AI by analyzing the limitations of current compensation frameworks used in leading jurisdictions. It argues that existing sandbox models, which primarily allocate civil and criminal liability to participating firms, are ill-equipped to address the distinct harms posed by AI technologies. These inadequacies arise from legal exemptions that undermine access to fault-based remedies, the technical opacity and distributed agency of AI systems, and the systemic nature of many AI-related harms. To address these challenges, the paper proposes a state-backed no-fault compensation scheme, modeled on Vaccine Injury Compensation Programs (VICPs), to more equitably and effectively distribute liability. This paper argues that the state's active involvement in sandbox governance renders it a co-responsible actor, warranting shared liability in harm cases. A state-backed no-fault approach would also help overcome the attribution challenges posed by opaque, distributed AI systems, reduce barriers for small and medium-sized enterprises (SMEs) in low- and middle-income countries (LMICs), and strengthen public trust in experimental regulatory regimes.

Introduction

First launched by the United Kingdom (UK) Financial Conduct Authority (FCA) in 2015, the regulatory sandbox model quickly attracted international attention as a tool for fostering innovation within regulatory systems (Cornelli et al. 2024; McCarthy 2023). While originally designed for Fin-Tech, it has since broadened to support a wider array of

emerging technologies, most notably artificial intelligence, or AI (Qiu et al. 2025; Kilian, Jäck, and Ebel 2025; Moraes 2025).

The rising adoption of regulatory sandboxes highlights their growing role in adaptive governance, amid the profound uncertainty surrounding frontier technologies, especially advanced AI systems. The rapid pace of AI development increasingly outstrips the capacity of traditional legal and institutional frameworks, creating regulatory lag and undermining the efficacy of conventional governance mechanisms (Dixon 2023; Ahern 2025). This challenge is intensified by the inherent unpredictability and unexplainability of complex AI systems, especially sub-symbolic models, whose opaque, black-box, and nonlinear nature hinders comprehensive risk assessment (Arrieta et al. 2020; Karnow 2016). In the context of AI advancement, these conditions exemplify Knightian uncertainty, situations in which neither outcomes nor their probabilities can be reliably determined. Under such conditions of “unknown unknowns,” conventional ex-ante regulatory models, which rely on stable risk assessments and fixed legal categories, may either fail to offer sufficient protection (Zhang et al. 2025; Scherer 2016; Rangone and Megale 2025) or become overly cautious (Carvão et al. 2025), thereby potentially stifling innovation (Abiri and Huang 2023; Arnal 2025).

Regulatory sandboxes, therefore, provide a pragmatic and adaptive alternative to AI governance by allowing selected entities to test their products, services, or business models in controlled, real-world environments (Truby et al. 2022; Qiu et al. 2025; Liao 2020). Examples of countries that have adopted regulatory sandboxes to govern their domestic AI landscape (or frontier technologies in general) include the European Union (EU) member states (through the EU AI Act), the UK, China, Canada, Japan, Singapore, and Indonesia (Qiu et al. 2025; Arnal 2025; Kataoka 2019; Morgan 2023; Tsai, Lin, and Liu 2020). Typically, these frameworks temporarily relax certain legal or compliance requirements, reducing barriers to innovation while maintaining regulatory oversight (Kilian, Jäck, and Ebel 2025; Morgan 2023). Through continuous monitoring and real-time evaluation, sandboxes generate empirical insights that help identify latent risks, emerging behaviors, systemic biases, and complex socio-technical dynamics often missed by theoretical models or ex-ante assessments (Yoshikawa 2019). As

*This research was conducted independently prior to submission. At the time of publication, the author is affiliated with The London School of Economics and Political Science, UK (c.h.tran@lse.ac.uk).

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

iterative learning tools, they support both policy refinement and more responsive, evidence-based regulation, while safeguarding space for innovation (Qiu et al. 2025; Boura 2024).

Despite the above promising benefits, regulatory sandboxes embody an inherent tension between their pro-innovation mandate and the core regulatory objective of safeguarding public and consumer interests. The temporary suspension or relaxation of legal and compliance obligations within regulatory sandboxes raises valid concerns about increased risk exposure for users and third parties that would typically be mitigated under fully regulated conditions (Allen 2020; Knight and Mitchell 2020; Ranchordas 2021). Although some contend that the associated trade-off risks are justifiable and potentially preventable by arguing that enhanced governmental oversight and the use of sandboxes as learning tools benefit both innovators and regulators in developing long-term safeguards (Qiu et al. 2025), the exemption-based structure nonetheless introduces significant legal and ethical vulnerabilities (Gromova and Ivanc 2020; Johnson 2023). Additionally, current liability frameworks remain inadequate in providing effective redress and fail to ensure sufficient protection or compensation for harms arising from activities conducted under regulatory exemptions, particularly in the context of AI innovation.

The structure of this paper is as follows. The next section analyzes accountability and liability challenges in AI regulatory sandboxes, focusing on the limitations of current compensation frameworks. It argues that current models, which primarily place liability on participating firms, are unprepared to address the complex, opaque, and often systemic harms associated with AI technologies. Following, the paper then advances a proposal for a state-backed, no-fault compensation scheme, drawing analogies to established public redress mechanisms such as Vaccine Injury Compensation Programs. This approach, it is argued, would more appropriately allocate responsibility, lower procedural and evidentiary barriers to redress, and foster public trust while enabling equitable innovation, particularly in low- and middle-income countries (LMICs).

It should be emphasized that the position advanced in this paper is not intended as an objection to the use of regulatory sandboxes for AI technologies. On the contrary, the author acknowledges their significant value as a policy instrument for navigating the uncertainties inherent in the current landscape of AI development. The aim of this paper is to, indeed, identify and critically examine specific features, both intrinsic to the regulatory sandbox model and inherent in the nature of AI technologies, that, if left unaddressed, may produce adverse consequences, particularly with respect to the rights of affected individuals to compensation and protection. Its purpose is to foreground these risks and to advance a remedial pathway through the adoption of a state-backed, no-fault compensation mechanism.

Yet, this paper does not seek to provide a detailed institutional design for such a scheme, acknowledging that optimal configurations will necessarily vary across jurisdictions according to their distinct legal traditions, welfare infrastructures, and policy priorities (Dickson et al. 2016). Future research, particularly case-specific analyses, could provide

valuable insights into the refinement of such mechanisms. However, this lies beyond the scope of this paper.

Compensation Liability Gaps in AI Regulatory Sandboxes

Regulatory sandboxes provide a controlled environment for the supervised experimentation of emerging technologies, including AI. However, they cannot eliminate the inherent risks posed by AI systems. These risks arise from the non-linear, unpredictable, and non-deterministic nature of AI behavior, characteristics further complicated by the system's capacity to adapt in unforeseen ways to dynamic, uncertain, and complex environments, leading to what are often described as emerging behaviors (Kolt 2023; The Alan Turing Institute 2024; Piantadosi, Marrone, and Sansone 2018; Yoshikawa 2019). Such properties pose fundamental challenges to formal methods of verification, testing, and validation, thereby limiting the reliability and effectiveness of these techniques in guaranteeing system safety (Yoshikawa 2019). Even when regulatory sandboxes employ rigorous pre-deployment evaluations, continuous monitoring, and layered safeguards, the successful completion of testing or risk mitigation efforts cannot be regarded as definitive assurance against the possibility of future harm, whether during the experimental phase or following broader market integration (Allen 2020).

Most regulatory sandboxes, therefore, do not exempt firms from liability toward third parties (Carneiro 2024; OECD 2025b), meaning that participants generally remain legally responsible for any harm caused during experimentation. This principle reflects a broader regulatory consensus that firms should bear civil or criminal liability for consumer harm, though enforcement varies across jurisdictions. Yet the issue is not merely whether firms should be held liable per se, but that they are often the sole bearers of responsibility, both within sandboxes and after sandbox approval (see Table 1). It remains unclear whether state agencies should share liability for harm incurred during or following sandbox trials. Compounding this problem is the inadequacy of traditional compensation mechanisms, which struggle to address the complexities of AI-related harm, such as opaque algorithmic processes, diffuse responsibility, and difficulties in proving causation.

