INTERNATIONAL JOURNAL OF MANAGEMENT AND ORGANIZATIONAL RESEARCH

Ethical AI in Financial Systems: A Risked- Based Framework for Responsible Innovation

Emmanuel Sampson
Lancaster University, United Kingdom

* Corresponding Author: **Emmanuel Sampson**

Article Info

ISSN (online): 2583-6641
Impact Factor (RSIF): 8.56
Volume: 05
Issue: 01
Received: 27-11-2025
Accepted: 25-12-2025
Published: 22-01-2026
Page No: 104-111

Abstract
The rapid integration of artificial intelligence (AI) into financial systems has transformed core financial functions, including credit allocation, fraud detection, algorithmic trading, and regulatory compliance. While AI-driven financial technologies promise enhanced efficiency, predictive accuracy, and financial inclusion, they also introduce significant ethical, legal, and systemic risks that challenge existing governance structures. These risks ranging from algorithmic discrimination and opacity to accountability gaps, privacy violations, and threats to financial stability are amplified by the scale, interconnectedness, and high-stakes nature of financial decision-making. Current ethical AI frameworks and regulatory responses, although valuable, often rely on principle-based or uniform governance approaches that fail to account for the heterogeneous risk profiles of financial AI applications. Moreover, compliance-oriented regulatory regimes typically establish minimum standards and may lag behind technological developments, limiting their effectiveness in managing emerging ethical risks in real time. This paper advances a finance-specific, risk-based framework for ethical AI governance that aligns the intensity of oversight with the likelihood and severity of potential harm. By embedding ethical considerations within established financial risk management and operational resilience practices, the proposed framework provides a structured and scalable approach to identifying, assessing, and mitigating ethical risks across the AI lifecycle. The framework emphasizes proportionality, accountability, transparency, and continuous monitoring, addressing both individual-level harms and system-wide stability concerns. Ultimately, the study argues that ethical AI governance in finance must move beyond compliance toward responsible innovation that sustains trust, resilience, and legitimacy in increasingly AI-driven financial systems.

DOI: https://doi.org/10.54660/IJMOR.2026.5.1.104-111

Keywords: Ethical artificial intelligence; Financial systems; Risk-based governance; Responsible innovation; Algorithmic accountability; Financial stability; AI regulation

Introduction

The increasing adoption of artificial intelligence (AI) within financial systems has brought about profound changes in the structure and functioning of modern finance. Across banking, insurance, asset management, credit evaluation, fraud detection, and regulatory compliance, AI-driven tools are reshaping how financial decisions are made. [1, 2] Advances in machine learning, large-scale data analytics, and computational power have enabled AI-based financial technologies to achieve meaningful improvements in operational efficiency, predictive performance, cost reduction, and financial inclusion. [3] Simultaneously, the increasing dependence on AI within financial systems has generated a range of ethical, legal, and systemic challenges that strain existing governance arrangements and raise concerns about institutional trust. [4] Financial systems hold a uniquely sensitive role in society, as they regulate access to capital, assess creditworthiness, shape economic opportunities, and influence patterns of wealth distribution. As a result, ethical failures in AI-driven financial applications can generate wide-ranging harms, including discriminatory lending practices, opaque automated decision-making, market manipulation, violations of data privacy, and risks to financial stability. [5, 6] Empirical evidence from credit scoring and insurance underwriting indicates that algorithmic systems can reproduce or intensify historical biases present in training data, frequently to the disadvantage of already marginalized groups. [7] Furthermore, the opacity of complex AI models particularly deep learning architectures has heightened concerns regarding explainability, accountability, and auditability in high-stakes financial decision-making. [8] In addition, the limited transparency of complex AI models, particularly those based on deep learning, has intensified concerns about explainability, accountability, and the ability to audit decisions in high-stakes financial contexts [8].

In response to these challenges, ethical AI has emerged as a central theme in both academic scholarship and policy development. International organizations, regulators, and professional bodies have articulated high-level ethical principles emphasizing fairness, transparency, accountability, robustness, and respect for fundamental rights.[3,4] While these principles provide valuable normative guidance, their translation into operational practice within financial institutions remains uneven. In practice, organizations frequently struggle to reconcile abstract ethical commitments with commercial pressures, technical constraints, and evolving regulatory demands that characterize real-world financial innovation. [9]

An expanding literature suggests that one-size-fits-all or principle-only approaches to ethical AI governance are inadequate for the financial sector, where AI applications carry markedly different risk profiles. For example, systems used in low-risk functions such as customer support do not pose the same ethical or systemic challenges as AI deployed in credit decisions, anti-money laundering surveillance, or algorithmic trading. [2] As a result, there is growing support for risk-based governance frameworks that calibrate ethical oversight And control mechanisms to the likelihood and magnitude of harm. [10]

