AI Law Competition and Collingridge Dilemma's Policy Dimension

Shen Wei
Professor of Law
Shanghai Jiao Tong University KoGuan Law School

The rapid advancement of artificial intelligence (AI), coupled with its inherent risks, necessitates the establishment of a robust global AI governance framework. Distinguished by its unique characteristics including rapid iterative development, black-box operational opacity, cross-sectoral scalability, and cascading systemic vulnerabilities, AI exacerbates the Collingridge Dilemma, a foundational paradox governing the regulation of emerging technologies, which presents a two-pronged governance impasse.

The Collingridge dilemma encapsulates the core challenge of imposing effective social control over technological innovation, and offers a cogent analytical framework for interpreting the regulatory predicament attendant to AI development. In the nascent stages of AI evolution, its societal ramifications are notoriously unforeseeable; yet once the technology matures and becomes deeply entrenched within societal infrastructures, its negative externalities evolve into increasingly intractable barriers to regulatory oversight and governance intervention.

Conventional doctrinal analyses predominantly situate the Collingridge Dilemma within the technical domain, emphasizing the inherent uncertainty and limited controllability of emerging technologies. However, this perspective is analytically incomplete. The dilemma is not only anchored in technological indeterminacy, but also is perpetually replicated and even amplified at the policy level by cognitive biases and strategic conduct among sovereign states. This phenomenon manifests principally in the goal distortion inherent to national AI legislative frameworks, thereby engendering novel risks to effective social control over AI systems. Accordingly, the Collingridge Dilemma manifests along dual dimensions: a technological dimension and a policy-induced dimension.

In recent years, states have implemented a litany of interventionist measures under the expansive guise of national security—a rubric that now encompasses supply chain resilience, data governance, technological risk mitigation, and semiconductor security. Such measures include export control regimes, foreign investment screening, industrial subsidization schemes, data security reviews and economic sanctions. While formally

predicated on risk prevention and mitigation, these policies have in practice fostered to a de facto decoupling across trade, investment, financial and data governance regimes. This policy trajectory not only ex ante constrains the trajectory of technological innovation but also ex post distorts cross-jurisdictional regulatory learning, thereby exacerbating the Collingridge Dilemma along its policy dimension.

A core catalyst underpinning this policy-layered dilemma resides in the inherent subjectivity of risk and security assessments. Existing international law imposes only narrow constraints on the manner in which sovereign states define and operationalize these malleable concepts. Consequently, decision-making is frequently driven by cognitive biases and strategic competitive imperatives rather than rigorous, objective risk evaluation. In pursuit of maximizing relative strategic and economic gains, states tend to adopt zero-sum approaches to AI governance, which risk precipitating collective irrationality and systemic institutional fragmentation at the global level.

The international legal architecture governing AI remains deeply fragmented, with major powers engaged in competitive rather than collaborative rule-making endeavors. At the international level, AI governance is characterized by greater strategic competition than multilateral cooperation, deeper institutional fragmentation than normative integration, and more pronounced soft-law attributes than hard-law binding obligations. The evolution of international law applicable to AI has reached a regulatory impasse, standing at a decisive normative juncture. Major powers leverage domestic legislative instruments to entrench their technological superiority, curtail the technological advancement of geopolitical competitors, and further widen pre-existing digital and AI governance divides.

In particular, the United States and the European Union, drawing on their dominant market and technological leverage, are competing for normative primacy in the formulation of global AI rules and standards. Through the extraterritorial spillover effects of their domestic legal regimes, these actors shape the regulatory models and normative trajectories of AI governance adopted by third states. They further embed human rights, democratic governance and allied normative priorities into the global AI rule-making process, ostensibly to preserve their technological hegemony and supply chain security. This dynamic has contributed to the escalating securitization and instrumental weaponization of AI-related regulatory norms.

To redress the policy-induced amplification of the Collingridge Dilemma, this study advances a three-pronged remedial framework. First, the international legal regime governing AI must be strengthened to curb discretionary and unilateral state conduct that undermines multilateral governance. Second, the cultivation of transnational

technical epistemic communities can foster normative coordination and mitigate security-driven policy divergence across jurisdictions. Third, sovereign states should develop proactive rule-making capacity to shape emerging global AI norms, thereby securing a more equitable and influential role in the construction of inclusive global AI governance.