Financial Risk and Management:
An International Journal
Volume 4, Number 1, 2025

AI Governance Mechanisms in Modern Financial Risk Management Systems

Shelli Melita[1]

[1] Universitas Diponegoro, Semarang, Indonesia

Abstract

Article history:
Received: January 11, 2025
Revised: February 23, 2025
Accepted: April 10, 2025
Published: June 30, 2025

Keywords:
AI Governance,
Explainable AI,
Financial Risk Management,
Model Risk Management,
Risk Governance.

Identifier:
Zera Open
Page: 42-55
https://zeraopen.com/journal/frmij

This article investigates how artificial intelligence (AI) is governed within modern financial risk management systems, asking under what conditions AI-based models can enhance, rather than undermine, risk control. The study conducts a systematic literature review of peer-reviewed articles published between 2019 and 2024, focusing on governance mechanisms associated with AI applications in credit, market, liquidity, and operational risk management. The reviewed evidence shows that AI governance is still emerging and uneven, with most institutions extending traditional model risk management frameworks while struggling to address data drift, feedback loops, bias, and systemic effects. The article discusses the literature through a narrative and thematic synthesis that maps governance practices across three main dimensions: AI-specific model risk management, the use of explainable AI as a governance tool, and organizational and ethical mechanisms such as human-in-the-loop oversight and legal accountability. The main findings highlight fragmented implementation, limited empirical evaluation of effectiveness, and the need for more coherent, testable governance architectures.

1. Introduction

Artificial intelligence is becoming a core component of modern financial risk management systems, supporting credit scoring, fraud detection, market and liquidity risk modeling, stress testing, and early-warning systems. Recent evidence shows that machine learning models can process high-dimensional and unstructured data, capture nonlinear risk patterns, and outperform traditional statistical techniques in predictive accuracy and responsiveness, especially in digital and internet-based financial environments (Ahmed et al., 2022; Tian et al., 2024). Systematic and bibliometric reviews document a rapid expansion of AI and machine learning applications across banking, capital markets, and FinTech after 2019, with particular emphasis on credit risk, operational risk, and real-time monitoring (Dianti, 2023; Fahrezi, 2024). These developments position AI not only as a set of analytical tools but also as critical infrastructure for contemporary financial risk management.

However, the integration of AI into high-stakes financial decisions introduces new types of risk that traditional risk management frameworks are not fully equipped to handle. Studies on algorithmic decision-making in consumer credit highlight concerns about opaque model logic, discrimination, and unequal outcomes for vulnerable borrowers, raising questions about the economic and normative legitimacy of AI-driven decisions (Sargeant, 2023). Work on AI ethics and systemic risks in finance argues that highly interconnected AI systems can amplify market volatility and create feedback loops that are difficult to detect and govern with existing tools (Svetlova, 2022). Research on ethical and legal dimensions of AI in financial services likewise points to gaps in accountability, transparency, and fairness, especially in credit scoring, robo-advisory, and financial inclusion initiatives (Uzougbo et al., 2024; Yang & Lee, 2024). Together, this literature suggests that AI can both mitigate and create financial risks, depending on how it is governed.

Against this backdrop, AI governance has emerged as a key concept linking technical model controls, organizational practices, and regulatory expectations. General AI governance frameworks propose integrated approaches that classify AI-related risks and translate them into guidelines for oversight, documentation, and control across the AI lifecycle (Wirtz et al., 2020; Wirtz et al., 2022). In the specific context of financial services, Floridi et al. (2020) outline a governance framework that embeds fairness, accountability, and human-in-the-loop oversight into financial risk management processes, emphasizing the need to align innovation with robust control architectures. Yet these frameworks are often developed from public-sector or general AI perspectives and only partially engage with the concrete mechanisms used by financial institutions to manage AI-related model risk, operational risk, and systemic risk.

A growing body of reviews maps AI applications in finance or surveys general AI governance principles, but few studies systematically integrate these strands to examine how AI governance mechanisms are conceptualized and implemented within modern financial risk management systems (Ahmed et al., 2022; Dianti, 2023; Tian et al., 2024). In particular, there is limited synthesis of how tools such as model risk management frameworks, explainable AI techniques, data governance, human-in-the-loop structures, and legal accountability mechanisms are used in practice to govern AI models that shape financial risk profiles. This article addresses that gap by conducting a systematic literature review of peer-reviewed studies published between 2019 and 2024 on AI governance mechanisms in financial risk management. By bringing together evidence on governance practices, risk taxonomies, and institutional arrangements, the study aims to clarify the current state of knowledge, identify converging design principles, and highlight unresolved tensions that should inform future research and regulation of AI-enabled financial risk management systems.

