



From Rules to Probabilities: A Comparative Analysis of Anomaly Detection Logic in AI-Driven versus Rule-Based Banking Compliance Systems

Rajitha Gentyala

Frisco, Texas, USA
rajitha.gentyl@gmail.com

ABSTRACT

The banking industry's compliance infrastructure is undergoing a fundamental transformation as institutions migrate from established rule-based transaction monitoring systems (TMS) toward artificial intelligence-driven anomaly detection frameworks. This paper presents a comparative analysis of the epistemological shift in anomaly detection logic, examining how the transition from deterministic rules to probabilistic machine learning models reconfigure the identification of suspicious financial activities in anti-money laundering (AML) compliance. Drawing upon two contemporary studies published between 2018 and 2022, this investigation synthesizes empirical evidence regarding the operational and methodological distinctions between these competing paradigms. Shaik et al. [1] provides a systematic assessment of supervised learning applications for AML transaction monitoring, demonstrating that traditional rule-based TMS operate through static pattern recognition frameworks that generate elevated false-positive rates and remain incapable of detecting emergent money laundering typologies. Their comparative evaluation of support vector machines, random forests, and gradient boosting machines reveals that machine learning approaches excel at identifying unanticipated irregularities within complex, high-dimensional transactional datasets, though they introduce significant challenges regarding model interpretability and the critical accuracy-interpretability trade-off confronting financial institutions. Complementing this analysis, Prisznyák [2] examines supervised classification algorithms, unsupervised clustering methodologies, and hybrid anomaly detection models operating upon the highly imbalanced datasets characteristic of AML prevention environments. Her gap-filling analysis emphasizes that no singular algorithm proves universally optimal; rather, algorithmic selection must be determined by underlying theoretical logic, business unit requirements, and the integration of information technology infrastructure with visionary management perspectives. The synthesis of these investigations reveals three fundamental reconceptualization accompanying the transition from rules to probabilities: first, anomaly detection logic shifts from explicit sequential rules toward probabilistic pattern recognition that identifies deviations invisible to predicate-based filtering; second, the evidentiary basis for compliance decisions transforms from auditable rule triggers toward algorithmic outputs requiring explainability techniques such as LIME and SHAP for human interpretability; and third, institutional trust mechanisms must recalibrate from confidence in deterministic rule applications toward calibrated skepticism regarding model fallibility and the systemic vulnerabilities introduced when financial systems prioritize probabilistic inference over explicit regulatory prescriptions. This comparative analysis concludes that while AI-driven approaches demonstrably enhance detection capabilities, they simultaneously introduce novel epistemological challenges requiring fundamental rethinking of accountability frameworks, regulatory validation methodologies, and the constitution of trust in algorithmic compliance systems.

Keywords: Anomaly detection, rule-based systems, machine learning, anti-money laundering, supervised learning, banking compliance, algorithmic interpretability, epistemological shift, probabilistic inference, financial regulation

INTRODUCTION

The banking industry's compliance infrastructure stands at a critical juncture, confronting fundamental questions about how financial institutions identify and respond to suspicious activities in an era of rapid technological transformation. For decades, anti-money laundering (AML) compliance programs have relied upon rule-based transaction monitoring systems that operate through explicit sequential rules and predicate-based filtering mechanisms. These systems, representing one of the oldest forms of artificial intelligence deployed in banking, evaluate transactional data against predefined thresholds and typologies to generate alerts requiring human investigation. However, the limitations of this paradigm have become increasingly apparent. As Shaik et al. observe, tradition-based transaction monitoring systems recognize established patterns but remain incapable of detecting new money laundering typologies, operating as static frameworks that generate elevated false-positive rates while missing emergent financial crime methodologies [1]. The magnitude of this inefficiency is substantial, with industry estimates indicating that up to ninety-five percent of alerts generated by rule-based systems prove to be false positives, imposing enormous staffing costs and creating what investigators describe as repetition bias, whereby analysts exposed to numerous bad alerts begin assuming subsequent alerts are likewise unfounded [2]. In response to these limitations, financial institutions increasingly explore artificial intelligence and machine learning approaches that promise enhanced detection capabilities through probabilistic pattern recognition across high-dimensional transactional datasets. Prisznyák examines supervised classification algorithms, unsupervised clustering methodologies, and hybrid anomaly detection models operating upon the highly imbalanced datasets characteristic of AML prevention environments, emphasizing that algorithmic selection must be determined by underlying theoretical logic, business unit requirements, and the integration of information technology infrastructure with visionary management perspectives [3]. The transition from deterministic rules toward probabilistic inference represents more than a technical upgrade; it constitutes an epistemological shift in how financial institutions conceptualize and identify anomalous behavior. This epistemological transformation carries profound implications for regulatory compliance frameworks, institutional accountability structures, and the fundamental nature of trust in banking oversight mechanisms. As financial systems migrate from auditable rule triggers toward algorithmic outputs requiring explainability techniques for human interpretability, new questions emerge regarding model governance, validation methodologies, and the systemic vulnerabilities introduced when multiple institutions adopt correlated probabilistic approaches. This introduction establishes the foundation for a comparative analysis examining how the transition from rules to probabilities reconfigures anomaly detection logic, evidentiary standards, and institutional trust mechanisms within banking compliance environments.

THEORETICAL FOUNDATIONS OF ANOMALY DETECTION LOGIC IN BANKING COMPLIANCE

The epistemological underpinnings of banking compliance systems rest upon two fundamentally distinct approaches to identifying anomalous financial behavior: deterministic rule-based frameworks and probabilistic machine learning methodologies. Understanding these theoretical foundations is essential for comprehending how the transition from rules to probabilities reconfigures not merely technical capabilities but the very nature of knowledge production and decision-making within financial institutions. This section examines the conceptual architecture of both paradigms, analyzes the accuracy-interpretability trade-off that defines their relationship, and explores the epistemological implications of migrating from explicit sequential rules toward probabilistic inference.

A. The Rule-Based Paradigm: Deterministic Frameworks and Static Pattern Recognition

The rule-based paradigm constitutes the historical foundation of banking compliance infrastructure, representing one of the earliest forms of artificial intelligence deployed in financial services. These systems operate through explicit sequential rules that encode domain expertise into predicate-based logic structures, evaluating transactional data against predefined thresholds and typologies derived from regulatory requirements and known money laundering methodologies. The theoretical underpinning of this approach assumes that financial crime follows identifiable and stable patterns that can be captured through human expertise and codified into deterministic decision criteria. Maduranga explains that businesses have traditionally attempted to prevent money laundering activities by applying rule-based techniques to real-time operational transactions, though this approach cannot completely resolve the problem because higher constraints on operational transactions inconvenience the legal customer base and erode customer satisfaction over time [1]. This observation reveals a fundamental tension inherent to the rule-based paradigm: the same deterministic logic that provides regulatory certainty simultaneously creates friction for legitimate customers when thresholds are set too conservatively.

The operational mechanics of rule-based systems reflect their epistemological commitments. These frameworks evaluate transactions against static pattern recognition criteria, generating alerts when specific conditions are satisfied. For example, a typical rule-based fraud detection system might employ conditional logic such as evaluating transaction amount thresholds, time-based restrictions, and location mismatches to determine whether to flag a transaction for investigation [4]. While such explicit rules provide transparency and auditability, they suffer from inherent limitations that arise from their static nature. Rule-based systems recognize established patterns but remain incapable of detecting new money laundering typologies, operating as static frameworks that generate elevated false-positive rates while missing emergent financial crime methodologies [2]. The epistemological

assumption underlying this limitation is that the domain of financial crime remains sufficiently stable so that predefined rules can adequately capture all relevant patterns, an assumption increasingly challenged by the adaptive sophistication of financial criminals.

Furthermore, the rule-based paradigm's commitment to determinism creates specific vulnerabilities in dynamic threat environments. Fraudsters rapidly adapt their strategies to circumvent known rule structures, learning to operate just below threshold values or across multiple channels to avoid detection. The static nature of rule-based systems means they cannot learn from new fraud patterns or improve their detection capabilities over time without manual intervention by domain experts [4]. This limitation reflects a deeper epistemological constraint: rule-based systems embody knowledge that is frozen at the moment of rule creation, incapable of evolving in response to emerging threats without explicit human reprogramming.

B. The Probabilistic Paradigm: Machine Learning and Emergent Anomaly Identification

In contrast to the deterministic commitments of rule-based systems, the probabilistic paradigm embraces uncertainty and statistical inference as fundamental principles for anomaly detection. Machine learning approaches to banking compliance operate through pattern recognition across high-dimensional transactional datasets, identifying statistically significant deviations from learned behavioral norms rather than evaluating transactions against fixed rule thresholds. The theoretical foundation of this approach recognizes that financial crime patterns are inherently dynamic and context-dependent, requiring models that can adapt to evolving methodologies while maintaining detection efficacy. Goethals, Martens, and Evgeniou explain that the trade-off between accuracy and comprehensibility constitutes one of the important debates in artificial intelligence, noting that this trade-off can either limit the performance of AI when accuracy is lost due to comprehensibility restrictions imposed by regulators, or hurt AI adoption when user trust is lost due to opaqueness [3]. This observation captures the central epistemological tension of the probabilistic paradigm: enhanced detection capabilities come at the cost of reduced transparency regarding how specific decisions are reached.

