Proceedings of the 4th International Conference on International Law and Legal Policy
DOI: 10.54254/2753-7048/2025.LD25883

The Balanced Role of AI Regulatory Sandbox in Innovation Incentives and Competing Safety Values

Zhichao Wang

College of Public Administration, Shandong Agriculture University, Taian, China
18866328363@163.com

Abstract: As artificial intelligence (AI) technologies rapidly penetrate key sectors such as healthcare, education, and public governance, traditional static, ex-post regulatory models can no longer keep pace with the risks and uncertainties arising from their swift evolution. Focusing on the "AI regulatory sandbox," the study clarifies its institutional logic, evaluates implementation outcomes, and identifies practical challenges, while exploring how the mechanism can balance innovation with the protection of the public interest. Employing a mixed approach of literature review and comparative case analysis, the study indicates that AI regulatory sandboxes provide a legally bounded, real-world environment that enables iterative testing, risk assessment, and rule refinement, thereby accelerating technological innovation and enhancing regulatory agility, and multi-stakeholder collaboration within the sandbox further strengthens the protection of user rights and social safety. Nonetheless, the mechanism remains challenged by major issues, including regulatory capture, inadequate transparency, fragmented cross-border coordination, and limited regulatory resources. As such, it recommends strengthening information disclosure and legal liability frameworks, establishing mechanisms for international mutual recognition and collaborative governance, and directing targeted investment toward regulatory talent and technical infrastructure.

Keywords: Regulatory Sandbox, AI Governance, Technology-Regulation Alignment, Risk and Compliance Mechanism, Institutional Innovation

1. Introduction

The rapid advancement of artificial intelligence (AI) technologies in key sectors such as healthcare, education, and public administration has exposed the limitations of traditional regulatory models, which are often too static and reactive to respond effectively [1]. In response, the AI regulatory sandbox has emerged as a promising governance innovation. Initially developed within the financial sector and subsequently adapted to the field of AI, the regulatory sandbox provides a controlled environment for iterative testing, risk assessment, as well as rule refinement. By fostering real-time collaboration, the sandbox enables developers and regulators to jointly build technical insight and regulatory understanding, thereby reinforcing a dynamic balance between innovation and oversight. However, unresolved challenges like transparency deficits, regulatory capture risks, jurisdictional fragmentation, and resource limitations continue to provoke discussion [2]. Through the analysis of relevant literature and case studies, this paper investigates the institutional advantages of the AI

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2025.LD25883

regulatory sandbox, including its potential to promote the parallel development of regulation and innovation, safeguard fundamental rights during testing phases, and boost institutional adaptability through multi-stakeholder collaboration. However, the implementation of the regulatory sandbox is not without controversy, as its legitimacy and effectiveness are challenged by risks like regulatory capture, cross-border inconsistencies, lack of transparency, and resource constraints. Accordingly, it seeks to address these core practical concerns by drawing on existing institutional experiences, and to provide targeted insights and policy recommendations for the future design and implementation of AI regulatory sandbox frameworks.

2. The institutional advantages of AI regulatory sandbox

2.1. Promoting the synchronous development of technology and regulation

The AI regulatory sandbox serves as a dynamic interface that enables the parallel advancement of technological innovation and regulatory capacity [3]. Despite innovation-oriented mandate set forth in Recital 139, the sandbox's greater institutional value, as revealed by empirical findings, lies in facilitating regulatory co-learning [4]. Under the sandbox framework, AI developers are required to disclose algorithmic structures and design rationales during the initial application phase. Meanwhile, regulators provide continuous oversight throughout the testing phase, and system operators offer iterative feedback on performance, risks, as well as anomalies. Through this dynamic exchange of information, regulators can monitor systems effectively and simultaneously gain sector-specific insights, which in turn bolsters their ability to design responsive and forward-thinking regulatory approaches. Instead of responding retrospectively, regulators can proactively contribute to refining standards, interpreting legal norms, and shaping enforcement logic, thus promoting the co-evolution of technological innovation and regulatory governance.