This study responds to these limitations by advancing a risk-based framework for ethical AI in financial systems that aims to balance innovation with responsibility. By integrating ethical considerations into the financial sector's established risk governance practices, the proposed framework offers a structured and scalable approach to identifying, evaluating, and mitigating ethical risks throughout the AI lifecycle. It emphasizes proportionality, contextual awareness, and continuous oversight, acknowledging that ethical risks evolve alongside technological, economic, and societal developments. [4]

Through its finance-specific and risk-oriented approach, this research advances contemporary discussions on ethical AI governance and demonstrates practical pathways for implementing responsible innovation in complex financial contexts. In this light, embedding ethical considerations into AI governance extends beyond regulatory compliance and represents a strategic necessity for preserving trust, resilience, and legitimacy as financial systems become more reliant on AI.

2. Artificial Intelligence in Financial Systems

2.1. Scope and Adoption of AI in Financial Services

Artificial intelligence (AI) is now firmly embedded in the operational and decision-making foundations of modern financial systems. Financial institutions increasingly rely on machine learning techniques and data-driven models to automate, enhance, or substitute forms of human judgment that were previously dependent on expert evaluation. Within retail and commercial banking, AI-based systems are extensively applied to credit assessment, loan approval processes, customer segmentation, and default risk forecasting. By integrating vast amounts of both structured and unstructured data, these technologies enable more granular risk evaluation, more precise pricing strategies, and the efficient scaling of lending decisions across large and diverse customer bases. [9, 2]

The insurance industry has likewise adopted AI technologies to improve underwriting decisions, premium determination, claims handling, and fraud detection. Machine learning–based predictive models are employed to estimate loss likelihoods, evaluate behavioral risk, and automate claims triage, leading to faster processing and reduced administrative burden. In capital markets, asset managers and trading firms make extensive use of algorithmic trading platforms and portfolio optimization models to detect market patterns, execute trades at high velocity, and actively manage portfolio risk. These systems frequently operate with limited direct human oversight, which introduces distinct challenges related to governance, accountability, and market stability. [11, 12]

2.2. AI in Financial Crime Prevention and Regulatory Compliance

Beyond their use in revenue-generating functions, AI technologies have become integral to financial crime prevention and regulatory compliance. Financial institutions increasingly rely on AI-driven systems for fraud detection, transaction monitoring, and anti-money laundering (AML) surveillance. By processing vast volumes of transactional data in real time, these systems identify anomalous patterns, flag potentially suspicious activities, and support the prioritization of cases for further investigation. Relative to traditional rule-based approaches, AI-enabled methods demonstrate greater flexibility and improved detection performance in complex and rapidly evolving risk environments. [13]

At the same time, the emergence of regulatory technology (RegTech) has further expanded the role of AI in compliance-related activities. RegTech solutions employ AI to automate regulatory reporting, conduct risk assessments, and facilitate interactions with supervisory authorities. These technologies enable continuous monitoring of regulatory requirements and reflect a broader transition toward data-driven and technology-enabled regulatory oversight. However, the growing dependence on AI for compliance and enforcement functions also heightens the importance of governance arrangements that promote transparency, accountability, and auditability particularly where automated outcomes carry legal or regulatory consequences. [14, 1]

2.3. Operational and Strategic Benefits of AI Adoption

The rapid uptake of AI across financial systems is largely motivated by its substantial operational and strategic advantages. By minimizing dependence on manual procedures and individual discretion, AI enables faster, more consistent, and more scalable decision-making. These systems can analyze data volumes far beyond human cognitive limits, uncover complex nonlinear patterns, and adapt quickly to shifts in customer behavior or market conditions. As a result, financial institutions can achieve cost efficiencies, strengthen risk management practices, and enhance their competitive position within increasingly data-driven markets. [3]

At the organizational level, AI facilitates more efficient resource allocation, improves the identification and mitigation of fraud and financial risk, and supports more personalized customer engagement. At the macroeconomic level, AI-enabled efficiencies have the potential to promote innovation and resilience within the financial sector as a whole. Nevertheless, the realization of these benefits is uneven and highly contingent on the availability of high-quality data, robust model design, and effective governance arrangements. Weaknesses in any of these areas can undermine performance and amplify ethical, operational, or systemic risks. [1]

2.4. AI and Financial Inclusion: Opportunities and Limitations

AI-enabled financial technologies are often presented as tools for promoting greater financial inclusion. By drawing on alternative data sources such as payment transactions, mobile phone usage patterns, or other forms of digital trace data AI-based credit assessment models may identify dimensions of creditworthiness that are not captured by conventional credit bureau records. This capability has been cited as a means of expanding access to credit and other financial services for underbanked or unbanked individuals who lack formal financial histories. [3, 14]

2.5. Scale, Interconnectedness, and Systemic Risk in Financial AI

A central characteristic of AI adoption in financial systems is the scale at which these technologies operate and their deep interconnectedness. Financial AI applications are often deployed across extensive customer populations and are embedded within networks of other automated systems both within and across institutions. As a result, decisions generated by these systems can simultaneously affect large numbers of individuals, while errors, biases, or model failures may spread rapidly throughout the financial ecosystem. In capital markets, interactions among algorithmic trading systems have been shown to create feedback mechanisms that intensify volatility, trigger market disruptions, and potentially compromise market integrity.