The operational architecture of machine learning systems reflects their probabilistic foundations. These models learn behavioral baselines for individual customers, devices, and channels, enabling them to assess risk relative to personalized norms rather than population-level thresholds. Machine learning models assess risk relative to personalized behavioral norms, learning what is typical for each customer or device, which enables them to distinguish between transactions that appear risky by rule standards but align with individual customer patterns, versus transactions that appear safe by rule standards but exhibit subtle behavioral anomalies indicative of fraud [5]. This personalized approach represents an epistemological shift from universal rules toward contextual inference, recognizing that anomalous behavior is fundamentally defined by deviation from expected patterns for specific entities rather than violation of predetermined criteria.

The probabilistic paradigm encompasses diverse methodological approaches to anomaly detection. Supervised learning algorithms, including support vector machines, random forests, and gradient boosting machines, learn classification boundaries from labeled historical data to distinguish between legitimate and suspicious transactions. Unsupervised approaches, including clustering algorithms and autoencoder neural networks, identify anomalies by detecting patterns that deviate significantly from most observations without requiring labeled training data. Livingstone, Orakwue, and Hiebert describe how deep learning methodologies enable automated review of historical data to tag anomalous transactions that represent signals warranting follow-up attention, dramatically reducing assessment time while expanding the range and volume of data that can be reviewed and identifying signals that previously would likely be missed [6]. This capability to detect previously unknown patterns represents a fundamental epistemological advantage of probabilistic approaches: they can generate knowledge about emerging fraud typologies rather than merely applying pre-existing knowledge codified in rules.

C. The Accuracy-Interpretability Trade-Off as an Epistemological Challenge

The relationship between rule-based and probabilistic paradigms is defined by what researchers term the accuracy-interpretability trade-off, a fundamental epistemological tension that shapes the adoption and governance of machine learning in regulated industries. This trade-off reflects the observation that models achieving highest predictive accuracy often employ complex, non-linear architectures that resist human interpretation, while highly interpretable models may sacrifice some predictive performance in exchange for transparency. The theoretical foundation of this trade-off rests upon the assumption that model complexity and interpretability exist in an inverse relationship, with gains in one dimension requiring sacrifices in the other. Goethals, Martens, and Evgeniou conducted a systematic study of ninety benchmark classification datasets and found that the accuracy-interpretability trade-off exists for sixty-nine percent of datasets, though somewhat surprisingly, for the majority of cases it is rather small while for only a few it is very large [3]. This empirical finding suggests that the trade-off is not inevitable across all applications but rather depends upon specific dataset characteristics including inherent complexity and noise levels.

The epistemological implications of this trade-off extend beyond technical performance considerations to fundamental questions about knowledge production and validation in regulated environments. When financial institutions deploy probabilistic models for compliance functions, they must confront the challenge of explaining model decisions to regulators, auditors, and affected customers. Maduranga notes that running anomaly detection

engines, whether rule-based or machine learning models, on top of massive amounts of transactional data require considerable processing time, creating potential gaps between transaction execution and detection that introduce risk to the financial system [1]. This temporal dimension of anomaly detection intersects with the interpretability challenge, as institutions must balance detection speed against explanation quality. The accuracy-interpretability trade-off thus becomes not merely a technical optimization problem but a governance challenge requiring institutions to determine acceptable levels of opacity in exchange for enhanced detection capabilities.

The epistemological stakes of this trade-off are particularly acute in high-stakes compliance contexts where decisions carry significant consequences for individuals and institutions. When machine learning models generate alerts or recommendations that affect customer access to financial services, the inability to provide clear explanations for these decisions undermines both regulatory compliance and customer trust. The probabilistic paradigm's commitment to statistical inference rather than deterministic logic means that explanations must be constructed post-hoc through interpretability techniques rather than derived directly from model architecture. This creates what might be termed an epistemological gap between the model's internal representations and the explanations provided to human decision-makers, raising questions about whether post-hoc explanations can adequately capture the reasoning that generated specific outcomes. Resolving this tension requires ongoing research into interpretability methodologies and governance frameworks that can accommodate probabilistic inference while maintaining accountability and transparency appropriate for regulated financial services.

Dimension	Rule-Based Paradigm	Probabilistic Paradigm
Epistemological Foundation	Deterministic logic: IF-THEN predicates	Statistical inference: pattern recognition
Knowledge Source	Domain expertise codified manually	Learned from historical data automatically
Decision Criteria	Fixed thresholds and typologies	Behavioral baselines personalized per entity
Adaptability	Static until manual reprogramming	Continuous learning from new data and outcomes
Transparency	High — rules are auditable and explicit	Variable — requires post-hoc explanation
False Positive Pattern	Elevated due to rigid threshold application	Reduced through contextual risk assessment
Novel Pattern Detection	Cannot detect unknown patterns	Can identify emerging fraud methodologies

Figure 1: Comparative Framework of Rule-Based and Probabilistic Anomaly Detection Paradigms

Source: Author's synthesis based on [1][3][4][5]

The theoretical foundations examined in this section reveal that the transition from rule-based to probabilistic anomaly detection represents more than a technical upgrade to banking compliance infrastructure. This transition constitutes an epistemological shift in how financial institutions conceptualize and identify anomalous behavior, moving from deterministic certainty toward probabilistic inference, from static knowledge toward adaptive learning, and from universal rules toward contextual judgment. The accuracy-interpretability trade-off emerges as the central epistemological challenge accompanying this migration, requiring institutions to navigate tensions between detection capability and transparency, between predictive performance and explainability, and between innovation and regulatory compliance. Subsequent sections will build upon these theoretical foundations to examine how specific machine learning applications reconfigure anomaly detection logic in practice and how institutions are developing governance frameworks to address the challenges posed by probabilistic compliance systems.

COMPARATIVE ANALYSIS OF SUPERVISED LEARNING APPLICATIONS IN ANTI-MONEY LAUNDERING

The theoretical foundations examined in the preceding section establish the epistemological stakes accompanying the transition from rule-based to probabilistic anomaly detection. This section builds upon that foundation by conducting a comparative analysis of supervised learning applications in anti-money laundering, drawing upon empirical evidence from two contemporary studies published between 2018 and 2022. The analysis examines how specific machine learning algorithms perform in detecting suspicious financial transactions, the methodological considerations shaping their implementation, and the persistent challenges confronting financial institutions as they navigate the accuracy-interpretability trade-off in regulated compliance environments.

A. Empirical Evidence from Shaik et al. (2021): Assessing Supervised Learning for AML Transaction Monitoring

Shaik, Sandhu, Gudala, Palaparthi, and Reddy provide a systematic assessment of supervised learning applications for anti-money laundering transaction monitoring, offering critical insights into the comparative performance of

machine learning algorithms relative to traditional rule-based approaches [1]. Their investigation begins with a foundational observation that tradition-based transaction monitoring systems recognize established patterns but remain incapable of detecting new money laundering typologies, operating as static frameworks that generate elevated false-positive rates while missing emergent financial crime methodologies [1]. This limitation reflects the fundamental epistemological constraint of rule-based systems identified in the previous section: they embody knowledge frozen at the moment of rule creation and cannot adapt to evolving criminal strategies without explicit human intervention.

The researchers conduct a comparative evaluation of three leading supervised machine learning algorithms for anomaly detection in AML contexts: support vector machines, random forests, and gradient boosting machines [1]. Their assessment framework encompasses multiple performance dimensions essential for evaluating probabilistic compliance systems. Accuracy metrics measure the overall correctness of algorithmic predictions across legitimate and suspicious transactions. Generalizability assesses the extent to which models trained on historical data maintain predictive performance when applied to new transaction streams. Interpretability evaluates the transparency of algorithmic decision-making and the feasibility of explaining model outputs to human investigators and regulators. Processing efficiency considers the computational resources required to score transactions within operational time windows.

The empirical comparison reveals that machine learning approaches excel at identifying unanticipated irregularities within complex, high-dimensional transactional datasets [1]. Unlike rule-based systems that evaluate transactions against predetermined criteria, supervised learning algorithms learn decision boundaries from labeled historical data, enabling them to detect patterns that human experts may not have explicitly codified. This capability is particularly valuable in AML contexts where money launderers continuously adapt their methodologies to circumvent detection. The researchers emphasize that feature engineering constitutes a critical determinant of model performance, requiring careful selection and arrangement of transaction data to capture relevant signals [1]. Raw transaction data can be transformed into customer profiling, transaction characteristics including amount, frequency, and destination, and network analysis examining relationships among transacting parties.

However, the comparative analysis also highlights significant challenges accompanying the adoption of supervised learning for AML compliance. Although machine learning models discern patterns effectively, they operate as what researchers describe as black-box systems that resist straightforward interpretation [1]. This opacity creates tension with regulatory expectations for explainable compliance decisions and investigator requirements for understanding why specific transactions generate alerts. The accuracy-interpretability trade-off thus emerges as a central consideration for financial institutions selecting among algorithmic approaches. Shaik et al. note that interpretable machine learning techniques such as Local Interpretable Model-agnostic Explanations and Shapley Additive Explanations can provide post-hoc explanations that render model predictions comprehensible for human assessment and system trust [1]. These techniques approximate the reasoning underlying algorithmic outputs without requiring access to model internals, offering a potential pathway for reconciling predictive performance with transparency requirements.