2.2. Protecting fundamental rights and social security within a controlled environment

By providing a legally bounded space for experimentation, the AI regulatory sandbox ensures that fundamental rights and public interests are not compromised during system trials [5]. In this context, it is imperative to recognize the potential implications for user rights and social security. Initiating experimental deployment in accordance with existing laws and established sandbox standards is therefore of paramount importance. Even when regulatory constraints are relatively limited during testing, there remains a risk of infringing upon users' legitimate rights and interests. However, as stipulated in Article 57(11) of the EU AI Act, regulatory authorities are empowered to promptly address risks arising from such infringements or to suspend the deployment of AI systems [6]. In addition, issuing non-enforcement letters or granting legal exemptions allows regulatory agencies to provide legal certainty for AI innovators and to define the limits of their rights and obligations within the controlled testing environment of the regulatory sandbox. This framework is instrumental in safeguarding the fundamental rights of users involved in the testing process, both those codified in current legislation and those implicating entitlement to compensation in cases where regulatory constraints are temporarily eased.

2.3. Enhancing institutional adaptability through multi-stakeholder collaboration

The sandbox model enhances regulatory adaptability by fostering multi-stakeholder cooperation, aligning legal oversight with the rapid and iterative nature of AI development [7]. To ensure both regulatory effectiveness and testing safety, it is crucial to engage a broad range of actors, including

DOI: 10.54254/2753-7048/2025.LD25883

AI innovators, regulatory bodies, and domain experts. Such collective engagement not only drives technological advancement but safeguards responsible and value-aligned innovation. For example, the UK's AI Airlock demonstrates how an AI regulatory sandbox can be established by convening multiple stakeholders, such as industry experts and supervisory authorities, to form a dedicated regulatory team. Once the team is in place, entities seeking entry into the sandbox work closely with it, supplying detailed information about their technologies and products. The team then analyses these submissions and subsequently reports its findings to the competent regulators. Through this deep interaction between AI developers and oversight bodies, the sandbox is expected to foster a more balanced trajectory for AI progress, as it seeks to reconcile the flexible, dynamic, and rapid evolution of technology with the comparatively reactive and slow-moving nature of traditional legislation. As such, the law becomes more adaptable, reducing the negative externalities that rigid rules can impose on innovation. This model reframes AI regulation from a paradigm of "absolute safety" to one of "reasonable cost," promoting the parallel advancement of legal supervision and AI technology.

3. Potential challenges and response strategies of AI regulatory sandbox

3.1. Transparency and public trust in algorithmic governance

Lack of transparency in AI regulatory sandboxes risks undermining public trust and triggering backlash against both the technology and its governance [8]. In recent years, the UK deployed an AI-based system to assess students' eligibility for examinations. However, the opaque nature of the algorithm and the undisclosed screening criteria gave rise to a crisis of trust, ultimately sparking widespread public protests. Consequently, ensuring transparency in the regulatory sandbox for AI systems, encompassing both the standards for admission and the behavior related to testing, exit, and market entry, has become imperative for both AI system innovators and the public to establish trust in and actively utilize the regulatory sandbox mechanism. However, as of now, the EU AI Act has made clear, in Article 72 of its preamble, the requirement for transparency [9]. On the one hand, this reflects a legal commitment to openness; on the other hand, transparency can be operationalized through publicly disclosing training data, providing summaries of model training methods, issuing compliance certificates, and similar practices. These measures allow the public to understand the basic working principles of AI systems, as well as their potential deviations. Thus, there is a greater likelihood that the public will accept and trust "successful" AI systems that have passed testing and entered the market, thereby advancing innovation in real-world applications.

3.2. Regulatory capture and the challenge of fair market access

The sandbox framework may inadvertently privilege large technology firms, raising concerns about regulatory capture and market inequality [10]. The concept of responsible innovation includes not only ensuring the safe use of AI systems for the public, but also maintaining a level playing field for developers of similar technologies. For example, large companies like OpenAI in the United States hold substantial financial resources, technical expertise, and social capital, which grant them a clear competitive advantage over small and medium-sized enterprises. Despite the explicit provisions in Articles 58(9)(v), 58(2)(iv), and 62 of the EU AI Act aimed at supporting the development of small and medium-sized enterprises, their practical impact remains to be fully assessed [11]. In addition, though some AI systems developed by competitive small and medium-sized enterprises such as China's emerging startup DeepSeek demonstrate strong potential, large companies are generally

DOI: 10.54254/2753-7048/2025.LD25883

more likely to create AI systems with greater "testing value," thus consistently securing a favorable position in the AI regulatory sandbox. Therefore, it remains necessary to examine whether large firms will persist in utilizing AI regulatory sandboxes with minimal regulatory constraints as a means of obtaining preferential policies, including legal exemptions. Meanwhile, regulatory bodies in jurisdictions such as the United States, the EU, and other national governments acknowledge the strategic role of large enterprises in advancing domestic AI development, boosting competitiveness, and strengthening their position in global markets. As a result, authorities may continue to extend preferential treatment to these corporations, raising concerns about potential regulatory capture. In response, some countries have taken steps to address this issue. For example, U.S. regulators must disclose decision-making processes, meeting records, and communications with stakeholders, all subject to public oversight. Besides, citizens' right to access government information is safeguarded by legislation such as the Freedom of Information Act.

