STAR MOUNTAIN
INTERNATIONAL PUBLISHING GROUP PET.LTD

Article

Generative Artificial Intelligence Governance in Financial Services: Managing Hallucination, Bias, and Compliance Risks

Anna Kowalska[1, *], James O. Morgan[2], Hiroshi Tanaka[3], Sofía L. Ramírez[4]

[1]Faculty of Management, University of Warsaw, Warsaw, 00-927, Poland; anna.kowalska@uw.edu.pl

[2]Said Business School, University of Oxford, Oxford, OX1 1HP, United Kingdom; james.morgan@sbs.ox.ac.uk

[3]Graduate School of Economics, Hitotsubashi University, Tokyo, 186-8601, Japan; hiroshi.tanaka@econ.hit-u.ac.jp

[4]Escuela de Negocios, Universidad del Pacífico, Lima, 15072, Peru; slramirez@up.edu.pe

***Correspondence:** Anna Kowalska. Email: anna.kowalska@uw.edu.pl

Abstract

Generative artificial intelligence has transitioned from experimental technology to embedded operational infrastructure across financial services, deployed for marketing, customer communications, anti-money laundering monitoring, and compliance functions. Yet the same capabilities that drive efficiency introduce novel risks that existing risk management frameworks are ill-equipped to address. This article examines the governance challenges posed by generative AI in regulated financial institutions, focusing on three risk categories identified as priorities by financial regulators: hallucinations and inaccurate outputs, algorithmic bias and concept drift, and the emerging autonomy of AI agents. The analysis draws on the Financial Industry Regulatory Authority's 2026 Regulatory Oversight Report, the European Banking Authority's implementation of the EU AI Act, the Monetary Authority of Singapore's proposed AI risk management guidelines, and the FINOS AI Governance Framework. Findings indicate that existing securities laws and supervisory rules apply with equal force to generative AI-powered operations, yet compliance gaps persist due to the novelty of the technology and the absence of standardised testing protocols. Persistent challenges include the difficulty of verifying output accuracy, the opacity of model decision-making, and the accountability vacuum created by autonomous AI agents. The analysis concludes by outlining a multi-layered governance framework encompassing cross-functional oversight, usage policies, testing and monitoring protocols, and recordkeeping practices tailored to generative AI.

Keywords: generative artificial intelligence; financial services; AI governance; regulatory compliance; hallucinations; algorithmic bias; AI agents

ARTICLE INFORMATION

Received: 2 April 2026; Accepted: 22 June 2026; Published: 24 June 2026

CITATION

Kowalska A, Morgan JO, Tanaka H, Ramírez SL. Generative Artificial Intelligence Governance in Financial Services: Managing Hallucination, Bias, and Compliance Risks.

Advances in Digital Finance. 2026; 1(1): 5.

1. Introduction

The integration of generative artificial intelligence into financial services represents one of the most consequential technological shifts in the sector since the advent of algorithmic trading. Within the span of three years, large language models have moved from research laboratories to production environments, where they now perform functions that were previously the exclusive domain of skilled professionals: drafting investor communications, synthesising regulatory filings, monitoring transactions for suspicious activity, and generating research summaries (FINRA, 2025; Bain & Company, 2025). Industry surveys consistently identify summarisation and information extraction as the predominant use cases, with firms deploying GenAI to condense vast quantities of unstructured text and extract actionable intelligence from documents that would otherwise require hours of human review (FINRA, 2025).

The efficiency gains are undeniable, yet they are accompanied by risks that are qualitatively different from those associated with earlier generations of financial technology. Generative AI systems produce novel outputs that are plausible and fluent yet may be factually incorrect or subtly biased. Unlike deterministic algorithms, these systems exhibit emergent behaviours that are difficult to predict, control, or fully understand—even for their developers. This opacity, combined with the scale and speed of deployment, creates governance challenges that existing risk management frameworks, designed for an earlier technological era, are ill-equipped to address (Bain & Company, 2025).