This section critically examines these shortcomings of traditional fault-based and strict liability frameworks in addressing the novel risks posed by AI. It highlights both the structural limits of these models and the difficulty of adapting liability rules to AI's unpredictability and opacity. Three key limitations drive this inadequacy. First, legal exemptions granted during sandbox participation may, in and of themselves, restrict access to traditional fault-based compensation mechanisms, thereby narrowing the legal avenues available to parties harmed by sandbox-tested innovations. Second, the opaque and distributed nature of AI systems, often characterized as "black-box" technologies, obstructs clear attribution of responsibility for compensational purposes, even when a causal link to the sandbox-tested product is evident. Third, numerous AI-associated harms are sys-

Regulatory Sandbox	Aspect Regulated	Participating Firms' Liability	Government Liability
EU AI Regulatory Sandboxes (EU Parliament 2024)	Innovative AI Systems	Providers are fully liable for any third-party damage resulting from experimentation in the sandbox.	Not explicitly stated.
UK FCA Regulatory Sandbox (FCA 2015)	Financial technology innovation, including AI applications in financial services	Regulation exemptions do not limit liabilities to consumers for any damage. Firms are expected to provide appropriate redress if required.	Not explicitly stated.
Utah (USA) AI Learning Lab (Utah AI Policy Act 2024)	Generative AI, AI systems interacting with consumers	Businesses cannot avoid liability under the Utah Consumer Protection Law on the basis that a violative statement or act was made by AI.	The state is not liable for any claims, damages, losses, or expenses arising from a participant's involvement in the learning laboratory.
Saudi Arabia Emerging Technology Regulatory Sandbox (Saudi Arabia CST 2023)	Innovative solutions built on advanced technologies such as AI, IoT, Blockchain, Cloud Computing, AR/VR	Participating firms remain liable for any damage caused to customers.	Not explicitly stated.
Brazil AI Regulatory Sandbox (Brazil MoDITS 2023)	Emerging technologies, especially AI and AI-related technologies	Compensation and dispute resolution mechanisms are needed to protect consumers, though it remains unclear whether responsibility lies with firms, the state, or both.	Not explicitly stated.

Table 1: Examples of Liability for Damage in AI and Emerging Technology Regulatory Sandboxes

temic, diffuse, and socio-economically far-reaching, rendering firm-level compensation liability alone insufficient to address structural dimensions of damages and ensure accountability. Such characteristics often exceed the capacity of corporate actors, particularly small and medium-sized enterprises (SMEs) and/or start-ups.

Legal Exemptions and Their Impact on Fault-Based Liability Systems

In the context of AI regulation, two principal liability frameworks dominate the legal discourse: fault-based liability and strict liability (Truby et al. 2022). Fault-based liability, which is particularly supported by legal scholars in the United States, requires the injured party to prove that the harm resulted from the defendant's negligence or a breach of a legal duty to obtain compensation (Coleman 2002). In contrast, strict liability, gaining increasing traction within European regulatory discourse, dispenses with the requirement to prove fault or intent, instead holding actors liable solely on the basis of causation and the inherently hazardous nature of certain AI-driven activities (Coleman 2002). Within this context, regulatory sandboxes, designed to promote innovation by temporarily exempting emerging technologies from certain legal requirements, may inadvertently undermine and dilute fault-based liability regimes. By relaxing regulatory constraints, these frameworks can weaken the legal avenues available to harmed individuals seeking redress.

When regulatory standards, particularly those related directly or indirectly to product safety, consumer protection, and industry compliance, are relaxed, the legal foundation

for seeking compensation based on breaches of these regulations is significantly weakened (Sherkow 2022; Allen 2020). In fault-based liability systems, establishing negligence typically requires demonstrating a breach of a legal duty or a failure to meet an established duty of care (Fonseca, Sequeira, and Xavier 2024; Muftić 2025; Goldberg and Zipursky 2016). Notably, in such systems, fault is not solely determined by the presence of harm, but also by the violation of a regulatory obligation (Goldberg and Zipursky 2016). However, when these standards are modified or suspended within regulatory sandboxes, determining whether a breach has occurred becomes more complex, which complicates the attribution of fault and the assignment of liability (Fonseca, Sequeira, and Xavier 2024). The suspension of those obligations erodes the legal structure used to attribute responsibility. Moreover, legal instruments such as safe harbors or no-enforcement letters can further diminish firms' incentives to adhere to safety and ethical standards, while simultaneously complicating post-hoc legal evaluations of responsibility in the event of harm (Allen 2020; Buocz, Pfothenauer, and Eisenberger 2023; Tsai, Lin, and Liu 2020).

A claimant may face significant challenges in establishing that a firm was "at fault" if the firm operated in accordance with modified rules approved within a regulatory sandbox (Allen 2020). The exemption from penalties in such contexts may inadvertently signal a degree of regulatory acceptance or endorsement of certain practices, thereby complicating the attribution of fault in subsequent civil proceedings, even in cases where liability formally remains. This dynamic introduces a potential disconnect between regulatory leniency

and the pursuit of civil redress, creating a substantial barrier to effective remedies for individuals who suffer harm. While such exemptions are intended to incentivize innovation and allow for the safe experimentation of emerging technologies, they simultaneously risk undermining the deterrent function of regulatory enforcement. Administrative fines serve not only a punitive purpose but also act as preventive signals, emphasizing the importance of legal compliance and the potential risks associated with emerging technologies, even in the absence of actual harm (Acciarri 2015). The suspension of these fines during sandbox participation may, therefore, compromise the broader protective objectives of the legal framework, potentially leaving accountability gaps during critical experimental phases.

The inherent trade-off in regulatory sandbox design, prioritizing innovation at the expense of legal burden, often results in the implicit weakening of remedial mechanisms and safeguards available to affected parties. Proponents of regulatory sandboxes may contend that these exemptions are principally intended to rationalize and simplify the legal and financial frameworks governing the interface between the state and participating enterprises, rather than to absolve firms of their enduring ethical and legal obligations toward consumers or third parties. Even in cases where consumer-related risks are implicated, firms are typically required to implement risk prevention and mitigation mechanisms as a precondition for sandbox participation (Alaassar, Mention, and Aas 2021; FCA 2017). From this perspective, the regulatory risk to third parties is considered minimal. However, such reasoning overlooks a fundamental characteristic of corporate behavior: the strategic pursuit of jurisdictions offering more favorable legal environments. This tendency can lead to “sandbox shopping,” wherein firms purposely select sandbox frameworks with looser regulatory requirements to reduce legal obligations (Truby et al. 2022). Over time, this behavior may contribute to a broader “race to the bottom,” whereby regulatory standards are incrementally lowered or selectively relaxed in the name of attracting innovation and investment, often at the expense of consumer protection and legal accountability (Allen 2020; Buckley 2023; Knight and Mitchell 2020; Ahern 2021). These developments highlight the risk that, if not carefully structured and monitored, sandbox initiatives may unintentionally promote deregulatory competition, undermining the legal frameworks designed to safeguard public interests.

Indeed, these concerns have already been raised and proven in the FinTech sector, where the regulatory sandbox model first emerged. Empirical evidence and theoretical arguments suggest that, aside from short periods following major associated crises, there is rarely a strong or organized constituency advocating for enhanced consumer protection or stricter financial (and technological) regulation (Allen 2020; Zetsche et al. 2018). In contrast, there is a consistently powerful incentive, both for private actors and public regulators, to loosen regulatory constraints in the name of innovation, economic competitiveness, or technological leadership (Curtis and Schulman 2006; Evans and Watson 2015). In the context of global technological competition, especially in transformative sectors such as AI, regulatory

tolerance is often framed as essential to national economic and strategic objectives (Mitchell 2025).

This imbalance of incentives may reflect a broader systemic trend among regulatory regimes for attracting innovative firms, rather than a problem limited to regulatory sandboxes only. However, the structure of regulatory sandboxes may exacerbate this dynamic by facilitating deregulation under the guise of fostering innovation. By temporarily suspending or diluting regulatory obligations, sandboxes effectively lower the burden of proof required to justify deregulation, shielding it behind the rhetoric of experimentation and progress (Buckley 2023; Ahern 2021; Allen 2020). This has significant consequences for fault-based liability regimes. When regulatory norms that ordinarily define duties of care or serve as benchmarks for assessing negligence are weakened or rendered inapplicable, the legal foundation upon which individuals rely to seek redress becomes increasingly unstable. In this context, the applicability and effectiveness of fault-based liability mechanisms are not merely diminished, but they may also be fundamentally undermined.

That is to say, while sandbox participation is often justified by the promise of more robust regulation in the future, such forward-looking assurances offer little recourse for those harmed during the interim experimental phase. Because traditional fault-based claims frequently depend on proving violations of specific regulatory duties, now relaxed or suspended, victims face significantly higher barriers in seeking compensation. The resulting structural inconsistency in supporting fault-based claims may prolong legal conflicts between injured parties and injurers, thereby delaying or even denying effective remedies. Ultimately, this raises broader concerns about whether the temporary legal leniency afforded by sandboxes is compatible with the enduring principles of justice and accountability in liability systems.

