The Transformation of AI Governance Paradigm from a Risk Regulation Perspective

Shuyuan Huang

Shanghai University of International Business and Economics, Shanghai, China
hshuyuan@hotmail.com

Abstract: The fast development of artificial intelligence (AI) technology has been the driving force behind the promotion of the development of new quality productive forces; however, it has also brought many complicated risks, including data breach, discrimination, and the loss of safety control. This has imposed severe limitations on the conventional legal governance model, which mainly focuses on ex-post accountability. This article, from the perspective of the theoretical model of risk regulation, will discuss the inherent limitations of the conventional legal governance model in addressing the risks of AI technology, including the problems of lag regulation, unclear liabilities, and inadequate compensation. Furthermore, this article will propose a new model of legal governance, namely the new governance pathway, which will shift from the conventional model of 'ex-post accountability' to 'risk regulation.' This paradigm shift will include three major elements: the shift of the focus of legal governance from 'ex-post remedy' to 'risk regulation,' the shift of the model of legal regulation from 'outcome-based liability' to 'whole-life-cycle governance,' and the shift of the subjects of legal governance from 'state-dominated regulation' to 'collaborative governance.' By establishing a model of risk regulation based on the Chinese context, this article will promote the healthy development of AI technology in a safe, trustworthy, and beneficial way.

Keywords: Artificial Intelligence; Risk Regulation; Paradigm Shift; Ex-post Accountability; Precautionary Principle

1 Introduction

We are in an AI-driven new wave of technological revolution and industrial transformation. China has over 100 large-scale AI models (≥10B parameters), with leading domestic open-source models like Tongyi Qianwen and DeepSeek. As AI applications expand, it has become a core driver of new-quality productive forces.[1]

However, AI's inherent risks—such as data privacy breaches, algorithmic discrimination, autonomous driving accidents,and generative AI disinformation—are complex and intertwined with technology and society. These uncertain, systemic, and concealed risks challenge traditional ex-post accountability-based legal frameworks and exceed traditional tort liability regulation.

Moreover, in more extreme cases, technical problems in AI systems, together with decision-making biases, could even directly jeopardize life safety and social stability.

S. Garcia-Esteban et al. (eds.), *Proceedings of the 2026 5th International Conference on Social Sciences and Humanities and Arts (SSHA 2026)*, Advances in Social Science, Education and Humanities Research 1014,
https://doi.org/10.2991/978-2-38476-577-5_96

To illustrate, decision-making problems in autonomous cars, such as algorithmic decision-making problems in unexpected road conditions, could even cause fatal accidents[2]. In the medical sphere, the risks brought about by AI-assisted medical diagnostic systems are usually more hidden, but their consequences are equally irreversible[3]. In the face of deepfakes, social trust is undermined, while the use of generative AI in critical sectors such as government, the judiciary, and medicine has created a fundamental dilemma, namely, how can victims prove fault, identify the responsible parties, and establish causation in traditional tort actions when victims are confronted with the difficulties of "algorithmic black boxes"?

This paper therefore aims to explore how to transform the legal governance paradigm when traditional ex post accountability is unable to address the endogenous risks brought about by AI, so that a new paradigm of AI-aligned risk prevention can be established to ensure safe, reliable, and socially aligned technological innovation.

2 Analysis of the Dilemma: Inadequacies of Traditional Legal Regulation and Risks of AI

The traditional system of regulation, based on ex-post accountability, case-based adjudication, and cause-based proof centered on redressing harms and attributing liability, is inadequate in addressing AI risks.

2.1 Slow Regulation and Ambiguous Liability

Regulation is a slow process, requiring decades or years, whereas AI is a fast-paced field with developments happening in a matter of weeks or months. For instance, the change from GPT 3.5 to GPT 4.0 happened in a year with tremendous progress. AI is penetrating the lives of people, the government, and the production sector, leading to tremendous and disruptive changes. Traditional regulation, in the form of case-based regulation or laws, is not capable of addressing AI risks, thereby making the entire system of AI risk management ineffective.

Generative AI combines elements of tech, services, and content creation, creating confusion between different types of legal issues[4]. In addition, there are a number of stakeholders in the field of AI, and the traditional system of liability is not applicable in this context. AI is autonomous and has a “black box,” and this makes it difficult to trace causation and liability in issues like AI-based medical diagnosis or copyright infringement in the training data used by AI.

2.2 Inadequacies of Traditional Regulation in Addressing AI Risks

The traditional system of regulation, based on the “harm and then seek redress and liability” process, is not effective in addressing AI risks, which are irreversible and non-compensable[5], such as fatal accidents caused by autonomous vehicles, discrimination caused by AI, and disinformation caused by generative. Australia’s 2016 Robodebt scandal is typical: an unmonitored automated welfare fraud detection system wrongly

targeted 500,000 low-income individuals, causing financial hardship, psychological trauma, and suicides.[6]

Personal information infringements are intangible, delayed, and diffuse, burdening individuals with heavy evidentiary requirements. While laws grant rights like algorithmic explanation and data deletion, rights-based regulation fails to solve core issues (algorithmic black boxes, biases, big data price discrimination).[7]This exposes structural flaws in traditional legal paradigms, making it urgent to introduce risk regulation theory and build an AI-specific governance framework.[8]

3 Logic, Theory, and Practice of Steering AI Governance Toward Risk Regulation Theory

The conventional legal approaches have been unable to mitigate the risks associated with AI, and hence there is a need to adopt the risk regulation theory in AI governance. This is a complete paradigm shift to the new AI-focused regulation theory, from "reacting to harm" to "preventing and controlling risks."

