Anamika, 2026, 14:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Bridging the Intelligence Gap: Identifying GenAI and XAI Deficiencies in Banking Risk Management and Regulatory Compliance

Anamika, Amit Butail
Department of Computer Science

Abstract- The banking sector's embrace of Generative Artificial Intelligence (GenAI) and Explainable Artificial Intelligence (XAI) is likely to change how financial services firms manage risk, interpret the impact of regulation on their business, and make decisions. Although there is clear commercial interest in both technologies, academic literature has paid comparatively little attention to providing a factual account of their capabilities and limitations. This paper aims to address that gap. Using a systematic review of the literature, practitioner reports, and regulatory guidance published between 2020 and 2025, we identify existing gaps in GenAI technology as applied to credit risk, market risk, operational risk, and anti-money laundering. We also identify existing gaps in XAI techniques, ranging from SHAP-based feature attribution to counterfactual explanation. GenAI technologies present a low-to-moderate risk of hallucination, a lack of transparency in training data, and limited explainability. XAI techniques help but do not fully resolve the explainability problem, particularly in high-frequency trading environments and complex, multi-agent regulatory contexts. We conclude by proposing a pragmatic research agenda for banking technology, risk, and regulatory professionals.

Keywords: Generative AI, Explainable AI, Banking Risk, Regulatory Compliance, Model Governance, Responsible AI, Basel IV, DORA.

I. INTRODUCTION

Banking has a long history of managing uncertainty. From the credit arrangements of medieval Florentine merchants to today's multi-trillion-dollar derivatives markets, the same underlying problem persists: how to make high-impact decisions when the future is not knowable. For centuries, the answer has relied on human judgment, supported by increasingly sophisticated quantitative models. Today, these challenges are compounded by machine learning and, more recently, large language models (LLMs). These tools add new capabilities to the existing toolbox, but also introduce unique challenges that the industry has only begun to address.

Generative AI is a new class of models that, in banking, can be applied to intelligent document processing, regulatory question-and-answer chatbots, generative stress-test scenarios, and fraud-narrative detection, among other uses. Explainable AI — encompassing post-hoc, ante-hoc, and other adaptive methods — does not aim to produce outputs itself, but rather to make the rationale of complex models transparent to humans.

The relationship between the two is more than incidental. As banking models grow more complex through the integration of GenAI, the need for XAI to interpret these systems grows correspondingly. Foundational to the European Union's AI Act, the Digital Operational Resilience Act (DORA), and several of the Basel Committee's principles for effective risk data aggregation is the requirement that complex automated model outputs be accompanied by a rationale and an auditable trail.

GenAI systems will inevitably increase demand for advanced XAI. Regulatory frameworks such as the EU AI Act, the Basel Committee's Principles for Effective Risk Data Aggregation, and DORA already require — actively or passively — those significant automated decisions be explainable, auditable, and contestable. Achieving these goals with currently available tools leaves much to be desired.

The question this paper attempts to answer is deceptively simple: where are the actual gaps? Section II describes the methodology. Section III surveys GenAI use cases in banking. Section IV identifies the principal risk and compliance gaps.

Section V analyses the capabilities and limitations of XAI. Section VI synthesises these findings into a structured gap map. Section VII outlines a research agenda, and Section VIII concludes the paper.

II. METHODOLOGY

This paper adopts a literature review approach based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, adapted for technology-policy research. Literature searches were conducted using Google Scholar, SSRN, the Bank for International Settlements working-paper repository, and selected practitioner databases. Search strings included "generative AI banking risk," "explainable AI financial compliance," "LLM regulatory risk," and "XAI credit risk model governance," among other variants.

Inclusion criteria required that sources: (i) be published or officially released between January 2020 and December 2025; (ii) directly address AI or ML use in banking or financial services; and (iii) provide descriptive detail on model performance, model governance, or regulatory engagement. Technical papers describing model architectures without a use context, and vendor grey literature without a disclosed methodology, were excluded.

From this process, 74 academic papers and 18 regulatory or industry reports were identified. The collection was analysed qualitatively across five risk domains and four explanation objectives. No claims are made about statistical representativeness; the focus is on qualitative coverage of the identified gaps rather than a quantitative meta-analysis.