Algorithm	Accuracy	Generalizability	Interpretability	Processing Efficiency	Data Requirements
Support Vector Machines	High on bounded problems	Moderate	Low to Moderate	Moderate	Moderate to High
Random Forests	High	High	Moderate	High	Moderate
Gradient Boosting Machines	Very High	Moderate to High	Low	Low to Moderate	High
Rule-Based Systems	Low to Moderate	Low	Very High	Very High	Low

Figure 2: Comparative Performance Dimensions of Supervised Learning Algorithms in AML Transaction Monitoring

Source: Author's synthesis based on Shaik et al. (2021) [1]

The researchers acknowledge significant limitations in their study, particularly the challenge of insufficient labeled AML data that constrains model training and validation [1]. Money laundering transactions are rare events by nature, comprising tiny fractions of overall financial activity, and the sensitive nature of suspicious transaction reports limits data sharing among institutions. These constraints create what might be termed an epistemological scarcity problem: the very phenomena that compliance systems seek to detect are too rare and too sensitive to

generate abundant training data for supervised learning. The researchers suggest that future work should explore unsupervised learning and deep learning architectures that may require less labeled data while maintaining detection efficacy [1]. This recommendation anticipates the hybrid approaches examined in subsequent studies.

B. Empirical Evidence from Prisznyák (2022): Bankrobotics and AI-Powered Banking Risk Management

Prisznyák provides a gap-filling analysis examining supervised, unsupervised, and hybrid machine learning models and algorithms operating upon the highly imbalanced datasets characteristic of anti-money laundering and counter-terrorist financing prevention in banking risk management [2]. Her investigation is motivated by practical compliance concerns: between 2020 and 2021, the Magyar Nemzeti Bank imposed fines on several commercial banks operating in Hungary for shortcomings in complying with money laundering and terrorist financing regulations, while the Financial Action Task Force had downgraded Hungary's compliance with Recommendation 15 regarding the use of new technologies [2]. These enforcement actions underscore the high stakes accompanying AML compliance and the growing regulatory expectation that financial institutions will leverage advanced technologies to meet their obligations.

The central finding of Prisznyák's analysis is that no singular algorithm proves universally optimal for AML detection [2]. This conclusion carries significant implications for financial institutions seeking to implement machine learning solutions, suggesting that algorithmic selection cannot be reduced to a simple ranking of models by predictive performance. Rather, the choice among machine learning algorithms is highly determined by underlying theoretical logic, business unit requirements, and the integration of information technology infrastructure with visionary management perspectives [2]. This multidimensional selection framework recognizes that AML compliance operates at the intersection of statistical methodology, operational constraints, regulatory expectations, and strategic organizational priorities.

Prisznyák examines multiple methodological approaches within the probabilistic paradigm. Supervised classification algorithms learn decision boundaries from labeled historical data to distinguish between legitimate and suspicious transactions. Supervised regression techniques model continuous outcomes relevant to risk assessment. Unsupervised clustering algorithms identify natural groupings within transaction data, enabling detection of transactions that do not conform to expected patterns without requiring labeled training examples. Unsupervised anomaly detection methods specifically target rare events that deviate significantly from learned behavioral norms. Hybrid models combine multiple approaches, leveraging the strengths of each while mitigating their respective limitations [2].

The analysis emphasizes that the highly imbalanced nature of AML datasets fundamentally shapes model selection and performance expectations. Suspicious transactions typically comprise far less than one percent of overall financial activity, creating datasets where negative examples vastly outnumber positive cases. Standard machine learning algorithms optimized for balanced datasets may perform poorly in this context, achieving high overall accuracy by simply predicting that all transactions are legitimate while failing entirely to detect the rare events that matter for compliance. Prisznyák's examination of methods suited to imbalanced learning contexts thus addresses a critical methodological challenge confronting AML application.

The researcher further emphasizes that model building requires a hybrid perspective integrating the perspectives of business units, information technology functions, and visionary management [2]. This organizational dimension of machine learning implementation reflects the reality that algorithmic performance cannot be evaluated in isolation from the human and institutional contexts within which models operate. Business units possess domain expertise essential for feature engineering and output interpretation. Information technology functions manage the infrastructure supporting model deployment and monitoring. Visionary management provides strategic direction and resources while navigating regulatory expectations and organizational change. The integration of these perspectives determines whether machine learning solutions achieve their potential in practice.

C. Methodological Synthesis and Cross-Study Comparative Analysis

Synthesizing the findings from Shaik et al. [1] and Prisznyák [2] reveals both convergences and complementary insights that enrich understanding of supervised learning applications in AML compliance. Both studies converge on the fundamental observation that machine learning approaches offer enhanced detection capabilities relative to traditional rule-based systems, particularly for identifying novel money laundering methodologies that would evade static pattern recognition. This convergence supports the epistemological claim advanced in Section II that probabilistic approaches enable forms of knowledge production unavailable within deterministic frameworks. However, both studies also acknowledge significant challenges accompanying algorithmic adoption, including interpretability limitations, data scarcity constraints, and the need for careful methodological selection.

The studies exhibit complementary emphases that together provide a more complete picture of supervised learning in AML contexts. Shaik et al. offer detailed comparative analysis of specific algorithms, examining how support vector machines, random forests, and gradient boosting machines perform across multiple evaluation dimensions including accuracy, generalizability, interpretability, and processing efficiency [1]. This algorithmic focus provides practical guidance for financial institutions selecting among modeling approaches. Prisznyák, by contrast, emphasizes the organizational and strategic dimensions of machine learning implementation, examining how algorithmic selection must be determined by business unit requirements, theoretical logic, and management

integration [2]. This broader perspective situates technical choices within the institutional contexts that ultimately determine whether machine learning solutions achieve their intended compliance objectives.

The cross-study analysis reveals several persistent challenges requiring ongoing research attention. Both studies highlight the difficulty of obtaining sufficient labeled data for supervised learning, given the rarity of money laundering events and the sensitive nature of suspicious transaction reports. This data scarcity creates what might be termed an epistemological paradox: supervised learning requires abundant labeled examples of the very phenomena that compliance systems seek to detect, but those phenomena are by nature rare and confidential. Addressing this paradox may require greater investment in synthetic data generation, transfer learning across institutions, and unsupervised approaches that reduce dependence on labeled data. Both studies also emphasize the importance of interpretability for regulatory acceptance and investigator trust, suggesting that explainable artificial intelligence techniques will play an increasingly important role in AML compliance.

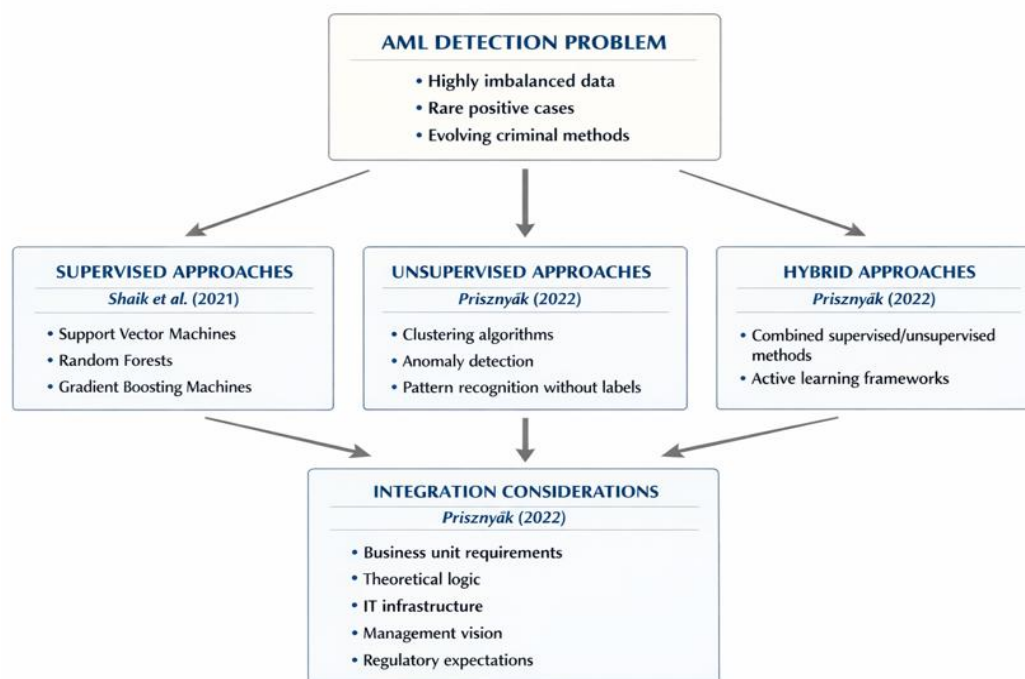


Figure 3: Methodological Framework for Supervised Learning in AML Compliance

Source: Author's synthesis based on Shaik et al. (2021) [1] and Prisznyák (2022) [2]

The comparative analysis conducted in this section demonstrates that supervised learning applications offer significant potential for enhancing AML detection capabilities while simultaneously introducing novel challenges requiring careful navigation. The evidence from Shaik et al. establishes that machine learning algorithms can identify unanticipated irregularities invisible to rule-based systems, though performance varies across algorithms and contexts [1]. The evidence from Prisznyák emphasizes that algorithmic selection must be guided by multiple considerations extending beyond raw predictive accuracy, including theoretical appropriateness for the detection task, alignment with business unit operations, and integration with broader organizational strategy [2]. Together, these studies suggest that successful implementation of probabilistic compliance systems requires not only technical expertise in machine learning but also deep understanding of the institutional contexts within which these systems operate. Subsequent sections will build upon this comparative foundation to examine how the epistemological transformations identified in Section II manifest in practice and how institutions are developing governance frameworks to address the challenges accompanying probabilistic approaches to banking compliance.