Opaque AI Systems and the Difficulties in Attribution for Compensation

When liability based on regulatory violations is reduced, whether through exemptions, relaxed compliance standards, or experimental contexts such as regulatory sandboxes, the burden of proof for establishing compensation and legal accountability often places more importance on demonstrating actual, concrete harm. This challenge is compounded by the black box nature of many advanced AI systems, which obscures whether there has been a breach of the duty of care and whether such a breach foreseeably caused the harm suffered, two core requirements under negligence and product liability regimes (Holm, Stanton, and Bartlett 2021).

The term “black box” refers to the inherent complexity and opacity of decision-making processes in AI models, particularly those built on deep learning and neural networks (Franzoni 2023). In such systems, even the developers and technical experts responsible for their design often struggle to fully explain how specific outputs are generated from given inputs (Ding et al. 2022; Franzoni 2023; Holm, Stanton, and Bartlett 2021). The vast scale and intricacy of these models, containing millions or even billions of algorithmic parameters, make it virtually impossible to trace which par-

ticular element or interaction within the system caused a harmful outcome (Rodríguez de Las Heras Ballell 2025). This epistemic opacity fundamentally undermines the ability of the legal system to assess intent and causation, core components of fault-based liability, especially within tort law (Holm, Stanton, and Bartlett 2021). It also weakens product liability frameworks that require a causal link between a design defect and a predictable harm (Funkhouser 2013).

That is to say, when an AI system's internal logic is inscrutable, establishing a causal connection between an action (or omission) and the harm caused becomes significantly more difficult. In many cases, the machine learning environment and data context are treated as superseding causes, intervening events that break the causal chain between the product's original design and the resulting injury (Yoshikawa 2019; Chopra and White 2011). As a result, assigning legal responsibility based on deliberate or negligent conduct is often unfeasible. Regulatory sandboxes, by design, accommodate the deployment of such opaque systems under looser regulatory constraints, thereby exacerbating the disconnect between technological complexity and legal accountability. Consequently, injured parties may find themselves with limited avenues for redress.

The black box nature of AI raises not only evidentiary barriers to proving fault but also conceptual challenges to the very framework of fault-based liability. Traditional legal notions of fault presuppose the identification of a human error or a breach of legal duty, usually tied to intentional misconduct or negligence. However, AI-induced harms frequently stem from emerging system behaviors, such as unforeseen data correlations or learned biases embedded within training datasets, none of which may have been consciously encoded or anticipated by developers (Yoshikawa 2019). Examples include Google Photos erroneously labeling Black individuals as gorillas (Chopra and White 2011), or high-frequency trading algorithms triggering flash crashes in financial markets in 2010 and 2015 (Yoshikawa 2019; Kirilenko A. et al. 2017). These systems often operate in ways that defy intuitive understanding, weakening the plausibility of attributing fault in the conventional sense. To date, computer scientists have been unable to predict such emerging harmful behaviors in advance; rather than ex-ante, most interventions occur ex-post, such as the implementation of circuit breakers to halt trading following a crash (Yoshikawa 2019).

Even when a causal relationship can be demonstrated between an AI-driven application and a harmful outcome, assigning responsibility to a specific individual or organization remains challenging. This complexity is often attributed to the "problem of many hands," a concept describing the diffuse and distributed nature of responsibility in collective processes (Coeckelbergh 2020; Slota et al. 2023; Khosrowi, Finn, and Clark 2024). AI development and deployment typically involve multiple stakeholders across various stages: data acquisition and pre-processing, model design, algorithm training, system integration, implementation, and end-user interaction (Khosrowi, Finn, and Clark 2024). Although each actor may perform their designated role in accordance with industry standards or regulations, the resulting harm may still arise due to cumulative or unanticipated

interactions among system components. As a result, it becomes highly contentious and legally ambiguous to identify which entity, or combination of entities, should bear primary responsibility. The complexity of assigning liability is further exacerbated when groups of actors involved span both within and outside the regulatory sandbox. This hybrid composition creates disparities in the applicable legal regimes, potentially resulting in fragmented or inconsistent liability assessments within the same factual scenario.

It is at this point that the structural limits of not only fault-based but also strict liability become apparent. Traditionally, strict liability imposes responsibility irrespective of fault, under the premise that manufacturers and operators are best positioned to control risk, gain relevant knowledge, and internalize harm-related costs. Yet, in the context of autonomous and black-box AI systems, these justifications become problematic. Developers and deployers frequently lack full insight into or control over AI behavior, especially when harmful outcomes stem from unexpected interactions with real-world data or environments (Kolt 2023; Spence, Behar, and Jayadev 2024). These effects might emerge only post-deployment, escaping even the most robust testing and risk mitigation protocols. Given AI's sensitivity to data and environmental inputs, testing mechanisms can capture only a narrow slice of possible scenarios, leaving many edge cases and hidden interactions undetected. Consequently, harms that do not manifest during development may arise unexpectedly, amplified by the epistemic opacity of black-box models. Under these conditions, imposing strict liability on participants merely for their involvement in the development/deployment process is neither analytically sound nor legally just (Holm, Stanton, and Bartlett 2021; Yoshikawa 2019).

Indeed, the structural challenges that AI poses to existing liability regimes have been explicitly recognized by the EU, as reflected in two legislative instruments introduced to complement the AI Act: the revised Product Liability Directive (revPLD) and the AI Liability Directive (AILD). Both instruments acknowledge the existence of a regulatory vacuum in which individuals harmed by AI systems face significant obstacles in securing compensation, owing to the epistemic opacity, adaptive behavior, and distributed agency that characterize modern AI. These features complicate the attribution of responsibility, increase litigation costs, and prolong proceedings, often to the detriment of timely and effective redress for victims.

Yet, while the revPLD successfully entered into force on 9 December 2024, the AILD, a proposed directive that aimed to ease the burden of proof in AI-related claims by introducing disclosure obligations and burden-shifting mechanisms, was ultimately withdrawn in February 2025, following widespread criticism regarding its limited scope and normative inadequacy (Duffour and Gerke 2023; Schütte 2025). Critically, scholars have argued that the AILD fails to address key legal challenges arising from harms caused by autonomous, black-box AI systems operating in unpredictable real-world conditions without an identifiable or foreseeable cause. As a result, these harms frequently fall outside the scope of both traditional fault-based liability and

strict liability frameworks.

The failure and withdrawal of the AILD thus reflect more than the limitations of a single legislative proposal; they underscore a deeper conceptual incompatibility between legacy liability frameworks and the technical realities of contemporary AI. Bridging this disconnect demands the development of novel liability models that transcend the conventional fault versus strict liability dichotomy and are specifically designed to accommodate the epistemic uncertainty, emergent risk, and distributed accountability embedded within today's AI systems.

Systemic AI Harms and the Inadequacy of Firm-Only Liability for Compensation

The prevailing model within regulatory sandboxes, which emphasizes the legal liability of private firms while vaguely mentioning the role of the state in addressing/compensating harm, may appear straightforward but rests on a problematic assumption: that AI-related harms can be effectively managed at the enterprise level. This notion is not only fundamentally challenged by the unique characteristics of AI systems, such as their opacity, complexity, and unpredictability, as mentioned in the previous subsection, but also reflects a misguided understanding of the nature, scale, and systemic risks associated with AI-induced damage. Many types of harm from AI are not limited to a single individual or transaction, but can affect large groups of people, entire markets, or even underlying social structures and values.

Typical examples of systemic and pervasive harm from AI include, but are not limited to (Slattery et al. 2024):

- **Algorithmic discrimination:** AI systems trained on biased data or designed without fairness constraints can systematically reinforce discrimination in domains such as hiring, credit, and criminal justice, thereby deepening existing social inequalities.
- **Disinformation and public manipulation:** Generative AI can rapidly disseminate disinformation at scale, threatening democratic institutions, exacerbating polarization, and eroding public trust in reliable sources.
- **Privacy degradation:** AI-driven surveillance and data analytics enable pervasive privacy intrusions, including mass profiling and data breaches, which compromise individual autonomy and societal freedoms.

As such, AI technologies are not monolithic, and neither are AI impacts; they are complex, adaptive, and frequently decentralized in both development and deployment. As mentioned, these systems often produce outcomes that are emergent rather than directly engineered. This structural complexity creates epistemological and legal challenges for measuring harm. In many cases, the impacts of AI are hidden, delayed, or ambiguous, raising difficulties not only in attribution but also in defining what constitutes harm (Thomas et al. 2025; Wilczek, Thäsler-Kordonouri, and Eder 2024; Nordström 2022). For instance, algorithmic decisions in credit scoring or predictive policing may produce outcomes that are legally defensible yet socially harmful, disproportionately affecting marginalized communities without clear markers of wrongdoing (Zarsky 2016). The logic

of firm-only liability, which presumes a discrete actor with control over risk-producing activities, collapses under the weight of such complexity.