3.3. Fragmentation of global regulatory standards and jurisdictional ambiguity

Divergent legal systems and regulatory thresholds across jurisdictions challenge the coherence and effectiveness of sandbox regimes on a global scale [12]. These structural and normative differences inevitably hinder the alignment and interoperability of AI regulatory sandboxes across countries and regions. As a result, limited cross-border collaboration and the lack of mutual recognition of testing outcomes further complicate global AI governance. In light of these challenges, AI innovators may prefer to apply for sandboxes in jurisdictions with less stringent testing requirements. However, a more pressing and unresolved issue lies in the allocation of responsibility for harm caused by AI systems that, after passing testing in one jurisdiction, are subsequently deployed across global markets. There is currently no clear consensus on whether liability should primarily rest with the developers who design and operate these systems, or with the regulatory authorities that approved their release. The situation becomes even more complex when sandbox host countries adopt lower safety thresholds than those of the jurisdictions where the systems are eventually used, potentially creating a gap in accountability. Without an internationally coordinated framework for cross-border responsibility, covering civil, administrative, and criminal dimensions where appropriate, firms may continue to exploit regulatory disparities, while affected individuals are left with limited avenues for redress. However, Article 2 of the EU AI Act explicitly defines its jurisdictional scope. It ensures that providers who place AI systems or general-purpose AI models on the EU market, or put them into use within the EU, are subject to the regulation, regardless of their geographical location. This includes providers based within the EU, in third countries, or elsewhere [13].

3.4. Resource constraints and the efficiency dilemma of regulatory agencies

The expanding scope and volume of AI sandbox applications place increasing strain on regulatory infrastructure and expert capacity [14]. As AI technologies proliferate, more developers are turning to regulatory sandboxes to test system performance, ensure compliance, and accelerate market entry. However, this surge in participation has revealed a growing tension, as limited regulatory resources, such as expert reviewers and oversight capacity, are under mounting pressure from the volume and complexity of sandbox applications. For example, in high-stakes applications like medical diagnosis, ensuring the accuracy of AI outputs usually requires expert oversight or secondary validation by specialists. As the scope of sandbox testing expands and more systems enter the pipeline, regulatory agencies face growing pressure not only in terms of available expert capacity but also in keeping pace with increasingly complex AI architectures. However, a well-designed AI

DOI: 10.54254/2753-7048/2025.LD25883

regulatory sandbox can serve as a critical platform for advancing emerging AI technologies, while simultaneously ensuring that their deployment remains safe, accountable, and aligned with the public interest. The need to allocate more time to the formulation of detailed and targeted regulatory policies, alongside increased investment in key areas such as expert staffing, technical infrastructure, and financial resources for regulatory agencies, is well justified when considered against the long-term societal and economic value that AI technologies are expected to generate.

4. Future prospects of AI regulatory sandbox

The AI regulatory sandbox is poised to become a cornerstone of next-generation AI governance, offering a dynamic mechanism to reconcile rapid technological advancement with the imperative of public interest protection [15]. In an era where both innovation and safety are held to increasingly high standards, the sandbox model holds significant potential to evolve into a globally coordinated, legally robust framework. Its future development will likely be driven by the institutionalization of more precise design elements. This includes a tiered approach to risk classification inspired by the EU AI Act's structure of minimal, limited, high, and unacceptable risk levels, dynamic oversight mechanisms that blend real-time monitoring with periodic algorithmic audits, and iterative policy development informed by performance metrics, public consultations, and post-deployment incident reporting. Each element should be contextually adapted to the AI application at hand. Through this refinement, the sandbox can function as both a space for innovation and a scalable regulatory tool. It helps preserve fair competition by ensuring proportional entry requirements for both start-ups and established firms, safeguards fundamental rights via ex-ante impact assessments and red-flag kill-switches, and supports responsible experimentation under clearly defined liability frameworks.