2. Literature Review

The existing literature on artificial intelligence in financial risk management is dominated by studies that document applications of machine learning across credit, market, liquidity, and operational risk, while only recently beginning to connect these applications to formal governance mechanisms. Bibliometric and systematic reviews show that research on AI and finance has grown rapidly since 2019, with a strong concentration on credit risk scoring, fraud detection, and internet-based financial risk management (Ahmed et al., 2022; Tian et al., 2024). Other systematic reviews emphasize how AI and machine learning improve predictive accuracy and early warning capabilities in financial risk management, but tend to treat governance and model risk as peripheral rather than central themes (Dianti, 2023; Fahrezi, 2024). Taken together, these studies establish the technical potential of AI in risk management, yet they provide only limited insight into how financial institutions structure governance arrangements around these systems.

A second strand of work develops broader AI governance frameworks that, while often not finance-specific, set out key principles and control mechanisms relevant for high-stakes financial applications. Governance frameworks proposed by Wirtz et al. (2020, 2022) and related AI governance reviews conceptualize responsible AI through risk classification, lifecycle controls, and guideline-based oversight, stressing fairness, accountability, transparency, and human oversight as core design principles. In financial services, Floridi et al. (2020) argue for a "control-by-design" approach that embeds fairness, human-in-the-loop decision rights, and robust documentation into risk management architectures. Complementary studies on ethical and legal aspects of AI in finance highlight issues of discrimination, opacity, accountability gaps, and regulatory uncertainty in credit scoring, robo-advisory, and financial inclusion contexts (Svetlova, 2022; Sargeant, 2023; Uzougbo et al., 2024; Yang & Lee, 2024). This conceptual and normative work provides a rich vocabulary for AI governance, but often remains detached from the concrete practices and model risk processes used in financial institutions.

More recent research begins to bridge AI governance and financial risk management by focusing on model risk and institutional implementation. Knowledge mapping work on model risk in banking shows that model risk management has evolved into a distinct research field, with growing attention to the implications of machine learning and AI models for risk identification, validation, and regulatory compliance (Cosma et al., 2023). Practice-oriented analyses of AI model risk illustrate how existing model risk management frameworks can be adapted to address data quality, validation, monitoring, and ethical concerns specific to AI models, while calling for enhanced governance policies and model classification schemes (Souza, 2023). Comparative studies of AI in risk management across jurisdictions further reveal how differences in infrastructure, regulation, and risk culture shape AI adoption and governance in banking sectors, underscoring the importance of supportive regulatory policies and institutional capacity (Nnaomah et al., 2024). Yet even this emerging literature tends to treat governance mechanisms piecemeal, leaving a gap for a systematic synthesis of how model risk management, explainability, data governance, human oversight, and legal accountability are jointly configured as AI governance mechanisms within modern financial risk management systems.

3. Methods

This study employs a systematic literature review approach to synthesize current knowledge on AI governance mechanisms in modern financial risk management systems. The review focuses on peer-reviewed journal articles published between 2019 and 2024 to capture the most recent wave of AI adoption and regulatory discussion in finance. Relevant studies were identified through structured searches in major academic databases such as Scopus, Web of Science, ScienceDirect, and Google Scholar, using combinations of keywords including "artificial intelligence", "machine learning", "AI governance", "model risk management", "financial risk management", "banking", and "regulation". The search was restricted to English-language articles. Conference papers, theses, book chapters, non peer-reviewed material, and purely technical papers without any discussion of governance, control, or institutional aspects were excluded. Reference lists of core articles were also screened to identify additional relevant studies not captured in the initial database queries.

A multi stage screening procedure was used to select and analyze the final set of studies. First, titles and abstracts were reviewed to exclude papers that did not concern AI applications in financial risk management or did not address any governance-related dimension, such as model risk oversight, explainability, data governance, compliance, or ethical control. Second, full text screening was conducted to retain empirical, conceptual, or review papers that explicitly discussed mechanisms, frameworks, or institutional arrangements for governing AI models in financial risk management contexts, including banking, capital markets, and financial regulation. The selected articles were coded using a structured template that captured publication details, type of AI application, type of risk, governance mechanisms described, stakeholders involved, and key findings. Given the diversity of methods, settings, and governance tools across studies, the evidence was synthesized using a narrative and thematic approach rather than a quantitative meta-analysis, with the aim of identifying common design principles, recurring challenges, and gaps that warrant further research.