The three steps of the risk regulation theory in AI governance include risk identification (identifying potential AI risks), risk assessment (measuring the probabilities of the risks), and risk intervention (steps taken to mitigate the risks based on their levels of risk). The definition of risk is based on the potential harm, the possibility of the harm, and the severity of the harm.

Theoretically, the risk regulation theory is appropriate for AI risks because it is proactive in the prevention of harm, focuses on systems rather than individual cases, and focuses on the processes rather than the legal responsibility of the entity involved in the AI system, considering the uncertainty and system nature of AI risks. The incorporation of the risk regulation theory in AI systems is likely to reduce the risks associated with AI systems significantly.

The AI governance framework globally is based on the balance between competition and collaboration, with the EU, Canada (through the proposed Artificial Intelligence and Data Act, AIDA, risk-based high-risk regulation framework), and the US working towards the development of the risk regulation framework in AI systems. The EU proposed the AI White Paper and the AI Act, the first AI regulation in the world based on the risk regulation theory with clear high-risk regulation, with limited room for the entities involved to make choices[9]. This is the new mainstream in the regulation of technology risks worldwide. The AI regulation framework in China is also evolving in the direction of the AI governance paradigm based on the trends and the need to participate in the global digital regulation framework.

4 Pathway Construction: The Threefold Transition from "Ex-Post Accountability" to "Risk Regulation"

4.1 From Damage Remediation to Risk Prevention

The conventional legal model is characterized by its inherent reactiveness, responding only after the event to ensure the restoration of justice through the mechanisms of liability and compensation. Conversely, the risk regulation theory focuses on the proactive management of risk sources, based on the Precautionary Principle[10]. The major breakthrough is the shift from ex-post remediation of damage to ex-ante prevention, addressing the irreversible risks of artificial intelligence (AI) through "Safety by Design" and the incorporation of risk control in the early stages of technological development. For instance, the Chinese "Interim Measures for the Management of Generative Artificial Intelligence Services" require AI services with the ability for public opinion mobilization and social mobilization to conduct security assessments and fulfill the requirement of algorithm filing[11], leading Baidu and Alibaba to set up ethical review systems to intercept risks before they occur. The "2025 AI Safety Governance Framework" (Version 2.0) also concretized the principle of "trustworthy application and loss of control prevention,"[12] requiring the incorporation of value constraints in technological processes to ensure the alignment of AI with human values from the start.

4.2 Process Control: From Outcome Accountability to Full-Lifecycle Supervision

In response to the systemic nature of AI risks, which is also hidden, the risk governance paradigm introduces full-lifecycle management, which replaces the traditional focus on the final outcome with the dynamic management of the evolution of technology. Such an oversight approach covers the entire lifecycle of AI, from development, deployment, application, and iteration, to address the black box problem.

Scholars have suggested the introduction of a three-stage autonomous system, namely pre-event, during-event, and post-event, for the comprehensive oversight of AI risks. Such an approach introduces the dynamic full-lifecycle risk governance mechanism. Relevant regulatory instruments also support this risk governance orientation. For instance, the Provisions on the Administration of Internet Information Service Algorithm Recommendation requires that internet information service providers regularly review and assess their algorithms, models, data, and results. Moreover, the AI Safety Governance Framework (Version 2.0) extends the scope of AI risk governance to cover "application-derived security risks," including the regulation of the social impacts of AI, as well as the phenomenon of deep fakes and the bias of algorithms.

At the enterprise level, technology companies are leveraging automated systems for AI-powered content moderation. Baidu continuously monitors text, videos, and images to flag risks, block pushed content, and filter search results. Its framework consists of three steps: early monitoring, proactive interception, and post-event tracking[13]. Alibaba Cloud and iFlytek use "ethical labeling preprocessing" to filter sensitive data

during AI model training. These compliance checks are built directly into the development cycle, ensuring risk management keeps pace with technological development.

4.3 Multi-Stakeholder Governance: From Government Solo to Collaborative Co-governance

Co-governance involves governments, businesses, civil society, and the public working together through clear rights and obligations to achieve coordinated cooperation. AI governance is too complex for a single entity, so a multi-stakeholder ecosystem is the only option.

Unlike the traditional ex post litigation model in which the government acts as the sole "public arbiter,"[14] risk regulation adopts a traditional ex post government-centric approach and extends the value chain to create synergy between the state, the market, and society. Based on the core concepts of "government leadership, platform accountability, and societal participation," risk regulation clarifies the roles of stakeholders and helps overcome the limitations of single-entity oversight.