These characteristics set financial AI apart from many AI applications in other sectors. Ethical concerns such as discriminatory outcomes or limited transparency are not confined to isolated cases but may escalate into system-wide issues with broader implications for financial stability and public trust. The Financial Stability Board has highlighted that the widespread use of AI in finance can introduce new forms of concentration risk, increase operational fragility, and generate correlated failure patterns, all of which call for heightened supervisory and governance attention. [1]

2.6. Implications for Governance of Financial AI

The convergence of high-stakes decision-making, stringent regulatory requirements, and deep systemic interconnectedness in financial systems calls for governance approaches that extend beyond generic ethical AI principles. Frameworks for governing AI in finance must deliberately embed ethical considerations within existing structures for financial risk management, operational resilience, and regulatory compliance. This, in turn, demands oversight mechanisms that are proportionate to the risk profile of individual AI applications, alongside continuous monitoring, clear accountability, and adaptive governance processes capable of responding to evolving risks. [4, 2]

Acknowledging these sector-specific characteristics forms the basis of the risk-based ethical AI framework advanced in this study. By grounding ethical governance within the financial sector's established risk culture, the proposed approach seeks to enable responsible innovation while simultaneously protecting core values such as fairness, transparency, and systemic stability.

3. Ethical, Legal, and Systemic Risks of Financial AI

3.1. Fairness and Discrimination

Fairness-related failures represent some of the most persistent and well-documented risks in the deployment of AI within financial systems. Machine learning models are typically trained on historical financial data that reflect existing social, economic, and institutional inequalities. When such data are used without adequate safeguards, AI systems may reproduce or even exacerbate discriminatory patterns, particularly where protected characteristics such as race, gender, or socioeconomic status are indirectly encoded through proxy variables. [5, 7]

In financial contexts such as credit scoring, loan approval, and insurance underwriting, discriminatory outcomes carry especially severe consequences. Adverse decisions may restrict access to housing, education, healthcare, or entrepreneurial opportunities, thereby reinforcing cycles of exclusion and disadvantage. Unlike discretionary human decision-making, algorithmic discrimination can occur consistently and at scale, affecting large populations with limited visibility or opportunity for redress. Moreover, performance-driven optimization objectives may incentivize accuracy at the expense of equity, unless fairness constraints are explicitly incorporated into model design and evaluation. [5]

These risks highlight the need for governance mechanisms that treat fairness not as an abstract ethical aspiration, but as a concrete risk dimension requiring measurement, monitoring, and mitigation throughout the AI lifecycle.

3.2. Transparency, Explainability, and Contestability

A central governance challenge associated with AI deployment in financial systems concerns the limited transparency of advanced machine learning models, particularly those built on deep learning architectures. These models frequently operate as opaque decision systems, generating outputs that are difficult to interpret, explain, or justify in terms that are accessible to human decision-makers. Such opacity constrains accountability and weakens the capacity of affected individuals, regulators, and financial institutions to meaningfully examine, contest, or validate automated outcomes. [8, 6, 15]

Within financial services, explainability extends beyond a technical design choice and constitutes a core governance requirement. Automated decisions relating to credit approval, insurance claims, fraud detection, or account restrictions often carry significant legal, financial, and social consequences. When decision rationales are unavailable, partial, or unintelligible, individuals may be unable to understand the basis of adverse outcomes or exercise effective rights to review and appeal. This undermines procedural fairness, weakens consumer protection, and erodes trust in both financial institutions and algorithmic decision-making systems. [5, 16]

Explainability also plays a critical role in regulatory oversight and institutional auditability. Supervisory authorities require sufficient visibility into model behavior to evaluate compliance with consumer protection laws, anti-discrimination requirements, and prudential risk standards. Lack of interpretability can impede supervisory assessments, limit effective model validation, and obscure emerging risks. As a result, transparency and contestability obligations should be proportionate to the risk and impact of specific AI applications, with systems that directly affect individuals' financial rights or market stability subject to enhanced explainability, documentation, and review requirements. [10, 1, 8]

