The Boundaries and Hierarchy of AI Explainability from a Legal Perspective

Luming Wang[1], Yuchen Wu[2]

[1]Zhongnan University of Economics and Law, [2]East China University of Political Science and Law

Abstract

The “ black box ” nature of AI algorithms presents a profound challenge to the foundational principles of modernlegal systems, specifically the attribution of liability andprocedural justice. This article addresses the legalboundaries and implementation mechanisms of explainability by proposing an integrated framework that combines a hierarchical model with indirect methods. We argue that the duty to explain must be governed by the principle of proportionality, dynamically calibrating its scope to the risk level of the algorithm.

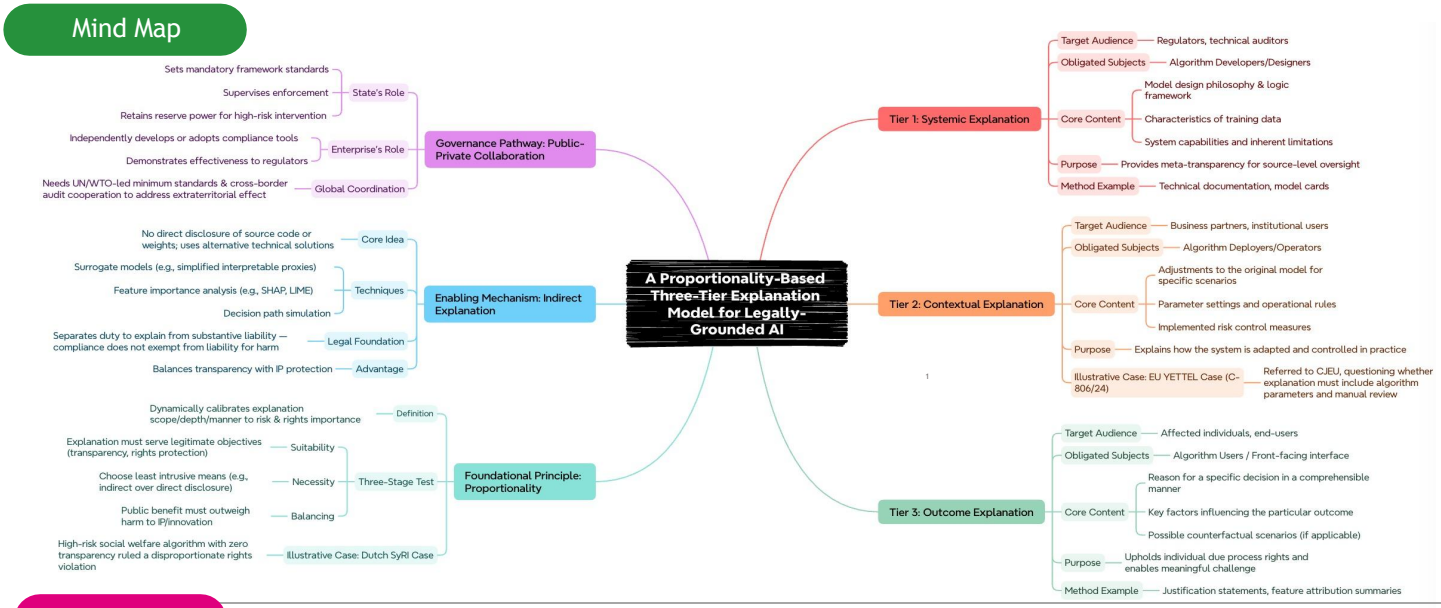
To resolve the inherent tension between transparency and intellectual property, an indirect explanation mechanism is proposed, utilizing alternative technical solutions for compliance. At the governance level, a public-private collaborative path is advocated, wherein the state sets mandatory framework standards and supervises enforcement, while enterprises independently develop compliance tools. This research provides critical theoretical support for advancing AI legislation and constructing a secure and self-governing AI governance framework for China.

Methodology and Results

Adopting literature analysis, case studies (e.g., SyRI, YETTEL) and theoretical construction, the study centers on the proportionality principle. It builds a three-tier explanation model and indirect mechanism, clarifies multi-subject obligation boundaries, and explores public-private collaborative governance and cross-border compliance coordination to balance transparency, innovation and other interests.

Algorithmic black boxes systematically challenge legal liability attribution and procedural justice, making AI explainability a legal necessity. It requires hierarchical, indirect mechanisms and the proportionality principle to define explanation boundaries, relying on public-private collaboration and global rule coordination. This provides theoretical support for China’s AI legislation and global trustworthy AI governance.

Mind Map



Final Remarks

Algorithmic opacity poses systemic threats to legal liability attribution and procedural justice, making AI explainability an indispensable legal cornerstone for trustworthy technology. This research addresses the gap by constructing a proportionality-guided framework—integrating a three-tiered explanation model, indirect compliance mechanisms, and public-private collaborative governance—to balance transparency, innovation, and intellectual property protection.

Looking ahead, legal frameworks must evolve dynamically with technological advancements, while interdisciplinary talent cultivation and strengthened auditing systems are critical for implementation. Globally, aligning with international standards via UN/WTO-led cooperation will mitigate cross-border compliance dilemmas. Ultimately, this work lays theoretical and institutional groundwork for China’s AI legislation and contributes to the global co-governance of accountable, rule-of-law-compatible AI.

Acknowledgements and References

Web sources:

Balakrishnan, A. 2024. Ethical and Legal Implications of AI Judges: Balancing Efficiency and the Right to Fair Trial. PhD dissertation, Utrecht University. chttps://studenttheses.uu.nl/handle/20.500.12932/48242cNaama, A., and Bourane, A. 2025. Law in the Age of Artificial

Intelligence: Challenges and Prospects. Lex Localis - Journal of Local Self- Government. https://www.proquest.com/openview/4f07684d458b51339650f23fb62138e4/1?pq-origsite=gscholar&cbl=55210.

Intelligent Rule of Law Research Institute, Tsinghua University. 2025. The EU Court of Justice Accepts the First Case Concerning the Application of the 'AI Act', Involving an Automated Calculation System for Contractual Phone Card Fees. WeChat Official Account Platform. https://mp.weixin.qq.com/s/19wdXHn1M3ye_3f2Zbuv0Q. Accessed: 2025-06-19.

Literature Articles:

Arefin, S.; Zannat, N. T.; and Global Health Institute Research Team United States. 2025. Securing AI in Global Health Research: A Framework for Cross-Border Data Collaboration. Clinical Medicine And Health Research Journal 5(2). doi.org/10.18535/cmhrj.v5i02.457

De Bruyne, J.; and Ooms, 2025. Tort Liability and Artificial Intelligence: Some Challenges and (Regulatory) Responses. In Artificial Intelligence and the Law, edited by De Bruyne, J.; and Vanleenhove, C., 158 - 173. Cambridge: Cambridge University Press. doi.org/10.1017/9781009367783.011.

Ding, X. 2022. Trust-Based Automated Decision-Making: Rethinking the Principle and Restructuring the Institution of the Right to Algorithmic Explanation. China Legal Science (1): 99- 118. doi.org/10.14111/j.cnki.zgfx.20211230.001.

Kahl, P. 2025. The Third Enclosure Movement: Rethinking AI Regulation through a Taxonomy of Restricted Knowledge. Preprint. Lex et Ratio Ltd. Retrieved from doi.org/10.5281/zenodo.17542063

Contact: Luming Wang | wangluming@stu.zuel.edu.cn, Yuchen wu | aron.wu2005@ecupl.edu.cn