4. Results and Discussion

The review shows that research on AI governance in financial risk management is still emerging and unevenly distributed across risk types, institutional settings, and jurisdictions. Most empirical studies focus on banking and credit risk, where AI and machine learning models are used for credit scoring, default prediction, and portfolio monitoring (Ahmed et al., 2022; Tian et al., 2024). Conceptual and review papers tend to concentrate on mapping the broader AI-in-finance landscape or high-level governance principles rather than on detailed institutional practices (Wirtz et al., 2022; Dianti, 2023; Fahrezi, 2024). Only a subset of contributions explicitly integrates governance mechanisms into the analysis of AI-enabled risk management, often framing them in terms of extended model risk management, explainability, and ethical or legal accountability (Floridi et al., 2020; Cosma et al., 2023; Uzougbo et al., 2024). This pattern suggests that governance issues are recognized as critical but are still treated as a secondary layer around technical innovation rather than as a core design dimension of AI-based risk systems.

A first cluster of findings concerns the adaptation of traditional model risk management frameworks to AI models. Studies on model risk in banking and AI model risk in financial institutions argue that existing three-lines-of-defence structures, model inventories, and validation processes provide a useful starting point but are insufficient to capture dynamic data issues, model drift, and feedback loops associated with AI (Cosma et al., 2023; Souza, 2023). Empirical and comparative work shows that banks typically incorporate AI models into existing model risk taxonomies, but that policies for data governance, monitoring, and performance thresholds are still evolving, especially in emerging market contexts (Ahmed et al., 2022; Nnaomah et al., 2024). Across jurisdictions, regulators emphasize the need for robust documentation, independent validation, and board-level oversight of AI risk models, yet supervisory expectations remain heterogeneous, leaving institutions to experiment with their own extensions to model risk management frameworks (Floridi et al., 2020; Wirtz et al., 2020, 2022).

A second cluster of studies highlights explainable AI as a central technical mechanism for AI governance in financial risk management. Work on explainable machine learning in credit risk management demonstrates how tools such as Shapley values and local explanation methods can be integrated into model development and monitoring to make complex models more interpretable for risk managers, auditors, and regulators (Bussmann et al., 2021). Systematic reviews on explainable artificial intelligence in finance and editorial overviews of xAI in financial applications confirm that explainability is increasingly viewed as a governance requirement rather than a purely technical add-on, with particular importance in credit scoring, trading, and portfolio risk models (Černevičienė & Kabašinskas, 2024; Klein & Walther, 2024). Recent empirical applications show that explainable AI can support credit decision-making and financial decision support by revealing key drivers of default risk and enabling more transparent communication with both internal stakeholders and customers (Nallakaruppan et al., 2024). At the same time, these studies caution that explanation techniques can be misused as "window dressing" if not embedded in rigorous validation, data governance, and escalation processes.

The third set of findings relates to organizational and ethical governance mechanisms, including human-in-the-loop structures, fairness controls, and legal accountability. Studies on algorithmic decision-making and AI ethics in finance emphasize that AI-based credit and risk models can entrench bias, undermine trust, and create new channels of systemic risk if human oversight and contestability are weak (Svetlova, 2022; Sargeant, 2023). Research on ethical AI in financial inclusion and legal accountability in financial services shows that organizations are experimenting with fairness metrics, audit trails, and escalation procedures that allow human reviewers to challenge or override AI-generated recommendations, especially in lending and financial inclusion contexts (Uzougbo et al., 2024; Yang & Lee, 2024). Comparative analyses of AI in risk management underline that institutional capacity, supervisory guidance, and risk culture strongly influence the maturity of these mechanisms, with some banking systems moving faster toward formal human-in-the-loop and fairness-by-design requirements than others (Nnaomah et al., 2024). Overall, the evidence points to a gradual convergence toward multi-layered AI governance architectures in which model risk management, explainable AI techniques, data governance, human oversight, and legal accountability are combined, but often in an ad hoc and fragmented way.

Taken together, the results suggest that AI governance mechanisms in modern financial risk management systems are developing along three interconnected dimensions: the extension of model risk management frameworks to AI-specific risks, the embedding of explainable AI as a core technical governance tool, and the institutionalization of human oversight and ethical controls in decision processes. However, the literature also reveals significant gaps. Few studies evaluate the effectiveness of different governance configurations in reducing concrete risk outcomes such as misclassification, discrimination, or systemic spillovers, and cross-risk perspectives beyond credit risk remain limited (Cosma et al., 2023; Tian et al., 2024). The findings therefore support calls for more empirical work on how specific combinations of technical, organizational, and regulatory mechanisms shape the risk profile of AI-enabled financial systems, and for clearer supervisory benchmarks that align innovation in AI-based risk management with financial stability and consumer protection objectives.