3.3. Accountability and Auditability

A significant governance challenge in the deployment of AI within financial systems arises from the diffusion of responsibility across technical, organizational, and commercial actors. Automated decision-making processes typically involve a complex network of stakeholders, including data scientists, model developers, third-party technology vendors, business units, senior management, and external service providers. When AI-driven systems generate harmful, biased, or unlawful outcomes, this fragmentation can obscure responsibility and complicate efforts to attribute accountability for errors or failures. [6, 17]

To address this challenge, effective governance frameworks must establish explicit lines of accountability, reinforced by strong auditability mechanisms. Comprehensive audit trails, rigorous documentation, and systematic version control are critical for enabling post hoc examination of automated decisions, supporting internal risk oversight, and facilitating regulatory supervision. Within the financial sector, established governance traditions particularly model risk management frameworks offer a robust institutional foundation for embedding accountability into AI systems. Supervisory guidance emphasizes the importance of independent model validation, continuous performance monitoring, and senior-level governance oversight as essential elements of responsible model deployment and use. [18, 19]

Extending these well-established practices to AI-driven decision systems helps ensure that automation does not displace human responsibility but remains subject to institutional control and ethical scrutiny. By anchoring AI governance within existing risk management structures, financial institutions can better align technological innovation with accountability, regulatory expectations, and public trust. [10, 4]

3.4. Privacy and Data Protection

Financial AI systems depend extensively on the processing of large volumes of sensitive personal and transactional data, placing privacy and data protection at the center of ethical and legal concern. These systems introduce a range of risks, including unauthorized access to personal information, repurposing of data beyond its original intent, security breaches, and the production of intrusive or unwarranted inferences about individuals' behavior, preferences, or socioeconomic characteristics. Such risks are further intensified when alternative data sources such as behavioral data, digital footprints, or online activity traces are incorporated into automated decision-making processes, often without clear visibility or understanding on the part of affected individuals. [5, 20]

Within the European Union, the deployment of AI systems that produce decisions with significant effects on individuals is closely regulated under data protection law, particularly through provisions governing solely automated decision-making and profiling in Article 22 of the General Data Protection Regulation (GDPR). These provisions emphasize the need for safeguards such as transparency, meaningful human intervention, and the right of individuals to obtain explanations and challenge automated outcomes.[22] Comparable expectations are reflected in UK GDPR guidance and enforcement practice, which stress accountability, fairness, and proportionality in the use of automated decision systems within regulated sectors such as finance. [23]

As a result, compliance with data protection regimes cannot be treated as a peripheral legal obligation but must be regarded as an integral component of ethical AI governance in financial systems. Privacy risks intersect with broader ethical concerns relating to fairness, transparency, and autonomy, and therefore require coordinated assessment rather than isolated treatment. Effective safeguards must be embedded across the entire AI lifecycle from data collection and feature selection to model development, deployment, and ongoing monitoring to ensure that financial innovation does not compromise fundamental rights or public trust. [3, 4]

3.5. Systemic and Financial Stability Risks

Beyond harms experienced at the individual level, several AI applications in finance pose material risks to financial stability and the functioning of markets as a whole. Algorithmic and high-frequency trading systems, in particular, can interact in complex and non-linear ways, intensifying market volatility, accelerating price movements, and exacerbating stress during periods of market turbulence. Empirical research has shown that automated trading strategies may reinforce feedback loops and herd behavior, thereby increasing the probability of abrupt market disruptions and destabilizing events. [1, 12]

Systemic vulnerabilities may also arise from the widespread adoption of similar AI models, data sources, or technological infrastructures across financial institutions. Reliance on shared vendors, standardized modeling approaches, or common cloud service providers can generate correlated behavior and concentration risk, creating potential single points of failure within the financial system. In such environments, localized model errors, cyber incidents, or infrastructure outages can rapidly propagate across institutions and markets, amplifying shocks and undermining system-wide resilience. [1, 23]

In response to these emerging risks, supervisory authorities have increasingly emphasized the concept of operational resilience, defined as the capacity of financial institutions to continue delivering critical services in the face of disruptions. Regulatory guidance highlights the importance of identifying critical business services, managing dependencies on third-party providers, and preparing for severe but plausible disruption scenarios.[24] These systemic considerations underscore how AI deployment in finance differs from AI use in many other sectors: ethical failures, operational weaknesses, or governance gaps can escalate quickly into broader threats to market integrity and public confidence. Consequently, financial AI governance must adopt a risk-sensitive and system-aware approach that accounts not only for individual harms but also for collective and stability-related effects.