In and of themselves, corporate firms are not inherently structured and capable of effectively addressing the broader socio-economic consequences of AI systems. Their institutional design prioritizes shareholder value maximization and market competitiveness over the mitigation of systemic societal risks. This structural limitation is especially pronounced among SMEs and startups, which often lack the financial resilience and organizational capacity to assume liability for harms (Koporcic et al. 2025; Durst and Henschel 2024), even when legal or ethical responsibility can be established, which is itself a complex and uncertain process. These constraints are particularly acute in LMICs, where local innovation ecosystems are frequently positioned as engines of inclusive technological development (Arcuri, Russo, and Gandolfi 2025; Sharma, Ajadi, and Beavor 2020). In such contexts, the imposition of strict liability regimes for harm compensation on emerging firms may have unintended consequences: discouraging innovation, deterring participation in regulatory experimentation, and inadvertently consolidating market dominance in the hands of large, well-capitalized multinational corporations.

This argument should not be interpreted as advocating for the wholesale exemption of firms from liability for harms resulting from their participation in regulatory sandboxes. Rather, it highlights the limitations of relying exclusively on individual firms, particularly those that are undercapitalized or in early stages of development, to shoulder the burden of compensating for harms that may be systemic, diffuse, or societally significant. It is crucial to distinguish between the two foundational principles underpinning compensation regimes: deterrence and redress (Marchisio 2021). Deterrence serves to prevent future negligence by holding firms accountable for harmful actions, while redress seeks to ensure that those who suffer damage are compensated in a proportionate and timely manner. These two objectives, while related, are not mutually dependent. Effective deterrence can, in principle, be maintained even in the absence of complete compensation, particularly if alternative liability mechanisms are in place.

Accordingly, it is insufficient to argue that reducing the compensatory burden on participating firms will necessarily increase the likelihood of negligence or diminish respect for third-party rights. On the contrary, an overly rigid liability framework may deter participation in regulatory sandboxes, stifle innovation, and entrench market dominance among already-powerful actors. Environmental law, indeed, provides a rich and instructive analogy for understanding this argument. A fundamental challenge in environmental governance is the issue of diffuse harm (Dauvergne 2018). Phenomena such as pollution, biodiversity loss, and climate change do not typically result from a single event or actor, but rather from the cumulative and often mutual effects of numerous discrete actions over time. In this context, the structural parallels between ecological and AI-related harms are not merely metaphorical. They reflect shared characteristics of distributed causality, deferred consequences, and

collective vulnerability. Environmental discourse has long grappled with the limitations of the polluter-pays principle (PPP), which posits that those who cause pollution should bear the costs of managing it (Baykal 2021; Bleeker 2009; García-Portela 2023). While appealing in theory, the PPP framework often fails to account for the significant asymmetries in capacity among polluters. Similarly, in the realm of AI, the imposition of strict liability for harms could unintentionally discourage experimentation and innovation, particularly by reinforcing monopolistic market structures dominated by well-capitalized actors with superior risk absorption capabilities (Mardhatillah 2023).

In response to the limitations of the PPP, environmental law has increasingly embraced public compensation schemes and risk-sharing mechanisms (Mardhatillah 2023). A comparable shift may be necessary in the governance of AI. If AI-induced harms, such as widespread job displacement, data-driven social polarization, or systemic algorithmic failures, are understood as collective action problems, then relying solely on individual firm liability is insufficient. Instead, liability frameworks must be complemented, and in some cases replaced, by publicly financed solutions, including state-backed liability-sharing schemes.

State-Backed No-Fault Compensation

This section advances a promising but underexplored proposal in AI governance: a state-backed, no-fault compensation scheme for harms arising from AI regulatory sandbox experimentation. This mechanism is intended to operate in parallel with, and as a complement to, existing strict liability frameworks. The proposed approach acknowledges the limitations of traditional fault-based liability principles, particularly the often-unrealistic assumption that compensation should depend on proving fault or negligence. It also responds to the potentially excessive deterrent effect of imposing strict liability solely on private enterprises in current regulatory sandbox frameworks.

Building on institutional precedents such as the Vaccine Injury Compensation Programs (VICPs), the proposed model advocates for a publicly funded, non-adversarial compensation scheme that dispenses with the need to prove fault, negligence, or product defect (Yang and Reiss 2023). The VICPs offer a persuasive precedent: established to counter the chilling effect of litigation on vaccine production, it provides a streamlined alternative to tort-based compensation. The program balances two public imperatives, ensuring fair redress for individuals harmed by vaccines and preserving public health by maintaining vaccine supply, by recognizing that socially beneficial activities can entail unavoidable risks (Yang and Reiss 2023; Mungwira et al. 2020). The rapid global expansion of no-fault compensation schemes for COVID-19 vaccine injuries, from 28 national programs in December 2019 to more than 130 countries in two years (Macleod, Uberti, and Kameni 2025), further underscores the growing recognition of such mechanisms as vital tools for managing systemic risks and sustaining socially valuable innovation. This pragmatic reconciliation of accountability and innovation makes the VICPs an instructive model for the governance of AI sandboxes, which similarly require

mechanisms to navigate the tension between experimental innovation and potential harm. Just as the VICPs institutionalize a regime of responsibility without the attribution of fault, states should likewise recognize and operationalize their obligation to address harms that may arise from the experimental regulatory environments they construct and oversee.

The justification for this approach rests on four interrelated grounds. First, by actively designing and supervising sandbox environments, the state becomes a co-producing actor in the creation of regulatory risk, and thus bears a principled responsibility to provide compensatory mechanisms. Second, a state-backed no-fault scheme avoids the evidentiary and legal complexities associated with the technical opacity and distributed agency characteristic of AI systems, which often hinder claimants in traditional liability regimes. Third, this model reduces compliance burdens on SMEs, particularly in LMICs, thereby fostering more inclusive and equitable innovation ecosystems. Fourth, by establishing predictable and accessible routes for redress, such a framework can enhance public trust and legitimacy, thereby supporting the responsible development and deployment of AI technologies within experimental regulatory contexts.

State Involvement Refers to Co-Liability

The emergence of AI regulatory sandboxes marks a significant shift in the governance of technological innovation. These controlled experimental environments, established and managed by public authorities, fundamentally reconfigure the relationship between state and market actors. Far from being passive regulators, states are deeply embedded in the design, oversight, and operational phases of AI sandbox frameworks. This multifaceted engagement renders the state not merely a facilitator, but a co-responsible actor in managing the regulatory risks that these experimental settings inherently generate. As such, there is a compelling normative and legal basis for the state's role in compensating harms that may arise from the operation of AI sandboxes or from the deployment of systems approved through such mechanisms.

The EU's AI Act, for example, requires the establishment of at least one AI regulatory sandbox by August 2026 (Article 53). These sandboxes are not permissive zones of deregulation, but rigorously managed environments in which AI systems are developed and tested under close supervision by national competent authorities. This direct, ongoing involvement not only affirms the state's institutional centrality in sandbox operations but also repositions it as a co-designer and risk manager, rather than a distant overseer. Indeed, the ability of authorities to waive administrative fines for sandbox participants who follow their guidance signals a collaborative, co-regulatory approach.

State co-responsibility is further underscored by its foundational role in defining the experimental parameters within which sandbox activities occur. In the case of AI, national and supranational authorities, including the EU AI Office, often provide the methodologies, benchmarks, and evaluative criteria used to assess sandbox projects (OECD 2023). In doing so, the state effectively sets the rules of the game

and shapes the underlying risk architecture. This is a constitutive act of governance, one that actively constructs the conditions of risk, rather than merely responding to them.

The justification for state involvement in compensation finds further support in established legal theories of state responsibility and duty of care. Legal scholarship indicates that states can be held responsible for failures to ensure that systems comply with legal obligations, including international law, even during the development and acquisition stages of any regulatory (Boutin 2023; Pacholska 2023). The principle of compliance-by-design, which calls for the integration of legal and ethical norms into technological development from the outset (Boutin 2023), implies a proactive duty on the part of the state to ensure that sandbox environments uphold legal standards. If the design of a sandbox is flawed, due to inadequate risk parameters, insufficient oversight mechanisms, or lenient approval criteria, then resulting harms may be traceable to the state's failure to institutionalize compliance at the meta-regulatory level. In such cases, the state could reasonably be held co-responsible for downstream impacts. This framework of shared responsibility imposes a clear duty on the state to create sandbox environments that are both reasonable and robust. In doing so, it plays a key structural role in preventing a regulatory race to the bottom, where jurisdictions compete to attract AI advancements through deregulation and minimal oversight.