5. Conclusion

The AI regulatory sandbox represents a significant step forward in reimagining governance models for emerging technologies. By enabling iterative, context-sensitive, and collaborative regulation, the sandbox bridges the gap between innovation and oversight in ways that traditional regulatory approaches cannot. It serves not only as an experimental testing ground but also as a foundational tool for building regulatory capacity and public trust in AI deployment. Nonetheless, the success of this model depends on its institutional robustness and the ability of regulators to respond to inherent risks. The full potential of sandbox regimes depends on robust transparency mechanisms, effective safeguards against regulatory capture, the harmonization of international standards, and sustained investment in regulatory infrastructure.If these conditions are fulfilled, the AI regulatory sandbox will be well-positioned to achieve its dual mandate of fostering responsible AI innovation while safeguarding the fundamental interests of society.

References

[1] Bouderhem, R. (2024). Shaping The Future Of AI In Healthcare Through Ethics And Governance. Humanit Soc Sci Commun, 11, 416.
[2] Wang, T. (2024). The Path Selection Of AI Regulation: The Paradigm, Controversies, And Impact Of The EU Artificial Intelligence Act. European Studies, 42(3), 1-30.
[3] Truby, J., et al. (2022). A Sandbox Approach To Regulating High-Risk Artificial Intelligence Applications. European Journal Of Risk Regulation, 13(2), 270-294.
[4] Ruschemeier, H. (2025). Thinking Outside The Box?. In B. Steffen (Ed.), Bridging The Gap Between AI And Reality. AISoLA 2023. Lecture Notes In Computer Science, 14129. Springer, Cham.

DOI: 10.54254/2753-7048/2025.LD25883

[5] Yordanova, K. and Bertels, N. (2024). Regulating AI: Challenges And The Way Forward Through Regulatory Sandboxes. In H. Sousa Antunes, P. M. Freitas, A. L. Oliveira, C. Martins Pereira, E. Vaz de Sequeira, & L. Barreto Xavier (Eds.), Multidisciplinary Perspectives On Artificial Intelligence And The Law. Law, Governance And Technology Series, 58. Springer, Cham.
[6] European Union. (2024). Artificial Intelligence Act, Article 57. https: //artificialintelligenceact.eu/article/57/
[7] Gonzalez Torres, A.P. and Sawhney, N. (2023). Role Of Regulatory Sandboxes And MLOps For AI-Enabled Public Sector Services. Rev Socionetwork Strat, 17, 297-318.
[8] Pangandaman, H.K. (2024). Advantages And Challenges In Using AI (Artificial Intelligence) In Research: A Literature Review. Scope, 1709-1701.
[9] Svitych, O. (2025). Blind Transparency: A Critical Discourse Analysis Of The EU AI Act. Critical Policy Studies, 0(0), 1–17.
[10] Prastowo, R.D., Sensuse, D.I., Lusa, S. and Putro, P.A.W. (2024). Navigating Regulatory Sandbox Initiatives For Innovation Diffusion In Fintech Lending: A Systematic Review. Asian Management And Business Review, 4(2), 324-339.
[11] European Union. (2024). Supportive Measures For SMEs In The EU AI Act: A Larger Than Small Or Medium-Sized Task. KU Leuven CiTiP Blog. https: //www.law.kuleuven.be/citip/blog/supportive-measures-for-smes-in-the-eu-ai-act-a-larger-than-small-or-medium-sized-task/
[12] Gromova, E. (2020). Regulatory Sandboxes (Experimental Legal Regimes) For Digital Innovations In BRICS. BRICS Law Journal.
[13] Fuster, G.G. (2022). The EU Artificial Intelligence Act: A Closer Look At The Legal Framework For AI In Europe. European Journal Of Law And Technology, 13(2).
[14] Mustonen, S. (2024). Developing Ethical Guidelines For Artificial Intelligence In Healthcare [Master's Thesis]. University Of Oulu. https: //oulurepo.oulu.fi/handle/10024/50797
[15] Ristovska, T., et al. (2025). A Review On AI In Cybersecurity: Ethical Challenges And Regulatory Frameworks. Environment. Technology. Resources. Proceedings Of The International Scientific And Practical Conference, 2, 285-291.