The regulatory community has responded with growing urgency. In December 2025, the Financial Industry Regulatory Authority published its 2026 Annual Regulatory Oversight Report, which included a dedicated section on generative AI for the first time in the organisation's history (FINRA, 2025). The report articulated a principle that compliance teams cannot afford to overlook: the regulatory frameworks governing traditional business activities apply with equal force to

GenAI-powered operations (Saifr, 2026). The European Banking Authority has established a dedicated workstream to map the EU AI Act against relevant provisions in banking and payments sectoral frameworks (European Banking Authority, 2025). The Monetary Authority of Singapore has proposed guidelines on AI risk management covering generative AI and emerging AI agents (MAS, 2025). The FINOS AI Governance Framework has provided a production-ready playbook that financial institutions can integrate into existing three-lines-of-defence models (FINOS, 2025).

The governance challenges posed by generative AI in regulated financial institutions are examined here, with a focus on three risk categories identified as priorities by financial regulators: hallucinations and inaccurate outputs, algorithmic bias and concept drift, and the emerging autonomy of AI agents. Regulatory reports, supervisory guidance, and industry frameworks are drawn upon to assess the regulatory landscape and identify governance practices for responsible deployment.

2. Materials and Methods

2.1 Analytical Framework

A qualitative, comparative approach is adopted to examine generative AI governance in financial services. The framework is organised around three analytical dimensions: the risk categories that financial regulators have identified as priorities for GenAI oversight; the regulatory and supervisory expectations articulated in official guidance; and the governance practices recommended by industry frameworks.

2.2 Regulatory Sources

Four primary regulatory and industry sources are drawn upon: the Financial Industry Regulatory Authority's 2026 Annual Regulatory Oversight Report (FINRA, 2025); the European Banking Authority's mapping of the EU AI Act against banking and payments sectoral frameworks (European Banking Authority, 2025); the Monetary Authority of Singapore's proposed Guidelines on AI Risk Management (MAS, 2025); and the FINOS AI Governance Framework v1.0 (FINOS, 2025). Additional sources include the Bank of England's approach to AI innovation and the SEC's 2025 examination priorities regarding AI use in investment strategies and operations.

2.3 Risk Categorisation and Governance Practice Classification

Risks are categorised according to the taxonomy developed in FINRA's 2026 report. Three primary risk categories are examined: accuracy and hallucinations, bias and concept drift, and AI

agent autonomy. Governance practices are classified into four functional categories: oversight structures, usage policies, testing and monitoring protocols, and recordkeeping and auditability.

3. Results

3.1 Regulatory Risk Categories

3.1.1 Accuracy and Hallucinations

The most immediate risk identified by regulators is the generation of inaccurate or fabricated information—colloquially referred to as hallucinations. Large language models generate text based on statistical patterns rather than factual understanding. They have no internal mechanism for distinguishing between accurate and inaccurate information in their training data, nor do they possess a reliable method for verifying the truth of their outputs (FINOS, 2025). The result is a system that can produce outputs that are syntactically fluent and rhetorically persuasive yet entirely factually incorrect.

In a financial services context, the consequences can be severe. A chatbot that fabricates performance data can mislead investors and expose firms to enforcement actions. An AI system that misinterprets a regulatory rule can result in unsuitable product recommendations or compliance failures (Saifr, 2026). When hallucinations appear in investor communications, marketing materials, or compliance recommendations, they can mislead customers and result in incorrect interpretations of rules (FINRA, 2025). Techniques such as retrieval-augmented generation can reduce hallucinations by providing factual context, but they cannot fully prevent the model from introducing errors (FINOS, 2025). Hallucinations remain a persistent and unresolved challenge.

3.1.2 Bias and Concept Drift

Algorithmic bias and concept drift introduce more subtle but equally serious challenges. AI models trained on historical data may perpetuate existing biases in marketing targeting, modelling and simulations, and risk assessments (Saifr, 2026). These biases can result in skewed outputs that disproportionately affect certain demographic groups or market segments. FINRA's 2026 report emphasises that firms must consider the integrity, reliability, and accuracy of AI models when relying on GenAI tools as part of their supervisory systems (DLA Piper, 2025).

Concept drift compounds this problem as models trained on older data become less accurate over time, particularly in rapidly changing markets. An anti-money laundering system trained on pre-pandemic transaction patterns may fail to identify emerging fraud schemes or generate excessive

false positives (Saifr, 2026). The combination of historical bias and concept drift creates a dynamic risk environment in which models that were fair and accurate at deployment may become neither over time.