4. Limitations of Existing Ethical AI Approaches in Financial Systems

4.1. Principle-Based Ethical AI Frameworks: Strengths and Shortcomings

Beyond impacts at the individual level, the deployment of AI in finance introduces risks that can threaten financial stability and market integrity. Certain applications most notably algorithmic and high-frequency trading systems may interact in complex, nonlinear ways that intensify market volatility, accelerate price dynamics, and exacerbate stress during periods of market disruption. Empirical evidence indicates that automated trading strategies can reinforce feedback loops and collective behavior, thereby increasing the likelihood of abrupt market dislocations and destabilizing events. [1, 12, 26]

Systemic vulnerabilities are further heightened when financial institutions converge on similar AI models, datasets, or technological infrastructures. Widespread reliance on shared vendors, standardized modeling approaches, or common cloud service providers can generate correlated behavior and concentration risk, creating potential single points of failure within the financial system. In such settings, localized model errors, cyber incidents, or infrastructure outages may propagate rapidly across institutions, amplifying shocks and undermining system-wide resilience. [1, 27]

In recognition of these emerging threats, supervisory authorities have increasingly emphasized the importance of operational resilience, understood as the capacity of financial institutions to continue delivering critical services in the face of severe but plausible disruptions. Regulatory guidance highlights the need to identify critical business services, manage dependencies on third-party providers, and conduct stress testing that accounts for technological and operational failures alongside traditional financial risks. [23, 24]

These systemic considerations distinguish financial AI from AI deployment in many other sectors. Ethical lapses, governance deficiencies, or operational breakdowns in financial AI systems can escalate rapidly into broader threats to market integrity and public confidence. This reinforces the necessity for risk-sensitive and system-aware governance frameworks that address not only individual-level harms but also collective dynamics and stability-related effects inherent in AI-driven financial systems.

4.2. Inadequacy of Uniform Governance Approaches

A key weakness of many existing ethical AI frameworks lies in their tendency to impose uniform governance requirements across a wide range of AI applications, regardless of context or risk level. In financial systems, however, AI use cases display markedly heterogeneous risk profiles. Low-impact applications such as customer service chatbots or internal process automation pose fundamentally different ethical, legal, and systemic risks from high-stakes systems used in credit approval, anti-money laundering (AML) surveillance, insurance underwriting, or algorithmic trading. Treating these applications as ethically equivalent obscures meaningful differences in the likelihood and severity of harm, thereby undermining effective prioritization of governance resources and controls. [1, 2]

The application of uniform governance standards can lead to two problematic outcomes. First, imposing stringent oversight and compliance burdens on low-risk systems may inhibit innovation, raise implementation costs, and discourage experimentation with technologies that could deliver efficiency gains or consumer benefits. Second, applying insufficient or generic oversight to high-risk systems may allow serious ethical failures such as discrimination, opacity, or systemic instability—to occur without adequate safeguards. In both scenarios, the absence of proportionality weakens the practical effectiveness and institutional credibility of ethical AI governance within financial systems. [4, 10]

Regulatory and policy developments increasingly recognize this limitation. Risk-based approaches, such as those reflected in the European Union's Artificial Intelligence Act and financial supervisory guidance, explicitly acknowledge that governance obligations should scale with the potential impact of AI systems. [10, 22] These developments underscore the need for ethical AI frameworks in finance that move beyond one-size-fits-all models and instead adopt differentiated, risk-sensitive governance aligned with the sector's established risk management traditions.

4.3. Limited Integration with Financial Risk Management Traditions

Another critical shortcoming of existing ethical AI framework is their weak integration with established financial risk management practices. Financial institutions operate within mature governance structures designed to manage credit risk, market risk, operational risk, and model risk. These frameworks emphasize clear accountability, documentation, validation, monitoring, and escalation procedures. However, ethical AI guidance is often treated as an external or parallel compliance layer, rather than being embedded within these existing governance mechanisms.

For example, ethical risks such as bias, opacity, or unfair treatment are rarely incorporated explicitly into model risk management processes, despite their potential material impact on financial and reputational outcomes. This separation limits the practical influence of ethical considerations on decision-making and reduces the likelihood that ethical risks will be identified, monitored, and mitigated alongside other forms of risk. [6, 11]

Embedding ethical AI governance within familiar financial risk structures would enhance institutional uptake and enable more systematic oversight, yet many existing frameworks do not provide guidance on how such integration should occur.

4.4. Compliance-Oriented Regulatory Approaches and Their Limits

Regulatory responses to the governance of artificial intelligence have increasingly adopted risk-based approaches, most prominently reflected in the European Union's Artificial Intelligence Act, which categorizes AI systems according to risk levels and assigns differentiated regulatory obligations accordingly. [10] This approach represents a significant advance toward proportional oversight, explicitly recognizing that AI applications vary widely in their potential to cause harm and therefore should not be subject to uniform regulatory treatment.