III. GENAI IN BANKING: A LANDSCAPE SURVEY

Credit and Lending

Large language models are now commercially available for use in loan-origination pipelines. HSBC and JPMorgan Chase have developed LLM-based systems for generating credit memos, formulating collateral-valuation narratives, and drafting borrower communication strategies. Several large Indian public-sector banks have developed similar systems. Synthetic-data approaches for small-business lending show even greater promise, using Generative Adversarial Networks or diffusion models to create data for defaulted loans where historical data is sparse and the class is imbalanced.

The advantages include lower processing times and reduced labour costs, as well as the ability to evaluate more diverse data — social media text, satellite imagery of agricultural land, trade receivables, and similar sources — that are largely unusable in traditional scorecards. However, models that synthesise narratives for credit requests may produce text that is formally convincing yet substantively incorrect. This is a significant concern, particularly where loan officers do not take the time to verify the underlying data.

B. Market and Liquidity Risk

Traders and treasury staff are using GenAI to formulate stress-test scenarios that fall outside the historical bounds mandated by regulators. This shows promise, since historical data will always miss truly unprecedented shocks, such as a cyberattack on clearing infrastructure or a central bank's sudden pivot to digital currency. In principle, generative models trained on extensive macroeconomic data can construct scenarios for situations of this kind.

Employing LLMs to synthesise research — ingesting everything from earnings calls and central bank statements to commodity news — to produce concise risk summaries for portfolio management has demonstrated clear efficiency gains. The risk, however, is that a model trained predominantly on pre-2022 interest-rate environments may overweight outdated rate assumptions and embed underlying biases into risk estimates, often without the oversight team noticing.

C. Fraud, AML, and Financial Crime

The greatest challenge lies in anti-money-laundering (AML) and counter-terrorist-financing operations. Financial intelligence units are inundated with alerts, the majority of which are false positives generated by legacy rule-based systems, while genuinely suspicious transactions are sometimes overlooked.

GenAI and graph neural networks show promise in identifying laundering typologies that require complex, multi-hop analysis.

Some regtech firms have achieved working models that reduce false positives by 40–60% while maintaining detection capacity. However, financial criminals continue to evolve their methods. A generative model may capture a typology with great confidence at one point in time, only for that typology to evolve and render the model erroneous. The time required to close the gap between a model and the evolving criminal methodology cannot be predicted, so model accuracy must be maintained on an ongoing basis.

IV. RISK AND COMPLIANCE GAPS IN CURRENT GENAI DEPLOYMENTS

A. Hallucination and Factual Reliability

Perhaps the most notorious failure mode of large language models is hallucination: LLMs produce convincing outputs and arguments that appear correct yet are factually false. The general AI literature documents this failure mode, but its consequences in banking are unique and especially concerning. For example, a credit analyst using an LLM might receive a loan summary that correctly describes a borrower's covenant breach and revenue flow, but incorrectly attributes the breach to the wrong reporting period. Such an error — likely to go unnoticed by a casual reader — could lead to a damaging credit decision.

The risk is compounded in regulatory compliance contexts. A compliance LLM that cites a superseded version of the Financial Action Task Force recommendations, or incorrectly states the Basel III capital-ratio formula, creates an exposure that is difficult to defend against. Retrieval-augmented generation techniques have reduced hallucination rates, but an error rate verified as acceptable for high-stakes banking applications has not yet been established.

B. Opacity and Model Governance

Across most jurisdictions, regulations require that models used in credit decisions, capital assessments, and compliance documentation be validated, documented, and continuously monitored. This requirement is embedded in the US Federal Reserve's SR 11-7 guidance, the EBA Guidelines on Internal Governance, and the Reserve Bank of India's model-risk-management guidelines.

A significant gap exists in validation techniques and methodologies for large model architectures, such as transformers with millions of parameters. Classical validation approaches — back-testing against historical records, sensitivity analysis, champion-challenger frameworks — are decades old and do not transfer cleanly to these architectures. A model validated as performing well may still degrade if the underlying data distribution shifts, and more problematically, existing monitoring techniques may fail to signal such a shift.

C. Adversarial and Security Risks

Unlike traditional cybersecurity threats, advanced GenAI systems in banking face continuous adversarial pressure. In a prompt-injection attack, malicious instructions embedded in documents that an LLM is asked to process could cause a compliance assistant to disregard suspicious activity reports, or a credit model to bypass risk thresholds. During fine-tuning, data poisoning can introduce systematic biases that remain undetectable until an adversary chooses to exploit them.