5. Conclusion

The review concludes that AI governance in modern financial risk management systems is still at an early but rapidly evolving stage. While AI and machine learning are now widely used for credit scoring, fraud detection, and other risk functions, governance mechanisms often lag behind technical innovation. Existing work shows that governance is frequently treated as an add-on to AI adoption, rather than as a core element designed in parallel with model development. As a result, many institutions extend their existing model risk management frameworks to AI, but struggle to fully address issues such as data drift, feedback loops, systemic effects, and fairness.

At the same time, the literature reveals three main dimensions through which AI governance is developing in financial risk management: the adaptation of model risk management to AI-specific risks, the growing use of explainable AI techniques as governance tools, and the emergence of organizational and ethical mechanisms such as human-in-the-loop oversight, fairness metrics, and legal accountability structures. These strands are complementary, yet they are often implemented in a fragmented and ad hoc way across institutions and jurisdictions. There is little empirical evidence on which combinations of technical, organizational, and regulatory mechanisms are most effective in reducing misclassification, discrimination, or systemic vulnerabilities.

The study is subject to limitations, including its focus on English-language, peer-reviewed articles published between 2019 and 2024 and the use of narrative synthesis rather than meta-analysis. Even so, the findings point to clear implications for practice and policy. Financial institutions and regulators need to move from high-level principles to concrete, testable governance architectures that integrate model risk management, explainability, data governance, human oversight, and accountability into a coherent whole. Future research should provide more comparative and empirical analyses of governance configurations across risk types, markets, and regulatory regimes, in order to support the design of AI-enabled financial risk management systems that are not only innovative, but also robust, fair, and aligned with financial stability and consumer protection objectives.

References

Ahmed, S., Alshater, M. M., El Ammari, A., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. *Research in International Business and Finance, 61*, 101646.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics, 57*(1), 203-216.

Černevičienė, J., & Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: a systematic literature review. *Artificial Intelligence Review, 57*(8), 216.

Cosma, S., Rimo, G., & Torluccio, G. (2023). Knowledge mapping of model risk in banking. *International Review of Financial Analysis, 89*, 102800.

Dianti, A. R. (2023). Enhancing financial risk management in the digital age: a systematic review. *Arthatama: Journal of Business Management and Accounting, 7*(2), 79-91.

Fahrezi, M. (2024). A Systematic Literature Review: The Use of Artificial Intelligence and Machine Learning in Financial Risk Management and Predictive Analytics. *International Journal of Research and Applied Technology (INJURATECH), 4*(2), 60-72.

Floridi, L., Lee, M. S. A., & Denev, A. (2020). Innovating with Confidence: Embedding AI Governance and Fairness in a Financial Services Risk Management Framework. *Berkeley Technology Law Journal, 34*(2), 1–19.

Klein, T., & Walther, T. (2024). Advances in Explainable Artificial Intelligence (xAI) in Finance. *Finance Research Letters, 70*, 106358.

Nallakaruppan, M. K., Chaturvedi, H., Grover, V., Balusamy, B., Jaraut, P., Bahadur, J., Meena, V. P., & Hameed, I. A. (2024). Credit risk assessment and financial decision support using explainable artificial intelligence. *Risks, 12*(10), 164.

Nnaomah, U. I., Odejide, O. A., Aderemi, S., Olutimehin, D. O., Abaku, E. A., & Orieno, O. H. (2024). AI in risk management: An analytical comparison between the US and Nigerian banking sectors. *International Journal of Science and Technology Research Archive, 6*(1), 127-146.

Sargeant, H. (2023). Algorithmic decision-making in financial services: economic and normative outcomes in consumer credit. *AI and Ethics, 3*(4), 1295-1311.

Souza, C. (2023). AI model risk: What the current model risk management framework can teach us about managing the risks of AI models. *Journal of Financial Compliance, 6*(2), 103-112.

Svetlova, E. (2022). AI ethics and systemic risks in finance. *AI and Ethics, 2*(4), 713-725.

Tian, X., Tian, Z., Khatib, S. F. A., & Wang, Y. (2024). Machine learning in internet financial risk management: A systematic literature review. *PLOS ONE, 19*(6), e0300195.

Uzougbo, N. S., Ikegwu, C. G., & Adewusi, A. O. (2024). Legal accountability and ethical considerations of AI in financial services. *GSC Advanced Research and Reviews, 19*(02), 130-142.

Wirtz, B. W., Weyerer, J. C., & Kehl, I. (2022). Governance of artificial intelligence: A risk and guideline-based integrative framework. *Government Information Quarterly, 39*(4), 101685.

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration, 43*(9), 818-829.

Yang, Q., & Lee, Y. C. (2024). Ethical AI in financial inclusion: The role of algorithmic fairness on user satisfaction and recommendation. *Big Data and Cognitive Computing, 8*(9), 105.