This co-responsibility is further reinforced by the public benefit rationale that frequently underpins the promotion of regulatory sandboxes. AI is widely touted as offering transformative potential across sectors, from healthcare and education to public administration (Gonzales 2023). In promoting AI development through sandboxes, especially when targeting public-interest applications, the state assumes a dual role: it facilitates innovation while also endorsing the legitimacy of associated risks. This dual role parallels other state-led public health interventions, such as vaccine programs, where governments simultaneously promote collective benefits and assume responsibility for individual harms through state-backed no-fault compensation schemes (Mello 2008). Just as these programs acknowledge the trade-offs inherent in advancing public goods, so too should states recognize their obligation to address harms resulting from the innovation environments they construct and manage.

Circumventing Legal and Evidentiary Barriers for Compensation Purposes

A state-backed no-fault compensation model offers a particularly compelling framework for addressing harms resulting from AI systems developed and tested within regulatory sandboxes. Such a model is uniquely positioned to overcome the profound legal and evidentiary challenges associated with AI's distinctive characteristics, namely, its technical opacity, autonomous behavior, and the distributed nature of agency along complex AI value chains.

Traditional fault-based liability regimes, as mentioned, are ill-equipped to manage the difficulties posed by these features. Core legal doctrines, which require claimants to establish fault, foreseeability, and causation, falter in the face of black-box algorithms whose inner workings are neither

transparent nor easily interpretable. The “many hands” problem further complicates fault attribution, as harms may arise from a combination of actions or omissions across multiple actors, including developers, deployers, data providers, and regulators. As a result, victims may be unable to identify the source of harm or prove the necessary elements of liability under conventional legal standards (Lior 2024; Duffourc and Gerke 2023).

A state-backed no-fault compensation model fundamentally reorients the focus of redress away from establishing blame and toward recognizing harm and its causal relationship with an AI system. This shift offers a number of advantages, particularly salient to the sandbox context. First, it acknowledges the legitimate risks associated with promoting AI innovation through regulatory flexibility. Sandboxes are designed to facilitate experimentation with emerging technologies, which inherently involves uncertainty and the possibility of failure. A no-fault scheme aligns with this regulatory philosophy by offering victims a more accessible and less adversarial path to compensation, without undermining the broader objective of fostering innovation (Mello 2008).

Second, the historical analogy with other no-fault compensation programs in preventive healthcare, such as the U.S. National VICP, underscores the pragmatic logic of a state-backed no-fault approach. Just as the VICPs were established to balance public health innovation with fair compensation in the face of evidentiary uncertainty (Mello 2008), a similar rationale applies to experimental AI. The black box nature of AI systems echoes the evidentiary challenges in vaccine cases, where injuries may be real but causality is difficult to prove. In both contexts, state-backed no-fault schemes serve to ensure efficient, equitable, and timely redress without imposing prohibitive burdens on claimants or innovators.

Finally, by obviating the need to litigate fault, a state-backed no-fault system reduces the likelihood of prolonged, costly, and contentious legal disputes (Holm, Stanton, and Bartlett 2021; Marchisio 2021). This is particularly important within sandboxes, where swift resolution and continued regulatory engagement are essential. A compensation framework that prioritizes responsiveness and certainty over legal formalism is better suited to the experimental ethos of sandboxes and the evolving nature of AI technology.

Fostering Inclusive Innovation by Reducing SME Compliance Burdens

The creation of the U.S.'s VICP in 1986 illustrates how deficiencies in liability and compensation systems can erode market participation and jeopardize public welfare. By mid-1980s, a surge of lawsuits over the diphtheria–pertussis–tetanus (DPT) vaccine led to increased legal costs, prolonged litigation, and substantial jury awards (Klein P. 2014). By 1984, only one U.S. manufacturer remained producing the DPT vaccine, while other vaccine lines experienced similar attrition (Miller and Carroll 2022). This market collapse created acute shortages and risked the resurgence of vaccine-preventable diseases, prompting the urgent establishment of a state-backed, no-fault compensation mechanism to secure continued vaccine availability.

A parallel dynamic is emerging in the context of AI. As governments increasingly regard technological advancement of AI as integral to socio-economic development and global competitiveness, the burdens of liability and compliance risk might discourage participation, especially among SMEs and innovators in LMICs. A state-backed, no-fault compensation scheme holds transformative potential for building more equitable and inclusive AI innovation ecosystems. By relieving innovators, particularly SMEs and those based in LMICs, from the potentially prohibitive burdens of liability and compliance, such a scheme can enable broader participation in AI development, particularly within the experimental settings of regulatory sandboxes.

SMEs are widely recognized as key drivers of economic growth, yet they remain acutely vulnerable to external shocks, as demonstrated by the recent debt crisis and the COVID-19 pandemic (Sebnem et al. 2020). This vulnerability is further compounded by the prospect of liability for unforeseeable and uncontrollable harms caused by AI, a burden that is especially prohibitive for SMEs in LMICs. Laws such as the EU's AI Act impose stringent requirements related to bias mitigation, data governance, transparency, and oversight (Musch, Borrelli, and Kerrigan 2023). While these regulations aim to safeguard public interest, they translate into substantial legal, technical, and administrative burdens. For SMEs lacking the legal and financial infrastructure of larger firms, these obligations can be disproportionately stifling (Rasdi and Baki 2025). Penalties for non-compliance are severe, up to 7% of global annual turnover in the EU AI Act (Article 99), creating a powerful disincentive for small innovators to engage in high-risk, experimental AI development. Moreover, SMEs often lack the risk tolerance necessary to absorb potential liability for harms that might arise from testing novel technologies (Rasdi and Baki 2025; OECD 2021), even within ostensibly supportive regulatory environments, such as sandboxes.

This chilling effect is particularly pronounced for SMEs in LMICs, where structural limitations further compound the difficulties of AI innovation (OECD 2025a). These include chronic underinvestment, limited access to high-speed internet and AI-specific hardware, data scarcity, especially for local language and culturally relevant datasets, and a persistent talent gap exacerbated by brain drain (OECD 2025a). Additionally, unclear or underdeveloped intellectual property regimes add legal uncertainty to an already complex innovation landscape. In such contexts, the risk of being held financially liable for unintended harms caused by emerging AI systems, especially when such harms are difficult to foresee or prove, is not just a deterrent; it is often a non-starter.

While SME-oriented initiatives such as the EU's Startup Nations Standard, the EUSAiR regional sandbox program, and Singapore's AI Verify Foundation provide valuable compliance support, they do not extinguish the fundamental legal exposure to which SMEs remain subject when engaging in the sandbox environment. Nor do they address liabilities that arise *ex post*, once AI systems exit the sandbox environment and circulate in open markets, where latent defects or unforeseen safety risks may manifest. A state-backed no-fault compensation model would close this gap by decou-

pling access to redress from fault-based or defect-based attribution, thereby reallocating liability in a manner consistent with principles of equity, proportionality, and legal certainty, providing a safety net for innovators and end-users.

In this regard, state-backed no-fault compensation can function as a form of innovation subsidy. By socializing the costs of unforeseen harms, costs that would otherwise fall directly on small innovators, the state assumes a portion of the externalized risk that is inherent in experimental AI development. This risk-sharing mechanism encourages more substantive innovation by reducing the need for overly defensive legal or technical design strategies aimed solely at mitigating liability. It allows SMEs to redirect scarce resources toward more productive uses such as data acquisition, infrastructure investment, and capacity-building. In this way, a state-backed no-fault scheme does not simply reduce regulatory friction; it becomes a foundational component of AI-ready infrastructure that supports long-term ecosystem development.

Crucially, the goal of such a model is not to eliminate risk but to render it manageable, particularly for those actors whose participation is essential to a globally representative AI landscape. A carefully calibrated state-backed no-fault scheme provides the necessary assurance for SMEs to pursue ambitious and socially impactful innovations without the disproportionate fear of financial ruin. This fosters a more inclusive innovation ecosystem for a broader spectrum of actors, particularly resource-constrained SMEs and those in developing economies.