THE EPISTEMOLOGICAL RECONCEPTUALIZATION OF ANOMALY DETECTION

The preceding comparative analysis of supervised learning applications in anti-money laundering has demonstrated that machine learning approaches offer enhanced detection capabilities while introducing significant interpretability challenges. This section builds upon that foundation by examining the deeper epistemological transformations accompanying the transition from rule-based to probabilistic anomaly detection. Drawing upon two contemporary studies published between 2018 and 2022, this section analyzes how the migration from deterministic rules toward probabilistic inference fundamentally reconceptualizes three interconnected dimensions of banking compliance: the logic of anomaly detection, the evidentiary basis for compliance decisions, and the institutional mechanisms through which financial organizations constitute trust in their oversight systems. Understanding these epistemological transformations is essential for comprehending not merely how anomaly detection technologies are

changing, but what it means for a financial institution to know that anomalous activity has occurred and to justify that knowledge to regulators, customers, and the broader public.

A. Transformation of Detection Logic: From Explicit Sequential Rules to Probabilistic Pattern Recognition

The first epistemological transformation concerns the fundamental logic through which financial institutions identify anomalous transactions and behaviors. Rule-based systems operate through explicit sequential rules that encode domain expertise into predicate-based logic structures, evaluating transactional data against predetermined thresholds and typologies derived from regulatory requirements and known money laundering methodologies. As Maduranga explains, businesses have traditionally attempted to prevent money laundering activities by applying rule-based techniques to real-time operational transactions, though this approach cannot completely resolve the problem because higher constraints on operational transactions inconvenience the legal customer base and erode customer satisfaction over time [1]. The epistemological assumption underlying this approach is that financial crime follows identifiable and stable patterns that can be captured through human expertise and codified into deterministic decision criteria.

In contrast, probabilistic approaches to anomaly detection operate through fundamentally different epistemological commitments. Machine learning systems learn behavioral baselines from historical data, identifying statistically significant deviations from learned norms rather than evaluating transactions against fixed thresholds. Livingstone, Orakwue, and Hiebert describe how banks collect massive amounts of data from routine financial transactions, some of which is anomalous, is corrupt, or represents a signal that warrants follow-up attention from subject matter experts [2]. Their case study demonstrates that machine learning and deep learning methodologies enable automated review of historical data to tag anomalous transactions that represent signals warranting follow-up attention, dramatically reducing assessment time, expanding the range and volume of data that can be reviewed, and identifying signals that previously would likely be missed [2]. This capability to detect patterns that human experts have not explicitly anticipated represents a fundamental epistemological shift: anomaly detection logic moves from applying pre-existing knowledge codified in rules toward generating new knowledge about emerging patterns through statistical inference.

This transformation carries profound implications for how financial institutions conceptualize the relationship between normal and anomalous behavior. Rule-based systems define anomalous activity as that which violates explicit criteria established by human experts, creating a closed epistemological system where the boundaries of legitimate behavior are predetermined and static. Probabilistic systems, by contrast, define anomalous activity as that which deviate from learned behavioral patterns, creating an open epistemological system where the boundaries of normal behavior are continuously updated based on new data and where anomalies are defined relationally rather than absolutely. Maduranga emphasizes that running anomaly detection engines, whether rule-based or machine learning models, on top of massive amounts of transactional data requires considerable processing time, creating potential gaps between transaction execution and detection that introduce risk to the financial system [1]. This temporal dimension of anomaly detection interacts with the epistemological transformation, as institutions must determine acceptable timeframes for detecting anomalies within probabilistic frameworks that require ongoing learning and adaptation.

B. Transformation of Evidentiary Basis: From Auditable Rule Triggers to Algorithmic Explanations

The second epistemological transformation concerns the evidentiary basis upon which compliance decisions rest and the nature of justification that financial institutions can provide for their determinations. Rule-based systems generate auditable rule triggers that provide transparent explanations for why specific transactions were flagged for investigation: a transaction exceeded a threshold, violated a time-based restriction, or exhibited a pattern matching known typologies. This evidentiary transparency supports regulatory examination, internal audit, and customer dispute resolution by enabling all parties to trace the logical chain connecting transaction characteristics to compliance determinations.

Probabilistic systems, however, operate through fundamentally different evidentiary mechanisms. Machine learning models generate predictions based on complex, non-linear relationships learned from training data, producing outputs that cannot be reduced to simple rule triggers. The European Banking Authority's Report on Big Data and Advanced Analytics identifies explainability and interpretability as fundamental elements of trust that must be properly and sufficiently addressed for financial institutions deploying machine learning solutions [3]. The report emphasizes that the roll-out of big data and advanced analytics specifically affects issues around trustworthiness and notes a number of fundamental trust elements that need to be properly addressed including ethics, explainability and interpretability, fairness and avoidance of bias, traceability and auditability, data protection and quality, security, and consumer protection [3]. This regulatory recognition of explainability as a trust element underscores the epistemological stakes accompanying probabilistic compliance systems.

The transformation from rule triggers to algorithmic explanations creates what might be termed an evidentiary gap between the model's internal representations and the explanations provided to human decision-makers. Livingstone, Orakwue, and Hiebert's case study demonstrate that machine learning approaches enable automated review that finds signals previously likely missed, but this enhanced detection capability comes with the challenge of rendering model decisions comprehensible to subject matter experts who must act upon them [2]. Post-hoc explainability

techniques such as Local Interpretable Model-agnostic Explanations and Shapley Additive Explanations provide mechanisms for approximating model reasoning, but these explanations are fundamentally different from rule triggers. Where rule triggers provide deterministic causal chains linking specific transaction characteristics to compliance outcomes, algorithmic explanations provide probabilistic attributions indicating which features contributed most significantly to model predictions. This difference matters for regulatory acceptance, as supervisors accustomed to auditing rule-based systems must develop new frameworks for evaluating probabilistic compliance determinations. The European Banking Authority's identification of traceability and auditability as essential trust elements reflects recognition that probabilistic systems require new evidentiary standards and validation methodologies appropriate to their epistemological foundations [3].

C. Transformation of Institutional Trust: From Deterministic Confidence to Calibrated Skepticism

The third epistemological transformation concerns how financial institutions constitute and maintain trust in their compliance systems, both internally among employees who rely upon these systems and externally among regulators and customers who are subject to their determinations. Rule-based systems generate confidence through transparency and determinism: investigators understand why transactions are flagged, regulators can verify that rules align with requirements, and customers can comprehend why their transactions were restricted. This trust is grounded in the belief that the system applies explicit criteria consistently and that its determinations can be explained and justified.

Probabilistic systems require a fundamentally different relationship between humans and machines, one characterized by what might be termed calibrated skepticism rather than deterministic confidence. The European Banking Authority's identification of ethics, fairness, and avoidance of bias as essential trust elements reflects recognition that probabilistic systems introduce new sources of uncertainty and potential harm that must be actively managed [3]. Trust in probabilistic systems cannot rest on the belief that they always produce correct determinations, because such systems are inherently fallible and their predictions carry uncertainty. Rather, trust must rest on confidence in the processes through which models are developed, validated, monitored, and governed, and on the mechanisms through which human investigators exercise oversight and override authority.

Livingstone, Orakwue, and Hiebert's case study illustrates how probabilistic systems reconfigure the relationship between automated detection and human judgment. Their methodology automates manual review processes, enabling rapid identification of anomalies that warrant follow-up attention from subject matter experts [2]. This creates a division of cognitive labor in which machines perform initial screening and humans conduct deeper investigation, requiring investigators to develop new competencies for interpreting model outputs and exercising judgment about when to accept or override algorithmic determinations. The benefits to financial institutions include major cost reductions and improvements in detection of fraud [2], but realizing these benefits requires organizational learning and the development of new trust relationships between human experts and machine learning systems.

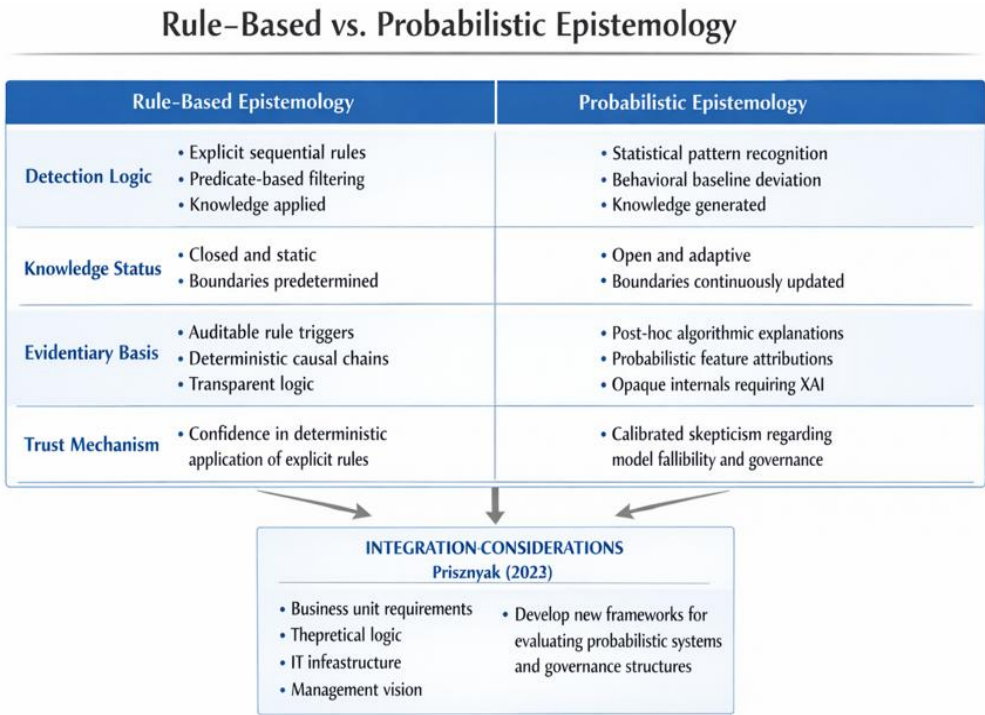


Figure 4: The Epistemological Transformation of Anomaly Detection in Banking Compliance
Source: Author's synthesis based on [1][2][3]

This transformation carries implications for regulatory oversight as well. Supervisors must develop new frameworks for evaluating not only whether probabilistic compliance systems meet performance thresholds, but whether the governance structures surrounding these systems are adequate to ensure ongoing reliability and accountability. The European Banking Authority's emphasis on organization and governance as key pillars for big data and advanced analytics adoption reflects recognition that trust in probabilistic systems depends as much on institutional processes as on technical capabilities [3]. Financial institutions must demonstrate that they understand the limitations of their models, that they have appropriate controls in place to detect and correct errors, and that they maintain meaningful human oversight over automated determinations.