Despite these advances, regulatory measures alone are insufficient to address the ethical governance challenges posed by AI in financial systems. Regulatory frameworks typically define minimum compliance thresholds rather than encouraging ethical leadership or proactive risk anticipation. As a result, financial institutions may prioritize formal adherence to legal requirements without substantively engaging with broader ethical considerations, such as fairness, accountability, or societal impact. [2, 4] Moreover, regulatory instruments are, by necessity, general in scope and often struggle to keep pace with rapid technological innovation, limiting their effectiveness in addressing emergent or context-specific risks. Jurisdictional variation further compounds these challenges, creating regulatory fragmentation and compliance complexity for financial institutions operating across multiple legal regimes. [1]

Accordingly, effective ethical AI governance in finance requires regulatory oversight to be complemented by robust internal governance frameworks. Such frameworks must translate legal and regulatory expectations into concrete operational practices that are integrated with institutional risk management processes, organizational culture, and strategic objectives. Embedding ethical considerations within internal governance structures enables financial institutions to move beyond compliance-driven approaches and to manage ethical risks dynamically, in alignment with both regulatory requirements and long-term institutional responsibility. [3, 24]

4.5. The Case for a Risk-Based, Finance-Specific Ethical AI Framework

Taken together, these shortcomings underscore the need for a more context-sensitive approach to ethical AI governance in financial systems. A risk-based framework offers distinct advantages in this regard. By aligning the intensity of oversight with the likelihood and magnitude of potential harm, such an approach enables proportional governance that balances the promotion of innovation with the protection of individuals, institutions, and markets. Moreover, situating ethical considerations within established financial risk management practices enhances practical feasibility, strengthens accountability, and increases institutional acceptance by leveraging familiar governance structures.[1,2]

Importantly, a risk-based perspective also accommodates the dual nature of ethical risk in finance, encompassing both individual-level harms such as unfair treatment or exclusion and system-wide concerns related to market integrity and financial stability. This reflects the distinctive role of financial systems as critical socio-economic infrastructures whose failures can have far-reaching societal consequences. [4, 24]

Viewing ethical risk as a dynamic and context-dependent phenomenon, rather than as a static checklist of principles, provides a more robust foundation for effective governance. Ethical risks evolve alongside changes in technology, data practices, market structures, and regulatory expectations, requiring continuous assessment and adaptive oversight. This understanding motivates the risk-based ethical AI framework advanced in this study, which seeks to translate ethical responsibility into operational governance practices that are responsive to the realities of financial innovation, regulatory complexity, and systemic interdependence. [3]

5. Methodological Approach

5.1. Research Design and Analytical Orientation

This study adopts a conceptual and design-science research approach to develop a practical framework for ethical AI governance in financial systems. Design-science methods are particularly appropriate where the objective is not hypothesis testing but the creation of an actionable artefact such as a framework, model, or governance architecture intended to address a real-world problem. [25] In the context of ethical AI, conceptual frameworks play a critical role in translating normative principles into implementable institutional practices. [4]

The analytical orientation of the study is interdisciplinary, integrating insights from ethical AI scholarship, financial risk management, regulatory governance, and financial stability analysis. Rather than treating ethics as an external constraint, the study conceptualizes ethical risk as a form of organizational and systemic risk that can be identified, assessed, and managed using tools familiar to financial institutions. [6, 27]

5.2. Justification for a Risk-Based Approach

A risk-based governance approach is particularly appropriate for financial systems for several interrelated reasons. First, financial institutions operate within well-established risk management cultures that prioritize proportionality, materiality assessment, and structured escalation mechanisms. These institutional traditions provide a natural foundation for integrating ethical oversight into existing governance processes rather than introducing parallel or externally imposed frameworks. Second, ethical risks associated with AI deployment in finance differ substantially in their severity, scale, and reversibility. While some harms may be limited and remediable, others such as discriminatory exclusion from credit or systemic market disruption can have enduring and far-reaching consequences, thereby warranting differentiated levels of oversight and control. [2, 4]

Third, supervisory and regulatory authorities increasingly expect financial institutions to demonstrate risk-sensitive and outcome-oriented governance practices, moving beyond formalistic or uniform compliance approaches. Emerging regulatory instruments and supervisory guidance emphasize the need for institutions to identify, assess, and manage AI-related risks in a manner commensurate with their potential impact, reflecting broader shifts toward proportional and principles-based regulation. [23, 24]

By conceptualizing ethical concerns as a category of risk management challenge, the proposed methodological approach facilitates institutional adoption and embeds ethical AI governance within established decision-making, control, and accountability structures. This integration supports more effective oversight while aligning ethical responsibility with the operational realities of financial innovation and regulatory compliance. [1, 3]