Enhancing Public Trust and Promoting Responsible AI Development

Public trust is the bedrock upon which the societal acceptance and successful adoption of AI technologies are built (Initiative 2025). Without this trust, the transformative potential of AI to improve public services, drive economic growth, and address complex societal challenges may remain unrealized (Aoki 2020). A lack of transparency regarding the functioning of AI systems, the outcomes of experimental programs like sandboxes, and the mechanisms for accountability can significantly undermine public confidence (Laplante and Kuhn 2022). Ethical AI principles, notably fairness, accountability, and transparency, are paramount in cultivating this trust (Aoki 2020). If the public perceives AI development, particularly during its experimental stages within sandboxes, as inherently risky and lacking adequate safeguards or clear avenues for redress when things go wrong, it can foster resistance and apprehension. Conversely, trust is nurtured when stakeholders are assured that their rights and interests are robustly protected. Sandboxes, while designed as controlled environments, are nonetheless sites of experimentation with technologies that have the potential for significant societal impact. Therefore, maintaining public confidence in these processes is vital for their legitimacy, continued operation, and ultimate success.

Redress mechanisms are regarded as a cornerstone of trustworthy AI governance and a tangible manifestation of accountability. Public trust is built and maintained through the establishment of clear, accessible, timely, and fair pro-

cesses to report grievances and seek remedies when AI systems cause harm. By design, a state-backed no-fault compensation scheme for sandbox-related AI harms serves as a powerful signal that the state and the innovators participating in these experiments are not abdicating responsibility for potential negative outcomes. Knowing that such a system exists can significantly bolster public assurance. This shifts the basis of public confidence: rather than relying solely on the often-unrealistic promise that no harm will occur from experimental technology (a “trust me” approach), trust is also built on the assurance that if harm does occur, there is a fair and accessible system for redress (a “we’ve got you covered” approach). This acknowledgment of potential fallibility, coupled with a demonstrable commitment to remedy, can be far more reassuring and build more resilient public trust than assertions of infallible safety, especially for technologies known for their complexity and unpredictability.

Furthermore, linking compensation schemes to the principles of responsible innovation reinforces their value. Responsible AI development inherently involves anticipating and mitigating risks, ensuring fairness, and maintaining accountability. A state-backed no-fault compensation system aligns with these principles by providing a structured mechanism to deal with unavoidable or unforeseen harms that can occur despite the best efforts and safety protocols within experimental settings (Holm, Stanton, and Bartlett 2021; Marchisio 2021). Regulatory sandboxes are themselves conceived as tools for fostering responsible innovation by allowing for learning in a controlled environment. By providing a safety net, a state-backed no-fault scheme can encourage a more responsible and open approach to experimentation. Innovators may be more forthcoming about potential risks and incidents if they know that a fair system for compensation is in place, rather than fearing punitive litigation that could stifle candid disclosure (Marchisio 2021). This fosters a culture of openness, learning, and continuous improvement, which is central to responsible AI development. The existence of such a scheme also subtly incentivizes the development of safer AI by clearly signaling the societal expectation that harms will be addressed (Marchisio 2021). The state’s role in promoting AI for the public good through sandboxes is thereby strengthened when coupled with a demonstrable commitment to compensate for associated harms, reinforcing the social contract aspect of technological innovation.

The establishment of a state-backed no-fault compensation framework can also shape the responsible development and deployment of AI within sandboxes. Best practices in AI governance are promoted by clear policies, robust stakeholder engagement, and comprehensive risk management, with transparency regarding AI’s use, limitations, and potential biases being essential. A compensation framework can serve as a vital accountability input, helping to understand causal connections between AI systems and harms, and enabling affected parties to pursue remedies. The knowledge that harms could trigger compensation may lead to more rigorous upfront risk assessments by sandbox participants and operators, and a greater emphasis on designing and implementing robust safety protocols during experiments. This encourages a proactive approach to risk management.

Moreover, a state-backed no-fault scheme can function as an invaluable feedback loop for promoting responsible AI development over time. Systems like the VICPs systematically track claims and compensated injuries (Mello 2008). Data collected through a similar scheme for AI sandbox harms, detailing the types of harms, the AI systems involved, and the specific contexts in which these harms occur, can provide crucial insights. This information can then be fed back to regulators, sandbox operators, and AI developers, highlighting recurring problems, identifying unforeseen risks, or pinpointing specific types of AI applications or experimental parameters that are proving to be particularly hazardous. This data-driven feedback can be used to refine sandbox protocols, update safety guidelines, inform necessary regulatory adjustments, and guide AI developers towards inherently safer design choices. In this manner, the compensation scheme transcends its function as a purely ex-post redress mechanism, contributing instead to ex-ante risk mitigation and the ongoing refinement of responsible AI practices. By ensuring that harm informs safer innovation, it creates a virtuous cycle in which technological advancement is coupled with ethical principles and a sustained commitment to public well-being, secured through a robust and transparent compensation framework.

Conclusion

This paper argues that existing liability frameworks in global AI regulatory sandboxes are ill-equipped to address the distinctive harms arising from complex and opaque AI systems. Firm-level liability models, often weakened by legal exemptions and the systemic nature of AI risks, fail to provide effective redress or sustain public confidence in AI governance. To address this accountability gap, the paper advances a state-backed, no-fault compensation scheme as a more legitimate and pragmatic alternative. By redistributing risk and enhancing access to justice, particularly for vulnerable actors and communities, such a scheme would promote inclusive innovation while reinforcing ethical and adaptive regulatory oversight.

Yet, the analysis remains primarily conceptual, drawing on scholarly critiques of the adequacy of traditional liability frameworks for addressing AI-specific risks. While empirical validation would strengthen these claims, relevant data are scarce given the nascent stage of AI regulatory sandboxes. Moreover, although AI has been deployed across multiple sectors, existing liability rules have not yet demonstrably impeded innovation or generated pressing real-world cases necessitating immediate reform. However, historical experience cautions that applying conventional liability regimes to rapidly evolving markets can produce serious negative consequences (Mello et al. 2010; Marchisio 2021; Di Gregorio et al. 2015). Rather than waiting for such outcomes to manifest in AI sandbox contexts, a more prudent course is to pre-emptively adapt the legal framework.

Acknowledgments

I am grateful to Mr. Duc-Lam Nguyen and Ms. Anh-Tho Tran for their thoughtful reviews and comments.