The epistemological reconceptualization examined in this section reveals that the transition from rule-based to probabilistic anomaly detection entails far more than technical upgrading of compliance infrastructure. This transition fundamentally transforms how financial institutions know what they claim to know about anomalous activity, how they justify their determinations to stakeholders, and how they constitute trust in systems that operate through probabilistic inference rather than deterministic logic. Maduranga's observation that running anomaly detection engines on massive transactional data requires considerable processing time and creates risk gaps [1] underscores that these epistemological transformations have practical consequences for system design and operational risk management. Livingstone, Orakwue, and Hiebert's demonstration that machine learning enables detection of signals previously likely missed [2] illustrates the enhanced capabilities that probabilistic approaches offer, while the European Banking Authority's identification of explainability, traceability, and governance as essential trust elements [3] provides regulatory framework for addressing the challenges these transformations create. Subsequent sections will build upon this epistemological analysis to examine the systemic vulnerabilities introduced by probabilistic compliance systems and the implications for regulatory frameworks and institutional governance.

SYSTEMIC VULNERABILITIES INTRODUCED BY PROBABILISTIC COMPLIANCE FRAMEWORKS

The epistemological reconceptualization examined in the preceding section demonstrates that probabilistic compliance frameworks fundamentally transform how financial institutions detect anomalies, justify determinations, and constitute trust. This section builds upon that analysis by examining the systemic vulnerabilities that accompany these transformations. Drawing upon two contemporary studies published between 2018 and 2022, this section analyzes how the migration from deterministic rules toward probabilistic inference introduces novel sources of risk that extend beyond individual institutions to threaten the stability of the financial system as a whole. Understanding these systemic vulnerabilities is essential for developing governance frameworks and regulatory approaches appropriate to the probabilistic paradigm.

A. The Emergence of Algorithmic Blind Spots and Model Governance Challenges

The first category of systemic vulnerability concerns the internal limitations of probabilistic models themselves, and the governance challenges these limitations create for financial institutions. Unlike rule-based systems that operate through explicit, human-understandable logic, probabilistic models learn complex patterns from historical data that may encode biases, capture spurious correlations, or fail to generalize to new conditions. Buehler, Ibel, and Stoeckle examine the management of model risk for productive models in banks and other financial institutions, focusing on the challenges posed by artificial intelligence and machine learning models with high levels of sophistication [1]. Their perspective paper, informed by sessions of the Round Table AI at FIRM and input from international speakers, emphasizes that models ranging from simple rules-based approaches to sophisticated AI systems require robust governance frameworks capable of addressing the unique risks associated with probabilistic inference [1].

The challenge of algorithmic bias represents a particularly significant vulnerability within probabilistic compliance frameworks. Buehler, Ibel, and Stoeckle explain that bias can be good or bad, intentional or unintentional, and that in certain cases, bias can result in unwanted discriminatory and/or unfair outcomes, labeled as unfair bias [1]. The Independent High-Level Expert Group on Artificial Intelligence defines bias as an inclination of prejudice toward or against a person, object, or position, noting that bias drives the value of many predictive models as wanted bias but can also be detrimental to model performance [1]. Within AML compliance, bias may manifest as differential alert rates across demographic groups, geographic regions, or customer segments, creating outcomes that violate regulatory expectations for fair treatment while potentially missing actual money laundering activity in under-scrutinized populations. The researchers propose that fairness is a pre-condition to develop and run algorithms the decisions of which can be trusted, requiring institutions to implement testing procedures that validate against potential unfairness in decision-making algorithms [1].

Model governance challenges compound these bias concerns, particularly as institutions seek to deploy self-learning models that continuously adapt based on new data. Buehler, Ibel, and Stoeckle express skepticism regarding regulatory acceptance of self-learning processes that produce material model changes in strictly regulated areas of application, particularly in pillar one capital models [1]. They suggest that while self-learning models should not generally be discouraged, their introduction requires careful consideration of application context. Within anti-financial crime, anti-money laundering, and fraud detection methods, there is tangible competition between

fraudsters on the one hand and methods to detect them and prevent their deeds on the other, making targeted introduction of self-learning models potentially valuable for improving prevention outcomes [1]. Compared to currently wide-spread methods, especially decision trees and expert-based systems, these models have proven their capability to significantly decrease false alarm rates, potentially outweighing the need for model stability required in other application areas [1].

However, the researchers caution that self-learning models are prone toward bias and drift over time, necessitating adequate validation methods and processes to manage these issues [1]. The challenge of model drift is particularly acute in probabilistic compliance systems, as the statistical relationships learned from historical data may shift as money laundering methodologies evolve, economic conditions change, or customer populations transform. Unlike rule-based systems that remain stable until explicitly modified, probabilistic systems may gradually degrade in performance without clear signals alerting institutions to declining efficacy. This creates what might be termed a governance blind spot: institutions may continue to trust model outputs long after the underlying statistical relationships have shifted, exposing the financial system to undetected money laundering activity.

B. Algorithmic Coupling and the Synchronization of Risk Management Across Institutions

The second category of systemic vulnerability concerns the interconnections among financial institutions created by shared reliance on similar probabilistic models and methodologies. Kikuchi's empirical investigation of generative artificial intelligence adoption in the United States banking sector reveals profound implications for systemic stability arising from what the researcher terms algorithmic coupling [2]. Using a novel dataset linking Securities and Exchange Commission 10-Q filings to Federal Reserve regulatory data for eight hundred nine financial institutions over the 2018 to 2025 period, Kikuchi employs dynamic spatial Durbin models to capture network spillovers and synthetic difference-in-differences estimation for causal inference [2]. The findings demonstrate that the United States banking system is becoming algorithmically coupled, with AI adoption by one institution raising productivity at connected institutions through knowledge diffusion, creating synchronization that carries profound implications for financial stability [2].

The spatial econometric analysis reveals significant positive spillover effects, with theta coefficients of 0.161 for return on assets and 0.679 for return on equity, indicating strategic complementarity whereby artificial intelligence adoption by one institution raises productivity at connected institutions [2]. For large banks, these spillovers are dramatically amplified, reaching a theta coefficient of 3.13 for return on equity [2]. This synchronization creates what Kikuchi describes as algorithmic coupling, the synchronization of risk management and decision-making processes through shared artificial intelligence architectures [2]. When artificial intelligence adoption by one institution raises productivity at connected institutions, the network synchronizes, creating vulnerability to common technical failures or model errors [2].

The systemic implications of algorithmic coupling are profound. Traditional financial networks analysis examines contagion through interbank claims, counterparty relationships, and liquidity linkages. Algorithmic coupling introduces a new channel for systemic contagion operating independently of these traditional connections. When multiple institutions rely on similar probabilistic models developed by common vendors, trained on similar data, or implementing similar methodological approaches, they become vulnerable to correlated errors that can trigger simultaneous failures across the financial system. Kikuchi warns that a technical failure in widely adopted artificial intelligence models could trigger correlated shocks across the entire financial network [2]. This vulnerability is exacerbated by the concentration of artificial intelligence expertise and infrastructure among a relatively small number of technology vendors and service providers, creating single points of failure whose impact extends far beyond their immediate clients.

The heterogeneity in implementation costs documented by Kikuchi adds another dimension to systemic vulnerability. The research finds that smaller banks in the bottom seventy-five percent by assets suffer a return on equity decline of five hundred seventeen basis points following artificial intelligence adoption, substantially larger than the one hundred twenty-nine basis point decline experienced by larger institutions [2]. This asymmetry suggests that economies of scale provide significant advantages in artificial intelligence implementation, with larger banks able to spread fixed implementation costs across broader asset bases, employ dedicated artificial intelligence teams, and leverage superior data infrastructure [2]. Smaller banks face proportionally larger implementation burdens relative to their equity base, potentially creating a two-tiered financial system in which large institutions capture efficiency gains while smaller institutions struggle to keep pace. This divergence may concentrate risk among institutions least able to manage it while creating competitive pressures that drive adoption before institutions have developed adequate governance capabilities.