6. Illustrative Applications of the Risk-Based Framework

The practical utility of the proposed risk-based ethical AI framework is illustrated through its application to several high-impact use cases in financial systems. In the context of credit scoring and loan approval, AI systems are generally classified as high-risk because they directly shape individuals' access to credit and broader economic participation. Applying the framework in this domain necessitates robust pre-deployment ethical risk assessment, including systematic bias evaluation, verification of data quality and representativeness, the implementation of explainability mechanisms to enable contestability, and structured human oversight for adverse or borderline decisions. These safeguards are consistent with prevailing ethical expectations and align closely with established supervisory approaches to model risk management in the banking sector. [2, 5, 24]

Anti-money laundering (AML) and fraud detection systems similarly entail heightened ethical and legal risks, reflecting their dependence on large-scale monitoring, the processing of sensitive personal data, and automated identification of suspicious activity. Although such systems are essential for preserving financial integrity, high false-positive rates may result in account restrictions, reputational damage, or exclusion from financial services. Under the proposed framework, these applications warrant enhanced governance controls, including continuous performance evaluation, levels of explainability adequate for internal governance and regulatory review, clearly defined accountability for escalation and enforcement decisions, and safeguards designed to address proportionality and data protection concerns. [6, 10, 11]

Algorithmic trading and market surveillance systems occupy the highest tier of ethical and systemic risk, given their capacity to influence market dynamics, liquidity, and stability. In this setting, the framework prioritizes systemic risk assessment, stress testing, and governance arrangements that explicitly account for model interactions, feedback mechanisms, and concentration effects arising from widespread adoption of similar strategies. Strong oversight, independent validation, and close integration with operational resilience planning are critical to mitigating the risk of market disruption or correlated institutional failures. [1, 12, 24]

Taken together, these examples demonstrate how ethical AI governance can be operationalized in a manner that is proportionate, context-aware, and compatible with existing financial risk management traditions. By calibrating oversight to the severity and scale of potential harm, the framework supports responsible innovation while protecting individual rights, reinforcing institutional accountability, and safeguarding systemic stability.

7. Conclusion and Future Research Directions

7.1 Conclusion

The integration of artificial intelligence into financial systems has reshaped critical functions including credit allocation, fraud detection, market trading, and regulatory compliance. While these technologies deliver substantial gains in efficiency, scalability, and analytical capacity, they also generate significant ethical, legal, and systemic risks that strain existing governance arrangements. Current ethical AI frameworks and regulatory responses, although valuable, frequently rely on high-level principles or uniform compliance approaches that insufficiently account for the heterogeneous risk profiles and systemic importance of financial AI applications. [1, 4]

This study responds to these limitations by advancing a finance-specific, risk-based framework for ethical AI governance. By framing ethical concerns as a category of organizational and systemic risk, the framework integrates ethical oversight into established financial risk management practices, emphasizing proportionality, accountability, and continuous monitoring. Through illustrative applications across high-impact use cases, the framework demonstrates how governance controls can be differentiated and scaled to address both individual-level harms and system-wide stability risks. [2, 24]

Overall, the findings highlight that ethical AI governance in finance should be understood not as a peripheral compliance obligation but as a strategic enabler of responsible innovation. Embedding ethical considerations within existing risk cultures enhances institutional trust, facilitates regulatory alignment, and strengthens the resilience and legitimacy of increasingly AI-driven financial systems. [3, 10]

7.2. Future Research Directions

Future research should empirically examine the implementation and effectiveness of risk-based ethical AI governance within financial institutions, including cross-jurisdictional and sectoral comparisons. Further work is also needed to develop standardized metrics for assessing ethical risk, particularly in relation to fairness, explainability, and systemic impact.

Additionally, as regulatory regimes such as the EU Artificial Intelligence Act evolve, research should explore how internal governance frameworks interact with formal legal requirements and supervisory practices. Finally, emerging technologies including generative AI, foundation models, and increased reliance on third-party AI providers raise new governance challenges related to accountability, concentration risk, and operational resilience that warrant focused investigation.

Together, these avenues will be critical to ensuring that AI-driven financial innovation remains aligned with ethical principles, regulatory expectations, and the stability of the financial system.