3.1.3 Autonomy of AI Agents

The emergence of autonomous AI agents represents a frontier of risk that existing regulatory frameworks were not designed to address. Advanced AI agents can independently execute tasks, make decisions, and take actions across multiple systems (FINRA, 2025). While this autonomy promises efficiency gains, it also creates accountability gaps that pose significant regulatory challenges (Saifr, 2026). The fundamental question—who is responsible when an AI agent initiates an unauthorised trade, sends non-compliant communications, or accesses restricted data—remains unresolved in many organisations.

FINRA's 2026 report recommends that firms consider a narrow scope of authority for AI agents, explicit permissions, audit trails of actions, and explicit human checkpoints before execution (Shumaker, 2025). The regulatory supervisory model requires registered human decision-makers at critical junctures, and firms must ensure that autonomy does not come at the cost of human accountability (DLA Piper, 2025)..

3.2 Regulatory and Supervisory Expectations

3.2.1 FINRA: Existing Rules Apply Without Exception

FINRA's 2026 report leaves no room for ambiguity: the regulatory frameworks that have long governed traditional business activities apply with equal force to GenAI-powered operations (Saifr, 2026). Using GenAI can implicate rules regarding supervision, communications, recordkeeping, and fair dealing (DLA Piper, 2025). Specifically, Rule 3110 supervisory obligations extend to GenAI outputs and model behaviours, and firms cannot delegate supervisory responsibility to algorithms (Saifr, 2026). Rule 2210, which governs marketing content and customer service responses, applies equally to machine-generated material—the fact that content is AI-produced does not reduce a firm's responsibility for its accuracy and appropriateness (Saifr, 2026). Recordkeeping obligations apply to GenAI systems as well. Firms must retain records of business-related communications and supervisory activities, including logs of AI prompts, outputs, model versions, and human oversight actions (Saifr, 2026).

3.2.2 EU AI Act and the European Banking Authority

The EU AI Act, which entered into force in stages beginning in 2025, classifies AI systems based on risk. The use of AI systems for creditworthiness assessment or credit scoring of natural persons is classified as high-risk under Annex III(5)(b) of the AI Act (European Banking Authority, 2025). In January 2025, the European Banking Authority established a dedicated workstream to map the AI Act against relevant provisions in EU banking and payments sectoral frameworks (European Banking Authority, 2025). The European Parliament, in its November 2025 resolution on the impact of artificial intelligence in the financial sector, recommended clear, proportionate rules and supervisory coordination for AI in financial services (European Parliament, 2025).

3.2.3 Monetary Authority of Singapore: Risk-Based Proportionality

The Monetary Authority of Singapore has proposed guidelines on AI risk management that apply to all financial institutions and set out supervisory expectations on oversight of AI risk management, key AI life cycle controls, and capabilities needed for the use of AI (MAS, 2025). The guidelines cover different AI applications and technologies, including generative AI and newer developments such as AI agents (MAS, 2025). MAS has set out expectations in several key areas: oversight by boards and senior management; establishment of clear identification processes for AI usage; and implementation of robust controls in areas such as data management, fairness, transparency, human oversight, and third-party risks (MAS, 2025). The guidelines are intended to be applied in a proportionate manner, commensurate with the size and nature of financial institutions' activities (MAS, 2025).

3.2.4 Bank of England and Prudential Regulation

The Bank of England has articulated a framework for responsible AI innovation that emphasises the importance of appropriate governance, stress testing, and consideration of risks to safety and soundness and to financial stability (Bank of England, 2025). The Prudential Regulation Authority has held roundtable sessions with regulated firms on the adoption of AI in the context of implementing the principles set out in SS1/23 'Model risk management principles for banks' (Bank of England, 2025). Participants noted key differences between the UK's regulatory approach, the US approach, and other jurisdictions (Bank of England, 2026).

3.3 Industry Governance Frameworks

3.3.1 FINOS AI Governance Framework

The FINOS AI Governance Framework v1.0, released in June 2025, provides a production-ready playbook that financial institutions can integrate into existing three-lines-of-defence models (FINOS, 2025). The framework identifies 18 top-level risk categories and 17 implementable controls aligned to the NIST Risk Management Framework, OWASP standards, and the EU AI Act (FINOS, 2025). The risk catalogue includes operational risks such as hallucination and inaccurate outputs, security risks such as supply-chain poisoning, and regulatory risks related to model concentration and compliance (FINOS, 2025). The framework recommends mitigations such as retrieval-augmented generation, human review of critical outputs, and continuous monitoring of model performance.