References

- Abiri, G.; and Huang, Y. 2023. A Red Flag? China's Generative AI Dilemma. *Harvard Journal of Law & Technology*, 37.
- Acciarri, H. A. 2015. Enforcement Deflection and Innovation. In *Law, Development and Innovation*, 39–63. Switzerland: Springer International Publishing AG.
- Ahern, D. 2021. Regulatory Lag, Regulatory Friction and Regulatory Transition as FinTech Disenablers: Calibrating an EU Response to the Regulatory Sandbox Phenomenon. *European Business Organization Law Review*, 22(3): 395–432.
- Ahern, D. 2025. The New Anticipatory Governance Culture for Innovation: Regulatory Foresight, Regulatory Experimentation and Regulatory Learning. *European Business Organization Law Review*, 26(2): 241–283.
- Alaassar, A.; Mention, A.-L.; and Aas, T. H. 2021. Exploring a New Incubation Model for FinTechs: Regulatory Sandboxes. *Technovation*, 103.
- Allen, H. J. 2020. Sandbox Boundaries. *Vanderbilt Journal of Entertainment and Technology Law*, 22(2): 299–321.
- Aoki, N. 2020. An Experimental Study of Public Trust in AI Chatbots in the Public Sector. *Government Information Quarterly*, 37(4).
- Arcuri, M. C.; Russo, I.; and Gandolfi, G. 2025. Productivity of innovation: The Effect of Innovativeness on Start-up Survival. *The Journal of Technology Transfer*, 50: 1111–1169.
- Arnal, J. 2025. AI at Risk in the EU: It's Not Regulation, It's Implementation. *European Journal of Risk Regulation*, 1–10.
- Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; and Herrera, F. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82–115.
- Baykal, F. Y. 2021. The polluter should... pay? *Heritage and Sustainable Development*, 3(1): 71–77.
- Bleeker, A. 2009. Does the Polluter Pay? The Polluter-Pays Principle in the Case Law of the European Court of Justice. *European Energy and Environmental Law Review*, 18(6): 289–306.
- Boura, M. 2024. The Digital Regulatory Framework through EU AI Act: The Regulatory Sandboxes' Approach. *Athens Journal of Law*, 10(3): 385–398.
- Boutin, B. 2023. State Responsibility in Relation to Military Applications of Artificial Intelligence. *Leiden Journal of International Law*, 36(1): 133–150.
- Brazil MoDITS. 2023. Regulatory Sandbox: Reference Guide. <https://www.gov.br/agu/pt-br/regulatory-sandbox-reference-guide.pdf>. Accessed: 2025-12-04.
- Buckley, R. P. 2023. Enabling a Balanced Proportional Approach to Innovation: Regulatory Sandboxes, Innovation Hubs, and Beyond. In *FinTech*, 164–176. Cambridge University Press.
- Buocz, T.; Pfothner, S.; and Eisenberger, I. 2023. Regulatory Sandboxes In The AI Act: Reconciling Innovation and Safety? *Law, Innovation and Technology*, 15(2): 357–389.
- Carneiro, G. 2024. *Moving Beyond the Innovation-Regulation Dichotomy: Regulatory Sandboxes for Artificial Intelligence Regulation*. LL.M. diss., Harvard School of Law, Harvard University, Boston, MA.
- Carvão, P.; Ancheva, S.; Atir, Y.; Jeloka, S.; and Zhou, B. 2025. Governance at a Crossroads: Artificial Intelligence and the Future of Innovation in America. Technical report, Mossavar-Rahmani Center for Business & Government, Harvard University.
- Chopra, S.; and White, L. 2011. *A Legal Theory for Autonomous Artificial Agents*. Ann Arbor : University of Michigan Press.
- Coeckelbergh, M. 2020. Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 26(4): 2051–2068.
- Coleman, J. L. 2002. Fault and Strict Liability. In *Risks and Wrongs*, 212–233. Oxford Academic.
- Cornelli, G.; Doerr, S.; Gambacorta, L.; and Merrouche, O. 2024. Regulatory Sandboxes and the FinTech Funding: Evidence from The UK. *Review of Finance*, 28(1): 203–233.
- Curtis, L. H.; and Schulman, K. A. 2006. Overregulation of Health Care: Musings on Disruptive Innovation Theory. *Law and Contemporary Problems*, 69(4): 195–206.
- Dauvergne, P. 2018. The Power of Environmental Norms: Marine Plastic Pollution and the Politics of Microbeads. *Environmental Politics*, 27(4): 579–597.
- Di Gregorio, V.; Ferriero, A.; Specchia, M.; Capizzi, S.; Damiani, G.; and Ricciardi, W. 2015. Defensive Medicine in Europe: Which Solutions? *European Journal of Public Health*, 25(3).
- Dickson, K.; Hinds, K.; Burchett, H.; Brunton, G.; Stansfield, C.; and Thomas, J. 2016. No-Fault Compensation Schemes: A rapid realist review to develop a context, mechanism, outcomes framework. Technical report, EPPI-Centre, UCL.
- Ding, W.; Abdel-Basset, M.; Hawash, H.; and Ali, A. M. 2022. Explainability of Artificial Intelligence Methods, Applications and Challenges: A Comprehensive Survey. *Information Sciences*, 615.
- Dixon, R. B. L. 2023. A Principled Governance for Emerging AI Regimes: Lessons From China, The European Union, and The United States. *AI and Ethics*, 3(3): 793–810.
- Duffourc, M. N.; and Gerke, S. 2023. The proposed EU Directives for AI Liability Leave Worrying Gaps Likely to Impact Medical AI. *NPJ Digital Medicine*, 6(1): 77.
- Durst, S.; and Henschel, T., eds. 2024. *Small and Medium-Sized Enterprise (SME) Resilience: Strategies for Risk and Crisis Management*. Cham: Springer.
- EU Parliament. 2024. EU Artificial Intelligence Act. <https://artificialintelligenceact.eu/article/57/>. Accessed: 2025-12-04.

- Evans, J. P.; and Watson, M. S. 2015. Genetic Testing and FDA Regulation: Overregulation Threatens the Emergence of Genomic Medicine. *JAMA: The Journal of the American Medical Association*, 313(7): 669–670.
- FCA. 2015. Regulatory Sandbox. <https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf>. Accessed: 2025-12-04.
- FCA. 2017. Regulatory sandbox lessons learned report. <https://www.fca.org.uk/publication/research-and-data/regulatory-sandbox-lessons-learned-report.pdf>. Accessed: 2025-12-04.
- Fonseca, A. T. d.; Sequeira, E. V. d.; and Xavier, L. B. 2024. Liability for AI Driven Systems. In *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, 299–317. Springer Nature.
- Franzoni, V. 2023. From Black Box to Glass Box: Advancing Transparency in Artificial Intelligence Systems for Ethical and Trustworthy AI. In *Computational Science and Its Applications - ICCSA 2023 Workshops, 2023, Vol.14107*, 118–130. Switzerland: Springer.
- Funkhouser, K. 2013. Paving the Road Ahead: Autonomous Vehicles, Products Liability, and the Need for a New Approach. *Utah Law Review*, 437.
- García-Portela, L. 2023. Backward-Looking Principles of Climate Justice: The Unjustified Move from the Polluter Pays Principle to the Beneficiary Pays Principle. *Res Publica*, 29(3): 367–384.
- Goldberg, J. C. P.; and Zipursky, B. C. 2016. The Strict Liability in Fault and the Fault in Strict Liability. *Fordham Law Review*, 85(2): 743–788.
- Gonzales, J. T. 2023. Implications of AI Innovation on Economic Growth: A Panel Data Study. *Journal of Economic Structures*, 12(1): 13–37.
- Gromova, E.; and Ivanc, T. 2020. Regulatory Sandboxes (Experimental Legal Regimes) for Digital Innovations in BRICS. *BRICS Law Journal*, 7(2): 10–36.
- Holm, S.; Stanton, C.; and Bartlett, B. 2021. A New Argument for No-Fault Compensation in Health Care: The Introduction of Artificial Intelligence Systems. *Health Care Analysis*, 29(3): 171–188.
- Initiative, D. 2025. Sandboxes for AI: Tools for A New Frontier. Technical report, DataSphere Initiative.
- Johnson, W. G. 2023. Caught in Quicksand? Compliance and Legitimacy Challenges in Using Regulatory Sandboxes to Manage Emerging Technologies. *Regulation & Governance*, 17: 709–725.
- Karnow, C. E. A. 2016. The application of traditional tort theory to embodied machine intelligence. In *Robot Law*, 51–77. Edward Elgar Publishing.
- Kataoka, S. 2019. Japan's New Policy for Testing Innovative Propositions for Growth with Government: The Regulatory Sandbox. *Banking & Finance Law Review*, 35(1): 187–196.
- Khosrowi, D.; Finn, F.; and Clark, E. 2024. Engaging The Many-hands Problem of Generative-AI Outputs: A Framework For Attributing Credit. *AI & Ethics (Online)*.
- Kilian, R.; Jäck, L.; and Ebel, D. 2025. European AI Standards - Technical Standardisation and Implementation Challenges under the EU AI Act. *European Journal of Risk Regulation*, 1–25.
- Kirilenko A., A.; Kyle S., A.; Samadi, M.; and Tuzun, T. 2017. The Flash Crash: High-Frequency Trading in an Electronic Market. *The Journal of Finance (New York)*, 72(3).
- Klein P., N. 2014. Licensed Pertussis Vaccines in the United States: History and Current State. *Human Vaccines & Immunotherapeutics*, 10(9): 2684–2690.
- Knight, B.; and Mitchell, T. 2020. The Sandbox Paradox: Balancing the Need to Facilitate Innovation with the Risk of Regulatory Privilege. *South Carolina Law Review*, 72(2): 445–475.
- Kolt, N. 2023. Algorithmic Black Swans. *Washington University Law Review*, 101(4).
- Koporcic, N.; Kukkamalla, P. K.; Markovic, S.; and Maran, T. 2025. Resilience of small and medium-sized enterprises in times of crisis: an umbrella review. *Review of Managerial Science*.
- Laplante, P.; and Kuhn, R. 2022. AI Assurance for the Public - Trust but Verify, Continuously. In *2022 IEEE 29th Annual Software Technology Conference (STC)*, 174–180. IEEE.
- Liao, F. 2020. Does China Need the Regulatory Sandbox? A Preliminary Analysis of Its Desirability as an Appropriate Mechanism for Regulating Fintech in China. In *Regulating FinTech in Asia. Perspectives in Law, Business and Innovation*, 81–95. Singapore: Springer.
- Lior, A. 2024. Holding AI Accountable: Addressing AI-Related Harms Through Existing Tort Doctrines. Technical report, The University of Chicago Law Review.
- Macleod, S.; Uberti, F.; and Kameni, E. 2025. No-fault Compensation Schemes for COVID-19 Vaccine Injury: A Mixed Bag for Claimants and Citizens. *Journal of Medical Ethics*, 51(2): 115–120.
- Marchisio, E. 2021. In Support of “No-fault” Civil Liability Rules for Artificial Intelligence. *SN Social Sciences*, 1(2): 54–54.
- Mardhatillah, S. R. 2023. A Concept of Liability for the Government's Negligence in the Prevention of Wildfires in Indonesia. *KnE Social Sciences*, 248–254.
- McCarthy, J. 2023. From Childish Things: The Evolving Sandbox Approach in The EU's Regulation of Financial Technology. *Law, Innovation and Technology*, 15(1): 1–24.
- Mello, M. M. 2008. Rationalizing vaccine injury compensation. *Bioethics*, 22(1): 32–42.
- Mello, M. M.; Chandra, A.; Gawande, A. A.; and Studdert, D. M. 2010. National Costs Of The Medical Liability System. *Health Affairs*, 29(9): 1569–1577.
- Miller, J. M.; and Carroll, R. S. 2022. An Informed Approach to Vaccine Hesitancy and Uptake in Children. *Delaware Journal of Public Health*, 8(1): 60–64.
- Mitchell, C. 2025. Conservative Think Tank Warns of ‘Overregulation’ of AI in Comments to White House. *Inside AI Policy Weekly Report*, 3(16).