8. References

1. Financial Stability Board (FSB). Artificial intelligence and machine learning in financial services: market developments and financial stability implications. Basel: FSB; 2017. p. 1.
2. Zetzsche DA, Buckley RP, Arner DW, Barberis JN. Regulating a revolution: from FinTech to smart regulation. Fordham J Corp Financ Law. 2018;23:1–46.
3. Organisation for Economic Co-operation and Development (OECD). OECD principles on artificial intelligence. Paris: OECD Publishing; 2019.
4. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, *et al*. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach. 2018 Dec;28(4):689–707.
5. Barocas S, Hardt M, Narayanan A. Fairness and machine learning. In: Ricci F, Rokach L, Shapira B, editors. Recommender systems handbook. 2nd ed. New York: Springer; 2020. p. 453–479.
6. Kroll JA. Accountable algorithms [dissertation]. Princeton (NJ): Princeton University; 2015.
7. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv. 2021 Jul 13;54(6):1–35.
8. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv. 2017. arXiv:1702.08608.
9. Ryll L, Seetharaman A. Ethical artificial intelligence in finance: governance, risk, and accountability. J Financ Regul Compliance. 2023;31(2):165–182.
10. Ulnicane I. Artificial intelligence in the European Union: policy, ethics and regulation. In: Cini M, Borragán NP, editors. The Routledge handbook of European

integrations. London: Routledge; 2022.
11. Opara IJ, Fatola ME, Adeoba MI, Sodje HA, Chibueze NT, Olorunkosebi MT, et al. AI-driven predictive microbiology with real-time sensors for next-generation food safety. J Life Sci Public Health. 2025;1(2):79-89. doi:10.69739/jlsph.v1i2.1299
12. Financial Stability Board (FSB). Final report: recommendations of the Task Force on Climate-Related Financial Disclosures. Basel: FSB; 2017.
13. Kirilenko AP, Stepchenkova SO, Kim H, Li X. Automated sentiment analysis in tourism: comparison of approaches. J Travel Res. 2018 Nov;57(8):1012–1025.
14. Pavlidis G. Financial Action Task Force and the fight against money laundering and the financing of terrorism: quo vadimus? J Financ Crime. 2021 Aug;28(3):765–773.
15. Arner DW, Barberis JN, Buckley RP. FinTech and RegTech in a nutshell, and the future in a sandbox. Charlottesville (VA): CFA Institute Research Foundation; 2017 Jul 31.
16. World Bank Group. Global economic prospects: January 2022. Washington (DC): World Bank Publications; 2022 Mar 4.
17. Burrell J. How the machine "thinks": understanding opacity in machine learning algorithms. Big Data Soc. 2016;3(1):2053951715622512.
18. Selbst AD, Barocas S. The intuitive appeal of explainable machines. Fordham Law Rev. 2018;87:1085–1139.
19. Rahwan I. Society-in-the-loop: programming the algorithmic social contract. Ethics Inf Technol. 2018 Mar;20(1):5–14.
20. Polak P, Nelischer C, Guo H, Robertson DC. "Intelligent" finance and treasury management: what we can expect. AI Soc. 2020 Sep;35(3):715–726.
21. Kambhu J, Peters B, Ruocco JJ, Sabado W, Greenlee JD, Fleming J, *et al*. Office of the Comptroller of the Currency notes on senior supervisors meeting with Citigroup. Washington (DC): OCC.
22. Mittelstadt B, Russell C, Wachter S. Explaining explanations in AI. In: Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT*); 2019 Jan 29. p. 279–288.
23. Gáti B. Some data protection issues of the EU regulation of artificial intelligence. In: Protivrječja savremenog prava. Vol II. Istočno Sarajevo: Faculty of Law, University of East Sarajevo; 2022. p. 588–605.
24. Butterworth M. The ICO and artificial intelligence: the role of fairness in the GDPR framework. Comput Law Secur Rev. 2018 Apr;34(2):257–268.
25. Abuhamra A, Chin PN, Ganesan Y. Examining the impact of board of director committees on the financial stability and performance of banks. Glob Bus Manag Res. 2024 Jul;16(3):56–71.
26. Badev A, Gutiérrez LB, Da S, Lopes R, Duellmann K, Endo Y, *et al*. Basel Committee on Banking Supervision. Basel: Bank for International Settlements.
27. Easley D, López de Prado MM, O'Hara M. Flow toxicity and liquidity in a high-frequency world. Rev Financ Stud. 2012 May;25(5):1457–1493.
28. Danielsson A. Knowledge in and of military operations: enriching the reflexive gaze in critical research on the military. Crit Mil Stud. 2022 Jul;8(3):315–333.
29. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q. 2004 Mar;28(1):75–105.
30. Acharya VV, Cooley TF, Richardson M, Walter I. The power of central banks and the future of the Federal Reserve System. In: Acharya VV, editor. Regulating Wall Street. Hoboken (NJ): Wiley; 2011. p. 51–68.
31. Financial Stability Board (FSB). The Basel Committee on Banking Supervision: Macroeconomic Assessment Group. Basel: FSB; 2010.

How to Cite This Article

Sampson E. Ethical AI in financial systems: a risked-based framework for responsible innovation. Int J Manag Organ Res. 2026;5(1):104-111. doi:10.54660/IJMOR.2026.5.1.104-111.

Creative Commons (CC) License