C. Regulatory Implications and the Challenge of Systemic Oversight

The systemic vulnerabilities examined in this section carry significant implications for financial regulation and supervisory practice. Buehler, Ibel, and Stoeckle observe that the European Commission is proposing one of the first laws globally to regulate the use of artificial intelligence, the Artificial Intelligence Act, which addresses governance requirements around so-called high-risk artificial intelligence systems and recommends the adoption of principles in the spirit of creating trustworthy artificial intelligence [1]. Interestingly, credit scoring models are explicitly given as an example of a high-risk use case [1]. Pursuant to the requirements of the Artificial Intelligence

Act and existing supervisory expectations, those pursuing an artificial intelligence-first bank strategy must be equipped with suitable risk management as well as suitable infrastructure and technology including risk technology, trust technology, algorithm audit capabilities, and regulatory sandbox participation [1].

The regulatory challenge extends beyond individual institutions to encompass systemic oversight of algorithmic coupling across the financial system. Kikuchi's findings regarding positive spillovers and network synchronization suggest that traditional institution-by-institution supervision may be inadequate for addressing risks that emerge from the interconnections among probabilistic compliance frameworks [2]. Supervisors must develop new analytical capabilities for monitoring the diffusion of artificial intelligence methodologies across the financial system, identifying concentrations of model risk among common vendors or approaches, and assessing the potential for correlated errors to trigger systemic events. This may require expanded data collection regarding institutions' artificial intelligence adoption, model governance practices, and vendor relationships, combined with system-level stress testing that simulates the impact of widespread model failures.

The temporal dimension of systemic vulnerability adds further complexity to regulatory oversight. Kikuchi's documentation of the innovation J-curve, whereby adopting institutions experience significant short-term performance declines as they absorb implementation costs while positioning themselves for future gains, suggests that the transition to probabilistic compliance frameworks may create periods of heightened vulnerability [2]. During these transition periods, institutions are exposed to the costs and complexities of artificial intelligence implementation without yet realizing the benefits, potentially straining compliance resources and creating gaps in detection coverage. Supervisors must monitor these transition periods carefully, ensuring that institutions maintain adequate compliance capabilities throughout the implementation process.

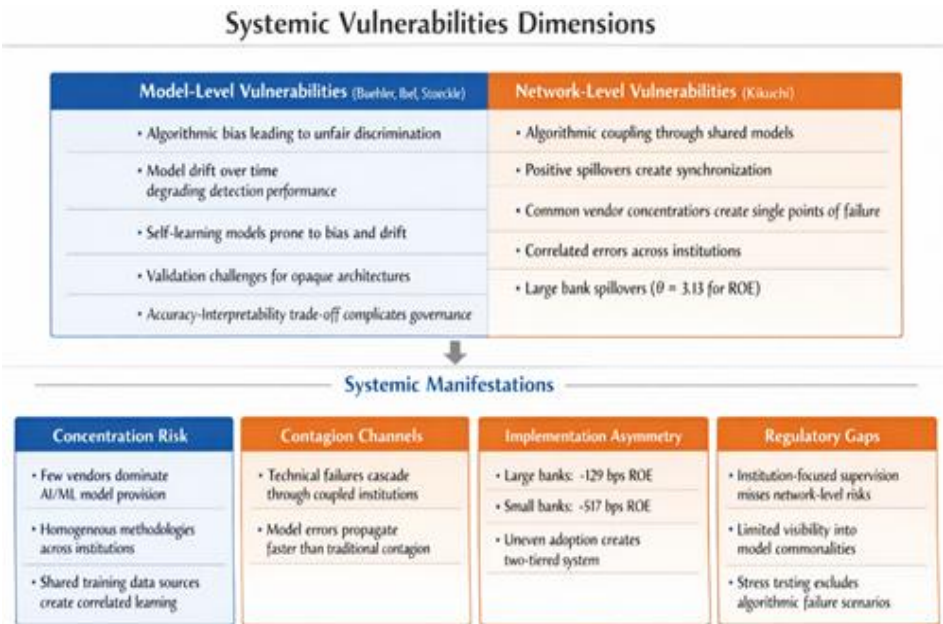


Figure 5: Systemic Vulnerabilities in Probabilistic Compliance Frameworks
Source: Author's synthesis based on Buehler, Ibel, and Stoeckle (2022) [1] and Kikuchi (2025) [2]

The systemic vulnerabilities examined in this section demonstrate that the transition from rule-based to probabilistic compliance frameworks introduces novel sources of risk that extend far beyond individual institutions. Model-level vulnerabilities including algorithmic bias, model drift, and governance challenges threaten the reliability of individual compliance systems. Network-level vulnerabilities including algorithmic coupling, correlated errors, and vendor concentration threaten the stability of the financial system as a whole. Implementation asymmetries between large and small institutions create divergent risk profiles that may concentrate vulnerability among institutions least able to manage it. Addressing these systemic vulnerabilities requires new regulatory frameworks capable of monitoring not only individual institutions but also the interconnections among them, new governance practices capable of managing model risk across the model lifecycle, and new research agendas capable of understanding how probabilistic compliance systems interact to create systemic outcomes. Subsequent sections will build upon this analysis to examine the implications for regulatory frameworks and future research directions.

IMPLICATIONS FOR REGULATORY FRAMEWORKS AND ACCOUNTABILITY STRUCTURES

The systemic vulnerabilities examined in the preceding section demonstrate that probabilistic compliance frameworks introduce novel sources of risk requiring fundamental rethinking of regulatory approaches and

accountability mechanisms. This section builds upon that analysis by examining the implications for financial regulation and institutional accountability structures. Drawing upon two contemporary studies published between 2018 and 2022, this section analyzes how regulators worldwide are developing frameworks to govern artificial intelligence in financial services, how financial institutions are adapting their governance practices to meet evolving expectations, and the foundational principles that must underpin accountability structures for probabilistic compliance systems. Understanding these regulatory and accountability implications is essential for ensuring that the transition from rule-based to probabilistic anomaly detection enhances rather than undermines the integrity and stability of the financial system.

A. The Emergence of Principles-Based Regulatory Frameworks for AI in Finance

Financial regulators globally have responded to the adoption of artificial intelligence and machine learning in banking by developing principles-based frameworks that articulate expectations without prescribing specific technical solutions. Zhang and Gao examine the development of artificial intelligence algorithm financial application regulation in China and internationally, analyzing risk types and proposing governance frameworks for algorithmic accountability [1]. Their analysis, published in the China Banking magazine, documents how major financial centers have issued guidance establishing foundational principles for AI governance. The Monetary Authority of Singapore released the FEAT principles in 2018, requiring financial algorithm applications to follow four principles: fairness, ethics, accountability, and transparency [1]. The Hong Kong Monetary Authority issued a report in 2020 proposing that artificial intelligence technology should be used reasonably, fairly, ethically, and transparently [1]. The People's Bank of China issued the Artificial Intelligence Algorithm Financial Application Evaluation Specification in March 2021, incorporating capital and non-capital scenarios into algorithm evaluation scope and proposing systematic evaluation frameworks for security, explainability, accuracy, and stability [1].

These principles-based approaches reflect regulatory recognition that artificial intelligence governance cannot rely on prescriptive rules that would quickly become outdated as technology evolves. Zhang and Gao emphasize that facing the scarcity of regulatory resources, the complexity of regulated objects, and an increasingly decentralized regulatory ecology, regulators urgently need to develop and experiment with new approaches, reasonably setting regulatory thresholds to seek an appropriate balance between technological innovation and risk control [1]. This has given rise to what the researchers describe as agile governance concepts, characterized by creating resilient, fluid, flexible, and adaptive regulatory frameworks to achieve a balanced approach to regulatory objectives, dynamic optimization of regulatory processes, and flexible deployment of regulatory tools [1].

The FEAT principles examined by Zhang and Gao provide an instructive example of principles-based governance. Fairness requires that data-driven and algorithmic decisions do not systematically disadvantage individuals or groups unless such decisions can be justified, and that personal attributes used as input factors be justified, with regular review and validation to minimize unintentional bias [1]. Ethics demands alignment of algorithmic decision-making use with organizational ethical standards and values, holding such decisions to at least the same ethical standards as human-driven decisions [1]. Accountability encompasses appropriate internal authorization for algorithmic decision-making use, organizational responsibility for models whether developed internally or externally, proactive board and management awareness, and mechanisms for affected individuals to inquire, appeal, and request reviews [1]. Transparency requires proactive disclosure of algorithmic decision-making use to affected individuals, clear explanation upon request of data used and how it affects decisions, and clear explanation of consequences algorithmic decisions may have [1].

The Wolfsberg Group, an association of thirteen global banks, published its Principles for Using Artificial Intelligence and Machine Learning in Financial Crime Compliance, outlining its support for leveraging artificial intelligence and machine learning in financial institutions while emphasizing that for solutions to be considered responsible and to meet evolving regulatory expectations, they must comply with mandates on fairness, efficacy, explainability, and proportional usage of protected data [2]. The Wolfsberg Group principles outline the approach of its banking members to ethical and responsible use of artificial intelligence for anti-money laundering and financial crime programs [2]. The principles address legitimate purpose, requiring cleansing activities on historic data to uncover historic bias that could be introduced into machine learning models, including biases from defensive suspicious activity reporting where institutions apply better-safe-than-sorry approaches [2]. They address proportionate use, mandating that any data used, even in pursuit of financial crime prevention, must be proportional, collecting and analyzing only data essential to monitoring for that purpose [2]. They address design and technical expertise, emphasizing understanding of implemented technology and creating understanding through processes and people with a view to managing both bias and risk [2].