D. Cyber Risk and Regulatory Resilience

The Digital Operational Resilience Act (DORA), which entered into force across the EU in January 2025, imposes strict requirements for digital operational resilience, including resilience testing of AI-enabled systems. Most banking institutions have not yet developed the internal capability to conduct such testing, leaving a gap between regulatory expectation and operational reality that is likely to attract supervisory scrutiny in the near term.

V. WHAT XAI OFFERS — AND WHERE IT REACHES ITS LIMITS

A. The Promise of Explainability

XAI is a multi-part discipline. SHAP (Shapley Additive Explanations) quantifies how each feature

contributes to a model's predictions. LIME (Local Interpretable Model-Agnostic Explanations) constructs locally interpretable approximations of complex, black-box systems. Counterfactual explanation engines identify which variables would need to change for a different outcome to occur. Concept activation vectors identify high-level concepts detectable within a neural network.

The introduction of XAI tools to banking offers substantial benefits. A letter explaining the rejection of a mortgage application — an adverse-action notice — is far more useful to the applicant than a bare risk score with no explanation. SHAP-based dashboards used by credit analysts have been shown to support faster decisions and better documentation for override cases.

B. The Limits of Post-Hoc Explanation

A key limitation of XAI is that post-hoc methods explain a model's behaviour without changing it. If a model produces a biased decision and SHAP is used to explain that decision, the explanation does not make the decision fair — it merely describes the bias. This distinction matters because regulators and advocates alike sometimes treat the ability to explain a model as evidence that the model is fair, which is not correct.

A second limitation concerns the fidelity of surrogate models. LIME, in particular, approximates complex models locally, but a local approximation may diverge substantially from the model's broader behaviour. Where this divergence is significant, the resulting explanation risks being misleading. Studies of LIME's fidelity on large, complex financial datasets have found enough variability that explanations for borderline cases may be untrustworthy.

C. GenAI-Specific Explanation Challenges

The rise of generative models presents XAI with challenges not previously encountered in conventional machine learning. Attention visualisation, often treated as a post-hoc explanation method, is not in fact an explanation method: studies have shown it correlates poorly with the features that actually drive model outputs. The scale of LLMs and transformer architectures also makes some post-hoc methods, such as layer-wise relevance propagation, computationally infeasible.

An LLM-generated compliance summary or credit narrative is not a single prediction over a discrete feature vector, but a multi-step text-generation process in which the model's attention to previously generated tokens is contextually dependent and variable. Explaining such outputs is substantially more complex than explaining why a boosted-tree model produced a particular score — the latter remains within established engineering practice, while the former does not.

VI. SYNTHESISED GAP MAP

Table I presents a structured view of GenAI application areas identified in our research, the gaps associated with each, and the areas in which XAI methods offer at least partial mitigation.

TABLE I
GenAI and XAI Gap Map — Banking Risk and Compliance Domains

Dimension	GenAI Application	Current Gap	XAI Role
Credit Risk	Synthetic data generation, scenario modelling	Black-box scoring; no audit trail	Feature-attribution maps for loan decisions
Market Risk	Volatility forecasting, stress-test generation	Overfitting to training regimes; regime shifts	Counterfactual explanations for VaR breaches
Operational Risk	Fraud-narrative detection, alert triage	High false-positive rates; bias amplification	Rule extraction from LLM decision paths
Regulatory Compliance	Policy Q&A bots, regulatory-change monitoring	Hallucination risk; citation accuracy	Source-grounded explanations, confidence scores
AML / KYC	Transaction-pattern analysis, network graphs	Entity-resolution errors; privacy leakage	Graph explainability, differential-privacy layers

The table shows that no risk domain has had its GenAI-governance issues fully resolved by existing XAI methods, with gaps remaining across every domain. The most significant compliance issues arise around hallucination and legal liability, and around AML/KYC, where privacy constraints and adversarial behaviour create a moving target that current generative and explanatory tools cannot reliably address.

VII. RESEARCH AGENDA

The following research directions, identified through our analysis, should be prioritised by academics and practitioners alike.