3.3.2 Practical Governance Practices

Forward-thinking compliance programmes are moving beyond reactive risk management towards comprehensive GenAI governance (Saifr, 2026). A strong foundation begins with establishing a cross-functional committee to review and approve all GenAI use cases prior to deployment and maintain an enterprise-wide inventory of AI applications (Saifr, 2026). Clear roles, responsibilities, and escalation procedures should be defined within this governance structure (Saifr, 2026).

Usage policies define both permitted and prohibited applications, specify acceptable use guidelines, and establish disclosure obligations when AI is used in customer interactions (RegTech Analyst, 2026). Testing protocols should address privacy, integrity, reliability, and accuracy (DLA Piper, 2025). Monitoring systems should track model performance over time, detect concept drift, and flag outputs that may contain hallucinations or biases. Recordkeeping practices must be adapted to GenAI systems. Firms should maintain logs of AI prompts, outputs, model versions, and human oversight actions (Saifr, 2026).

4. Discussion

4.1 Interpretation of Findings

The analysis reveals a governance landscape characterised by a fundamental asymmetry: the technology is advancing far more rapidly than the regulatory and governance frameworks needed to manage its risks. Regulators have made clear that existing rules apply, but they have not yet provided the detailed guidance that compliance teams need to implement effective controls. The EU AI Act provides a risk-based framework, but its implementation in the financial sector is still in its early stages. The FINRA 2026 report identifies risks and expectations but does not prescribe specific technical solutions. The result is a governance environment in which firms must navigate ambiguous regulatory expectations while managing novel risks that existing frameworks were not designed to address.

Compliance teams must interpret regulatory signals, anticipate supervisory expectations, and implement controls in the absence of clear precedents. The persistence of hallucinations, the difficulty of detecting bias, and the accountability gaps created by autonomous AI agents all point to the need for new governance capabilities that most institutions have not yet developed.

4.2 Comparison with Existing Literature

The findings align with prior analyses identifying generative AI governance as a critical challenge for financial services. The FINOS AI Governance Framework provides a comprehensive risk catalogue that extends earlier work on AI risk management in finance. The regulatory sources examined here—FINRA, EBA, MAS, and the Bank of England—reflect a growing international consensus that existing rules apply to GenAI. The present analysis extends this literature by providing an integrated assessment of regulatory expectations across multiple jurisdictions and by identifying specific governance practices that financial institutions can implement.

4.3 Persistent Challenges

Several challenges persist across jurisdictions and governance approaches. The difficulty of verifying output accuracy remains a fundamental limitation. Techniques such as retrieval-augmented generation can reduce hallucinations but cannot eliminate them entirely. The opacity of model decision-making makes it difficult to detect and mitigate bias. The accountability vacuum created by autonomous AI agents poses novel challenges for regulatory models that assume human

decision-makers at critical junctures. These are structural features of generative AI systems that require governance responses rather than technical fixes.

The economics of governance present a further challenge. Comprehensive testing and monitoring of GenAI systems require substantial investment in technical capabilities that many firms have not yet made. The rapid pace of technological change means that governance frameworks must be continuously updated. The absence of standardised testing and monitoring protocols makes it difficult for firms to benchmark their practices against industry standards.

4.4 Policy Implications

For regulators, the analysis suggests several priorities. First, the development of detailed guidance on acceptable governance practices for GenAI, including specific expectations for testing, monitoring, and recordkeeping. Second, the harmonisation of regulatory approaches across jurisdictions to reduce compliance burdens for global financial institutions. Third, the establishment of standardised testing and monitoring protocols. Fourth, the clarification of accountability for autonomous AI agents.

For financial institutions, governance cannot be an afterthought. Firms must integrate GenAI governance into existing supervisory, compliance, and risk management structures. This requires investment in technical capabilities for testing and monitoring, the development of cross-functional governance bodies, and the training of personnel on the risks and compliance obligations associated with GenAI use.