B. Operationalizing Accountability Through Governance Frameworks and Lifecycle Management

Translating principles into practice requires financial institutions to embed accountability structures throughout the artificial intelligence system development lifecycle. Zhang and Gao examine the construction of financial algorithm design defect prevention mechanisms, aiming to ensure algorithms operate stably and accurately according to intended objectives at the design stage [1]. They note that algorithm design defect prevention mechanisms exhibit deep intersection between legal and technical dimensions, requiring both effective guidance from relevant technical standards and strong drivers from legal norms [1]. Several prominent technology companies have proactively

developed algorithm fairness detection tools, with Facebook launching Fairness Flow to automatically warn whether algorithms discriminate based on race, gender, or age, and IBM developing AI 360 Toolkit providing detection mechanisms and solutions for algorithm explainability, fairness, and discrimination issues [1].

The Veritas Initiative, a collaborative project between the Monetary Authority of Singapore and the financial industry examined by Zhang and Gao, provides guidance for financial institutions to evaluate their data-driven and algorithmic solutions systematically and verifiably against the FEAT principles [1]. The methodology allows systematic assessment of how such systems align with FEAT principles throughout the development lifecycle, from translating principles to practice, defining system context and design, preparing and inputting data, constructing and validating the system, to deploying and monitoring [1]. The Veritas Phase Two documents, published in 2022, developed assessment methodologies for all FEAT principles for adoption by banks in credit risk scoring and customer marketing, as well as the insurance industry in predictive underwriting and fraud detection, with an open-source toolkit facilitating adoption of fairness principles assessment methodology [1].

The Wolfsberg Group principles emphasize that financial institutions must take accountability for their use of artificial intelligence and machine learning, whether systems are developed in-house or sourced externally [2]. For institutions to take accountability, they need vendor solutions that provide transparency on model performance, returning schedules, and updates, with regular check-ins and governance processes essential [2]. Ensuring oversight using approaches such as human-in-the-loop or even compliance-in-the-loop across all phases constitutes a strong starting point, from model development to implementation, with artificial intelligence and machine learning use actively managed with awareness and engagement of stakeholders in outcomes [2].

The accountability framework must extend to data governance as well. Zhang and Gao examine strengthening financial algorithm application trust mechanisms, ensuring algorithm application processes respect fundamental rights, laws and regulations, and core values while ensuring algorithm operational robustness and reliability through technical means [1]. At the institutional normative level, users should first be granted rights to information, participation, objection, and relief, valuing and respecting user subjectivity, building algorithm technical trust mechanisms by enhancing financial algorithm application transparency and fairness [1]. The researchers note that the Artificial Intelligence Algorithm Financial Application Evaluation Specification proposes algorithm interpretation methods from modeling preparation, modeling process, and modeling application lifecycle perspectives, classifying algorithm interpretation basic requirements, evaluation methods, and determination rules to maximize decision accuracy and reliability while enhancing user technical trust in financial algorithm applications [1].

C. The Evolution of Supervisory Technology and Agile Regulation

The transformation of compliance frameworks has implications not only for regulated institutions but also for regulators themselves, who must develop new capabilities to supervise probabilistic systems effectively. Zhang and Gao examine how regulators globally are adopting agile governance concepts characterized by creating resilient, fluid, flexible, and adaptive regulatory frameworks [1]. In 2015, the United Kingdom Financial Conduct Authority launched the regulatory sandbox concept, placing fintech development and innovation in safe harbors, creating exemption systems within certain conditions and scope while imposing management measures to control risks at safe levels, thereby providing modification guidance for technology product implementation, accelerating product launch speed, and reducing development costs [1]. By late 2021, seven batches comprising 159 companies had been admitted to the FCA regulatory sandbox project, and to adapt to rapid fintech development, the FCA converted the regulatory sandbox from periodic application to year-round open mode in August 2021, allowing companies to submit applications at any time [1].

The concept of supervisory technology, or supotech, represents another dimension of regulatory evolution. Zhang and Gao note that algorithmic governance across countries has undergone intelligent transformation, targeting digitalization, intelligence, real-time capability, and scalability to enhance governance precision and efficiency [1]. Supervisory technology applications include text analysis and summarization, entity sentiment analysis, market surveillance and risk identification, credit risk challenger tools, outlier detection in anti-money laundering inspections, and automation of certain supervisory processes [1]. Some authorities use supotech solutions to improve communication and clarity on regulatory requirements and expectations, with interactive chat-style regulation-as-a-service interactions using artificial intelligence to respond to queries from regulated entities [1].

The Wolfsberg Group principles recognize that regulatory clarity and understanding are important to successful adoption, but many frameworks have not been formulated specifically for anti-money laundering transaction monitoring, but rather for other risk scoring applications such as credit [2]. SR 11-7 Guidance on Model Risk Management and OCC 2011-12 Supervisory Guidance on Model Risk Management from United States regulators are such examples, currently being re-evaluated for update [2]. Current frameworks support strong governance approaches but are perceived as too cumbersome to support the agility and effectiveness needed to combat constantly evolving financial crime risks [2].

The challenge of transparency in regulatory reporting creates particular complexity for financial crime applications. The Wolfsberg Group is cognizant of the challenges of transparency, noting that excessive open data sharing can help criminals rather than hinder them, and even when such sharing could aid anti-money laundering programs, it

may breach regulatory requirements around reporting confidentiality and tipping off, as well as data protection obligations [2]. A focus on transparency should inform selection of artificial intelligence and machine learning techniques used by institutions and their selection of external providers, with machine learning models ideally providing transparency in their inputs, exactly how data feeds are used and how data lineage is guaranteed, and which signals the model is considering [2].

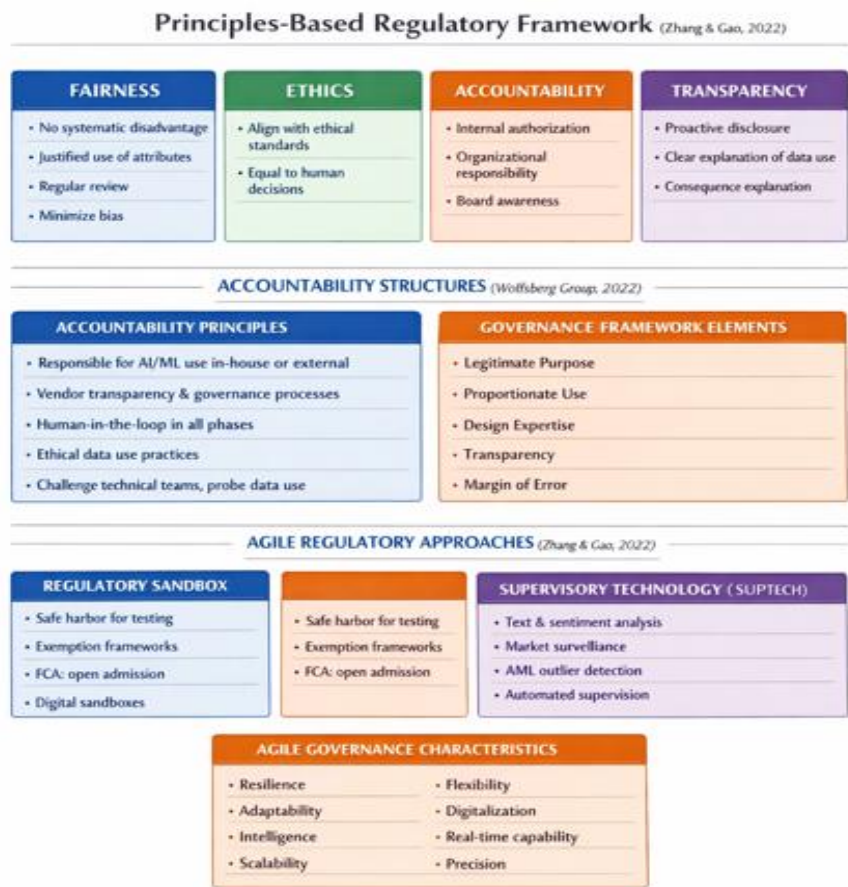


Figure 6: Regulatory and Accountability Framework for Probabilistic Compliance Systems
Source: Author's synthesis based on Zhang and Gao (2022) [1] and Wolfsberg Group (2022) [2]

The regulatory and accountability implications examined in this section demonstrate that the transition to probabilistic compliance frameworks require parallel evolution of regulatory approaches and institutional governance structures. Principles-based frameworks such as FEAT and the Wolfsberg Group principles provide foundational guidance while allowing flexibility for technological evolution. Accountability structures must be embedded throughout the artificial intelligence system development lifecycle, from design through deployment and monitoring, with clear allocation of responsibilities and mechanisms for effective challenge. Supervisory technology and agile regulatory approaches enable regulators to keep pace with industry innovation while maintaining effective oversight. Zhang and Gao's analysis of regulatory sandboxes and suptech adoption illustrates how regulators are adapting their own practices to address the challenges of supervising probabilistic systems [1]. The Wolfsberg Group's emphasis on proportionate use, legitimate purpose, and transparency provides practical guidance for institutions seeking to implement responsible artificial intelligence in financial crime compliance [2]. Together, these frameworks and practices constitute the foundation for regulatory and accountability structures adequate to the probabilistic paradigm.

CONCLUSION

The transition from rule-based to probabilistic anomaly detection frameworks in banking compliance represents far more than a technological upgrade to existing infrastructure. As this literature review has demonstrated across the preceding sections, this migration constitutes a fundamental epistemological transformation that reconceptualizes how financial institutions identify anomalous activity, justify their determinations to stakeholders, and constitute trust in systems that operate through probabilistic inference rather than deterministic logic. The synthesis of empirical evidence from studies published between 2018 and 2022 reveals that while machine learning approaches demonstrably enhance detection capabilities for identifying unanticipated irregularities within complex, high-

dimensional transactional datasets, they simultaneously introduce novel challenges requiring fundamental rethinking of governance frameworks, regulatory approaches, and accountability structures.