A. Calibrated Uncertainty Quantification for Banking LLMs

Current LLM systems either provide no uncertainty estimate often resulting in confidently stated but unsupported assertions — or rely on token-level probability scores that do not align with domain-specific confidence. Banking applications require calibrated uncertainty estimates tied to measurable outcomes, such as loan-default rates, fraud-detection rates, and rates of regulatory breach, expressed in terms that risk and compliance personnel can understand and act on. Approaches based on Bayesian neural networks and conformal prediction are promising, but their application to transformer models at banking scale remains largely unexplored.

B. Adversarial Robustness Testing Standards for Financial AI

The financial industry currently lacks common principles or best practices for adversarial robustness testing of GenAI systems, in contrast to the established norms of the cybersecurity field. Establishing such standards — covering prompt injection, data poisoning, model inversion, and membership-inference attacks — is essential for meaningful compliance with DORA and similar requirements in other jurisdictions. This is primarily a regulatory research priority rather than a purely technical one.

C. Causal XAI for Temporal Financial Data

The majority of existing XAI approaches are correlational: they identify features that correlate with predictions without establishing causal relationships. For policy and liability purposes in banking, the direction of the relationship between economic variables and credit outcomes matters, and correlational explanations are insufficient. Because banking data is inherently sequential, causal XAI techniques — such as those based on potential-outcomes frameworks and do-calculus — must be adapted to high-frequency, time-series data structures.

D. Inclusive AI for Emerging-Market Banking

There is a significant and under-studied disconnect between the populations on which banking AI models are trained and the actual customer bases of Indian and other emerging-market banks. A systematic study of GenAI performance gaps across linguistic, geographical, and economic dimensions — together with the development of mitigation strategies suitable for low-resource settings — would be of substantial academic and development-finance value.

VIII. CONCLUSION

GenAI and XAI are not quick-fix solutions to the problems of banking risk management and regulatory compliance. They are powerful instruments with significant constraints, deployed in a domain where the cost of failure extends beyond financial loss to systemic stability, regulatory integrity, and public trust. The gap analysis presented in this paper will remain incomplete as technology and regulation continue to evolve rapidly. The central conclusion, however, stands: the hardest problems lie at the intersection of generative capability and explainability requirements. These are not peripheral issues to be addressed after deployment; they are design constraints that must be incorporated from the outset.

Our research agenda is ambitious but not unattainable. Each of these problems — calibrated uncertainty, adversarial standards, causal explanation, and inclusive data practice — is

tractable given sufficient interdisciplinary attention and institutional commitment. We hope this paper contributes, in some small way, to mobilising that attention. Finally, we wish to highlight the unique contribution of the global academic community convening at forums such as ICFBE in Bangalore. Financial institutions deploying these systems face commercial incentives that do not always align with systematic gap identification, and their regulators face bandwidth constraints that limit in-depth technical scrutiny. Rigorous, independent, and openly published research is therefore essential as a corrective — and that discussion begins on the floor of the conference.

REFERENCES

1. Adadi, A., & Berrada, M. (2018). Peeking Inside the Black Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138–52160.
2. Arner, D. W., Barberis, J., & Buckley, R. P. (2020). FinTech, RegTech, and the Reconceptualization of Financial Regulation. Northwestern Journal of International Law & Business, 37(3), 371–413.
3. Bank for International Settlements. (2022). Supervisory Guidance on Model Risk Management. Basel Committee on Banking Supervision, Working Paper No. 34.
4. Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable Machine Learning in Credit Risk Management. Computational Economics, 57(1), 203–216.
5. European Banking Authority. (2023). EBA Report on Big Data and Advanced Analytics. EBA/REP/2023/01. Paris: EBA.
6. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council (AI Act). Official Journal of the European Union, L 1689.
7. Financial Stability Board. (2022). Artificial Intelligence and Machine Learning in Financial Services. FSB Report, November 2022.
8. Goodfellow, I., McDaniel, P., & Papernot, N. (2018). Making Machine Learning Robust Against Adversarial Inputs. Communications of the ACM, 61(7), 56–66.
9. Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.
10. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6), Article 115.
11. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. Proceedings of KDD 2016, 1135–1144.
12. Reserve Bank of India. (2023). Discussion Paper on Responsible AI in Regulated Financial Entities. RBI/2023-24/DP-05. Mumbai: RBI.
13. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 841–887.
14. Zhou, J., Gandomi, A. H., Chen, F., & Holzinger, A. (2021). Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. Electronics, 10(5), 593.