Governments set red lines, conduct inspections, and impose penalties. China has established an initial regulatory framework by enacting laws such as the Personal Information Protection Law and the Interim Measures for the Management of Artificial Intelligence Services[15]. The government also establishes a comprehensive framework, refines laws, issues ethics guidelines, conducts critical risk assessments, and coordinates the development of AI to prevent superficial development.

For companies, risk management agencies conduct self-assessments, record-keeping, and reporting. Most companies, including Baidu and Alibaba, have their own ethics committees. Industry organizations establish technical standards and conduct compliance certification.

Society, including the media, monitors for the public interest through labeling, complaints, reporting, and litigation. Multi-stakeholder platforms gather diverse needs and insights from relevant parties to achieve ethical consensus and oversight.

Through these structures, the government provides a comprehensive framework, companies implement self-regulation, and society participates, enabling agile governance. Governance is now able to adapt to the evolving challenges of AI. In short, a forward-looking perspective, process management, and the engagement of diverse stakeholders have contributed to establishing risk regulations that help achieve a global security-development balance, as reflected in the AI Safety Governance Framework (Version 2.0).

5 Conclusion

In the AI era, traditional governance models are increasingly inadequate. Building comprehensive whole-process supervision is therefore not just a theoretical preference, but an inevitable response to the era’s challenges. The metamorphosis involves moving from a perception of AI as a tool to a perception of it as a self-sustaining, evolving phenomenon that is transforming society, warranting particular attention and

regulation. Transitioning to a new paradigm of AI governance is a process, not a static goal. The pace of technological advancement is rapidly evolving, thus altering the risks associated with AI, making this a collaborative effort involving law, technology, and ethics.

We propose an agile, inclusive, and resilient global governance system that involves multi-stakeholder engagement, transcends national, cultural, and jurisdictional boundaries, and is a dynamic process of dialogue, experimentation, and learning from each other. This is a difficult process, but by embracing the transformative potential of AI, we can ensure human agency and harness the transformative potential of AI to ensure collective safety, well-being, and dignity.

References

1. Liu Jinrui & Li Mengqi, "The EU's Risk Regulation System for Large-Scale AI Models and Its Implications for China," China Information Security, no. 3 (2025): 78–82.
2. Tesla car that killed Seattle motorcyclist was in “Full Self-Driving” mode, police say. CNN, July 31, 2024. Available at: https://edition.cnn.com/2024/07/31/tech/tesla-full-self-driving-mode-seattle-motorcyclist-killed (last accessed October 5, 2025).
3. Tesla car that killed Seattle motorcyclist was in “Full Self-Driving” mode, police say. CNN, July 31, 2024. Available at: https://edition.cnn.com/2024/07/31/tech/tesla-full-self-driving-mode-seattle-motorcyclist-killed (last accessed October 5, 2025).
4. Zhang Linghan, "Legal Positioning and Tiered Governance of Generative Artificial Intelligence," Modern Law Science 45, no. 4 (2023): 126–141.
5. Mao, F. Robodebt: Illegal Australian Welfare Hunt Drove People to Despair. BBC News, July 7, 2023. Available at: https://www.bbc.com/news/world-australia-66130105.
6. Chen Junxiu & Xu Yuqin, "Data Pollution Risks Triggered by Large Language Models in Artificial Intelligence and Regulatory Approaches," Journal of Dalian University of Technology (Social Sciences)
7. Yang Zhangwen, "Ideological Attributes and Risk Regulation of ChatGPT-like Generative Artificial Intelligence," Inner Mongolia Social Sciences 45, no. 1 (2024): 57–64.
8. Zeng Xiong, Liang Zheng & Zhang Hui, "The Construction of China’s AI Risk Governance System and Theoretical Explanation Based on Risk Regulation Models—Taking Generative Artificial Intelligence as an Example," International Economic Review, no. 4 (2025): 131–152+7.
9. Chen Jidong, "Risk-Based Artificial Intelligence Governance," Research on Rule of Law, no. 5 (2023).
10. Interim Measures for the Administration of Generative Artificial Intelligence Services, promulgated by the Cyberspace Administration of China, implemented on August 15, 2023, art. 17.
11. Interim Measures for the Administration of Generative Artificial Intelligence Services (promulgated by the Cyberspace Admin. of China, Aug. 15, 2023) art. 17.
12. UNESCO. (2005). The Precautionary Principle. World Commission on the Ethics of Scientific Knowledge and Technology.

13. Cai Cuihong, "Cyberspace Governance in the Age of Artificial Intelligence: Technology Embedding, Human Leadership, and Human-Machine Collaboration," People's Forum · Academic Frontiers, no. 13 (2025): 34–48.
14. Yang Jianwu & Luo Feiyan, "Risk Challenges and Inclusive Prudential Regulation of False and Harmful Information in Generative Artificial Intelligence," Administration and Law, no. 7 (2025): 75–89.
15. Meng Fanqian & Wang Zhongyang, "Legal Risks and Governance of ChatGPT from the Perspective of Personal Information Protection," Media, no. 3 (2024): 51–54.