- Moraes, T. 2025. Regulatory Sandboxes For Trustworthy Artificial Intelligence - Global and Latin American Experiences. *International Review of Law, Computers & Technology*, 39(1): 55–74.
- Morgan, D. 2023. Anticipatory regulatory instruments for AI systems: A comparative study of regulatory sandbox schemes. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 980–981. AAAI Press.
- Muftić, N. 2025. Bases of Liability. In *Artificial Intelligence and Tortious Liability*, 97–151. Switzerland: Springer.
- Mungwira, R. G.; Guillard, C.; Saldaña, A.; Okabe, N.; Petousis-Harris, H.; Agbenu, E.; Rodewald, L.; and Zuber, P. L. F. 2020. Global Landscape Analysis of No-fault Compensation Programmes for Vaccine Injuries: A Review and Survey of Implementing Countries. *PloS one*, 15(5).
- Musch, S.; Borrelli, M. C.; and Kerrigan, C. 2023. Balancing AI Innovation with Data Protection: A Closer Look at the EU AI Act. *Journal of Data Protection & Privacy*, 6(2): 135–152.
- Nordström, M. 2022. AI Under Great Uncertainty: Implications and Decision Strategies for Public Policy. *AI & Society*, 37: 1703–1714.
- OECD. 2021. The Digital Transformation of SMEs. Technical report, The Organization for Economic Co-operation and Development.
- OECD. 2023. Regulatory Sandboxes in Artificial Intelligence: OECD Digital Economy Papers. Technical report, The Organization for Economic Co-operation and Development.
- OECD. 2025a. Enhancing Access to and Sharing of Data in the Age of Artificial Intelligence. Technical report, The Organization for Economic Co-operation and Development.
- OECD. 2025b. Regulatory Sandbox Toolkit: A Comprehensive Guide for Regulators to Establish and Manage Regulatory Sandboxes Effectively. Technical report, The Organization for Economic Co-operation and Development.
- Pacholska, M. 2023. Military Artificial Intelligence and the Principle of Distinction: A State Responsibility Perspective. *Israel Law Review*, 56(1): 3–23.
- Piantadosi, G.; Marrone, S.; and Sansone, C. 2018. On Reproducibility of Deep Convolutional Neural Networks Approaches. In *Reproducible Research in Pattern Recognition (RRPR 2018)*, 104–109. Springer.
- Qiu, Y.; Yao, H.; Ren, P.; Tian, X.; and You, M. 2025. Regulatory Sandbox Expansion: Exploring The Leap from Fintech to Medical Artificial Intelligence. *Intelligent Oncology*, 1(2): 120–127.
- Ranchordas, S. 2021. Experimental Regulations for AI: Sandboxes for Morals and Mores. *Morals & Machines*, 1(1): 86–100.
- Rangone, N.; and Megale, L. 2025. Risks Without Rights? The EU AI Act’s Approach to AI in Law and Rule-Making. *European Journal of Risk Regulation*, 1–16.
- Rasdi, M. R.; and Baki, U. N. 2025. Navigating the AI landscape in SMEs: Overcoming internal challenges and external obstacles for effective integration. *PloS one*, 20(5): e0323249.
- Rodríguez de Las Heras Ballell, T. 2025. Mapping Generative AI Rules and Liability Scenarios in the AI Act, and in the Proposed EU Liability Rules for AI Liability. *Cambridge Forum on AI: Law and Governance*, 1(5).
- Saudi Arabia CST. 2023. CST Emerging Technology Regulatory Sandbox: Application Guidelines. Accessed: 2025-12-04.
- Scherer, M. U. 2016. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies. *Harvard Journal of Law & Technology*, 29(2): 353.
- Schütte, B. 2025. Liability for Damage Caused by AI—The AILD and Existing Liability Concepts. In *The Challenges of Artificial Intelligence for Law in Europe*. Springer.
- Sebnem, K.-O.; Gourinchas, P.-O.; Penciakova, V.; and Sander, N. 2020. *COVID-19 and SME Failures*. Washington, D.C.: International Monetary Fund, 2020.
- Sharma, A.; Ajadi, S.; and Beavor, A. 2020. Artificial Intelligence and Start-Ups in Low- and Middle-Income Countries: Progress, Promises and Perils. Technical report, GSMA Mobile for Development.
- Sherkow, J. S. 2022. Regulatory Sandboxes and the Public Health. *University of Illinois Law Review*, 357–410.
- Slattery, P.; Saeri, A. K.; Grundy, E. A. C.; Graham, J.; Noetel, M.; Uuk, R.; Dao, J.; Pour, S.; Casper, S.; and Thompson, N. 2024. The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence. Technical report, Massachusetts Institute of Technology.
- Slota, S. C.; Fleischmann, K. R.; Greenberg, S.; Verma, N.; Cummings, B.; Li, L.; and Shenefiel, C. 2023. Many Hands Make Many Fingers to Point: Challenges in Creating Accountable AI. *AI & Society*, 38(4): 1287–1299.
- Spence, A. M.; Behar, A.; and Jayadev, A. 2024. Artificial Intelligence in the Age of Uncertainty: Michael Spence in Conversation with Anurag Behar and Introduced by Arjun Jayadev. *Information Systems Frontiers*, 26(6): 1971—1988.
- The Alan Turing Institute. 2024. Risks Posed to Performance and Reliability. <https://aiethics.turing.ac.uk/modules/safety/?modulepage=risks-posed-to-performance-and-reliability>. Accessed: 2025-12-04.
- Thomas, C.; Roberts, H.; Mökander, J.; Tsamados, A.; Taddeo, M.; and Floridi, L. 2025. The Case for a Broader Approach to AI Assurance: Addressing “Hidden” Harms in the Development of Artificial Intelligence. *AI & Society*, 40: 1469—1484.
- Truby, J.; Brown, R. D.; Ibrahim, I. A.; and Parellada, O. C. 2022. A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications. *European Journal of Risk Regulation*, 13(2): 270–294.
- Tsai, C.-H.; Lin, C.-F.; and Liu, H.-W. 2020. The Diffusion Of The Sandbox Approach To Disruptive Innovation And Its Limitations. *Cornell International Law Journal*, 53(2): 261–296.
- Utah AI Policy Act. 2024. Utah Artificial Intelligence Policy Act. <https://le.utah.gov/2024/bills/static/SB0149.html>. Accessed: 2025-12-04.

- Wilczek, B.; Thäsler-Kordonouri, S.; and Eder, M. 2024. Government Regulation or Industry Self-regulation of AI? Investigating the Relationships Between Uncertainty Avoidance, People's AI Risk Perceptions, and Their Regulatory Preferences in Europe. *AI & Society*, 40: 3797–3811.
- Yang, Y. T.; and Reiss, D. R. 2023. Vaccine Injury Compensation in the United States. In *Vaccine Law and Policy*, 43–54. Springer International Publishing.
- Yoshikawa, J. 2019. Sharing the Costs of Artificial Intelligence: Universal No-Fault Social Insurance for Personal Injuries. *Vanderbilt Journal of Entertainment and Technology Law*, 21(4).
- Zarsky, T. 2016. The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology, & Human Values*, 41(1).
- Zetzsche, D. A.; Buckley, R. P.; Barberis, J. N.; and Arner, D. W. 2018. Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation. *Fordham Journal of Corporate & Financial Law*, 23(1): 31–103.
- Zhang, X.; Zhou, W.; Hu, W.; Zhou, S.; Hu, X.; and Yang, L. 2025. Can Ex Ante Conformity Assessment Regulations Contribute to Trustworthy Foundation Models? An Evolutionary Game Analysis from an Innovation Ecosystem Perspective. *Technology In Society*, 82.