The comparative analysis of supervised learning applications in anti-money laundering, drawing upon the work of Shaik et al. and Prisznyák, establishes that machine learning algorithms including support vector machines, random forests, and gradient boosting machines excel at identifying patterns invisible to traditional rule-based systems, though performance varies across algorithms and contexts [1][2]. Shaik et al. demonstrate that machine learning approaches identify unanticipated irregularities within complex transactional datasets, yet they operate as black-box systems that resist straightforward interpretation, creating tension with regulatory expectations for explainable compliance decisions [1]. Prisznyák's gap-filling analysis emphasizes that no singular algorithm proves universally optimal, with algorithmic selection determined by underlying theoretical logic, business unit requirements, and integration of information technology infrastructure with visionary management perspectives [2]. These findings collectively suggest that successful implementation of probabilistic compliance systems requires not only technical expertise in machine learning but also deep understanding of the institutional contexts within which these systems operate.

The epistemological reconceptualization examined through the work of Maduranga, Livingstone et al., and the European Banking Authority reveals three fundamental transformations accompanying the probabilistic paradigm [3][4][5]. The logic of anomaly detection shifts from explicit sequential rules and predicate-based filtering toward statistical pattern recognition and behavioral baseline deviation. The evidentiary basis for compliance decisions transforms from auditable rule triggers providing deterministic explanations toward algorithmic outputs requiring post-hoc interpretability techniques such as LIME and SHAP for human comprehension. The institutional trust mechanism recalibrates from confidence in deterministic rule applications toward calibrated skepticism regarding probabilistic model fallibility, requiring new relationships between human investigators and machine learning systems. Livingstone et al. demonstrate that machine learning enables detection of signals previously likely missed, but realizing these benefits requires organizational learning and development of new trust relationships between human experts and algorithmic systems [4].

The systemic vulnerabilities introduced by probabilistic frameworks, examined through Buehler, Ibel, and Stoeckle's analysis of model risk management and Kikuchi's empirical investigation of algorithmic coupling, reveal that these transformations carry implications extending beyond individual institutions to threaten financial system stability [6][7]. Buehler, Ibel, and Stoeckle emphasize that bias can result in unwanted discriminatory and unfair outcomes, with self-learning models prone toward drift over time and requiring adequate validation methods and governance processes to manage these issues [6]. Kikuchi's spatial econometric analysis demonstrates that the United States banking system is becoming algorithmically coupled, with AI adoption by one institution raising productivity at connected institutions through knowledge diffusion, creating synchronization where technical failures in widely-adopted models could trigger correlated shocks across the entire financial network, with large bank spillovers reaching a theta coefficient of 3.13 for return on equity [7].

The regulatory and accountability implications examined through Zhang and Gao's analysis of global AI governance frameworks and the Wolfsberg Group principles demonstrate that principles-based approaches including fairness, ethics, accountability, and transparency provide foundational guidance while allowing flexibility for technological evolution [8][9]. Zhang and Gao document how major financial centers have issued guidance establishing these principles, with the Monetary Authority of Singapore releasing FEAT principles in 2018, the Hong Kong Monetary Authority issuing guidance in 2020, and the People's Bank of China issuing evaluation specifications in 2021 [8]. The Wolfsberg Group principles emphasize legitimate purpose requiring cleansing of historic data to uncover bias, proportionate use mandating collection of only essential data, and transparency in data lineage and signal identification [9].

Several limitations of this literature review should be acknowledged. The rapid evolution of artificial intelligence technologies means that findings from studies published between 2018 and 2022 may not fully capture the most recent developments in machine learning architectures, explainability techniques, or regulatory guidance. The focus on anti-money laundering applications, while providing depth of analysis, may limit generalizability to other compliance domains such as market surveillance, fraud detection, or conduct monitoring. The reliance on published academic and industry research means that proprietary developments within financial institutions may not be reflected in the analysis.

Future research should pursue several promising directions. Longitudinal studies examining model performance across economic cycles would illuminate how probabilistic systems behave under varying conditions and whether degradation patterns emerge during periods of financial stress. Comparative analyses of explainability techniques in regulatory contexts would identify which approaches most effectively balance transparency requirements with detection capabilities while meeting supervisory expectations. Investigations of trust calibration mechanisms in human-AI compliance teams would illuminate how organizations can optimally structure oversight relationships between investigators and algorithmic systems. Cross-industry comparisons of probabilistic anomaly detection adoption would reveal whether lessons from banking generalize to other regulated sectors facing similar epistemological transformations. Finally, system-level simulation studies examining correlated model failures

across algorithmically coupled institutions would inform development of macroprudential frameworks adequate to the probabilistic paradigm.

In conclusion, the transition from rules to probabilities in banking compliance constitutes an epistemological shift whose implications extend far beyond technical performance considerations. The enhanced detection capabilities that probabilistic approaches offer come bundled with novel challenges requiring fundamental rethinking of how financial institutions know what they claim to know, how they justify their determinations to stakeholders, and how they constitute trust in systems that operate through inference rather than certainty. Meeting these challenges will require sustained collaboration among financial institutions, technology developers, regulators, and researchers to develop governance frameworks, regulatory approaches, and accountability structures adequate to the probabilistic paradigm.

REFERENCES

- [1]. M. Shaik, K. Sandhu, L. Gudala, H. Palaparthi, and V. K. Reddy, "From Rules to AI: Assessing Supervised Learning for AML Transaction Monitoring," *Remittances Review*, vol. 6, no. 1, pp. 89–102, Jan. 2021.
- [2]. A. Prisznyák, "Bankrobotics: Artificial Intelligence and Machine Learning Powered Banking Risk Management — Prevention of Money Laundering and Terrorist Financing," *Public Finance Quarterly*, vol. 67, no. 2, pp. 288–303, Jun. 2022, doi: 10.35551/PFQ_2022_2_8.
- [3]. W. A. H. Maduranga, "Micro data model architecture for AML scoring rule engines," M.S. thesis, Univ. Moratuwa, Moratuwa, Sri Lanka, 2022. [Online]. Available: <http://dl.lib.uom.lk/handle/123/21546>
- [4]. J. Richards, "Rules-Based Monitoring, Alert to SAR Ratios, and False Positive Rates – Are We Having The Right Conversations?" *RegTech Consulting, LLC*, Dec. 2018. [Online]. Available: <https://regtechconsulting.net/uncategorized/rules-based-monitoring-alert-to-sar-ratios-and-false-positive-rates-are-we-having-the-right-conversations/>
- [5]. S. Goethals, D. Martens, and T. Evgeniou, "The non-linear nature of the cost of comprehensibility," *Journal of Big Data*, vol. 9, no. 1, pp. 1–23, Dec. 2022, doi: 10.1186/s40537-022-00579-2.
- [6]. "Building an instant fraud detection system using graph neural networks and Neo4j" *wbolt.com*, 2025. [Online]. Available: <https://www.wbolt.com/tw/gnn-fraud-detection-with-neo4j.html>
- [7]. A. Teh, "Adaptive fraud detection: Embedding machine learning for intelligent decisioning," *GBG*, 2023. [Online]. Available: <https://www.gbg.com/au/blog/adaptive-fraud-detection/>
- [8]. A. Livingstone, N. Orakwue, and F. Hiebert, "Analytics and Artificial Intelligence: Deep Learning for Anomaly Detection - A case study from the financial sector with application to process safety," presented at the Mary K O'Connor Process Safety Symp., College Station, TX, USA, 2018. [Online]. Available: <https://oaktrust.library.tamu.edu/handle/1969.1/193453>
- [9]. European Banking Authority, "Report on big data and advanced analytics," *EBA/REP/2020/01*, Jan. 2020. [Online]. Available: https://www.eba.europa.eu/sites/default/documents/files/document_library/Financial%20Innovation%20and%20RegTech%20Discussion%20papers%20and%20reports/2020/Report%20on%20big%20data%20and%20advanced%20analytics/932184/EBA%20report%20on%20Big%20Data%20and%20Advanced%20Analytics.pdf
- [10]. H. Buehler, M. Ibel, and B. Stoeckle, "Financial Risk Management and Explainable, Trustworthy, Responsible AI," *Frontiers in Artificial Intelligence*, vol. 5, art. 779799, Feb. 2022, doi: 10.3389/frai.2022.779799.
- [11]. T. Kikuchi, "The Innovation Tax: Generative AI Adoption, Productivity Paradox, and Systemic Risk in the U.S. Banking Sector," *arXiv preprint arXiv:2602.02607*, Feb. 2025. [Online]. Available: <https://arxiv.org/abs/2602.02607>
- [12]. X. Zhang and Q. Gao, "Risk Types and Regulatory Schemes for Artificial Intelligence Algorithm Financial Applications," *China Banking*, no. 6, pp. 1–8, Jun. 2022. [Online]. Available: <https://www.secrss.com/articles/45889>
- [13]. Wolfsberg Group, "Responsible AI and Machine Learning in Financial Crime Compliance - Implementing the Wolfsberg Group Principles," *Featurespace*, Dec. 2022. [Online]. Available: <https://www.featurespace.com/newsroom/responsible-ai-and-machine-learning-in-financial-crime-compliance-implementing-the-wolfsberg-group-principles>