4.5 Limitations and Future Research

The analysis focuses on regulatory sources from the United States, the European Union, Singapore, and the United Kingdom, and findings may not generalise to other jurisdictions with different regulatory approaches. The rapidly evolving nature of both the technology and the regulatory landscape means that some findings may become outdated quickly. Future research should conduct comparative analyses of GenAI governance practices across different types of financial institutions and jurisdictions. The governance of autonomous AI agents, in particular, represents a frontier area that is likely to receive increasing regulatory attention.

5. Conclusion

The governance of generative artificial intelligence in financial services has emerged as a critical challenge for regulators, compliance professionals, and financial institutions. The technology

has transitioned from experimental to operational with remarkable speed, yet the governance frameworks needed to manage its risks have lagged behind. Financial regulators have made clear that existing securities laws and supervisory rules apply with equal force to GenAI-powered operations, but detailed guidance remains in development.

Three risk categories have been identified as priorities: hallucinations and inaccurate outputs, which can mislead customers and expose firms to enforcement actions; algorithmic bias and concept drift, which can perpetuate discrimination and degrade model performance over time; and the autonomy of AI agents, which creates accountability gaps that existing regulatory models were not designed to address. Governance practices for managing these risks include cross-functional oversight bodies, usage policies, testing and monitoring protocols, and recordkeeping practices adapted to GenAI systems.

Generative AI offers substantial efficiency gains for financial services, but it also carries significant risks that, if unmanaged, could undermine investor protection, market integrity, and financial stability. Ensuring responsible GenAI deployment requires sustained commitment from regulators, compliance professionals, and industry practitioners alike—a commitment to governance that keeps pace with the technology it is designed to oversee. The challenge is not merely technical or regulatory but institutional: building the governance capabilities that will enable financial institutions to harness the benefits of generative AI while managing its risks effectively and responsibly.

Author Contributions

Conceptualization, A.K.; methodology, A.K. and J.O.M.; investigation, H.T.; writing—original draft preparation, A.K.; writing—review and editing, A.K., J.O.M., H.T., and S.L.R. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Funding

This research received no external funding.

Ethics Approval

Not applicable

Acknowledgements

The authors thank the Financial Industry Regulatory Authority, the European Banking Authority, the Monetary Authority of Singapore, the Bank of England, and the FINOS community for publicly available reports and frameworks on generative AI governance. AI language tools were used to assist with language polishing. The authors also gratefully acknowledge the open access policies of their respective institutions.

Conflicts of Interest

The authors declare no conflicts of interest to report regarding the present study.

References

A&O Shearman. (2025). *EBA factsheet on implications of EU AI Act for banking and payments sector*. A&O Shearman FinReg.

Bain & Company. (2025). *Generative AI in financial services: Eight risks and how to overcome them*. Bain & Company.

Bank of England. (2025). *The Bank of England's approach to innovation in artificial intelligence, distributed ledger technology, and quantum computing*. Bank of England.

Bank of England. (2026). *Summary of AI roundtables – February 2026*. Bank of England.

DLA Piper. (2025). *FINRA flags generative AI risks and governance expectations*. DLA Piper.

European Banking Authority. (2025). *EBA Chair letter to Mr Berrigan and Mr Viola on outcome of EBA's AI Act mapping exercise (EBA/2025/D/5384)*. European Banking Authority.

European Parliament. (2025). *Resolution on the impact of artificial intelligence in the financial sector*. European Parliament.

FINOS. (2025). *FINOS AI Governance Framework v1.0*. Fintech Open Source Foundation. https://www.finos.org

FINRA. (2025). *2026 FINRA Annual Regulatory Oversight Report*. Financial Industry Regulatory Authority. https://www.finra.org

MAS. (2025). *Guidelines on artificial intelligence risk management*. Monetary Authority of Singapore. https://www.mas.gov.sg

RegTech Analyst. (2026). *Building a GenAI governance framework for FinTech firms*. RegTech Analyst.

Saifr. (2026). *Building a GenAI governance framework: Takeaways from FINRA's 2026 Oversight Report*. Saifr.

Shumaker, Loop & Kendrick, LLP. (2025). *Generative artificial intelligence in financial services: A practical compliance playbook for 2026*. Shumaker, Loop & Kendrick, LLP.