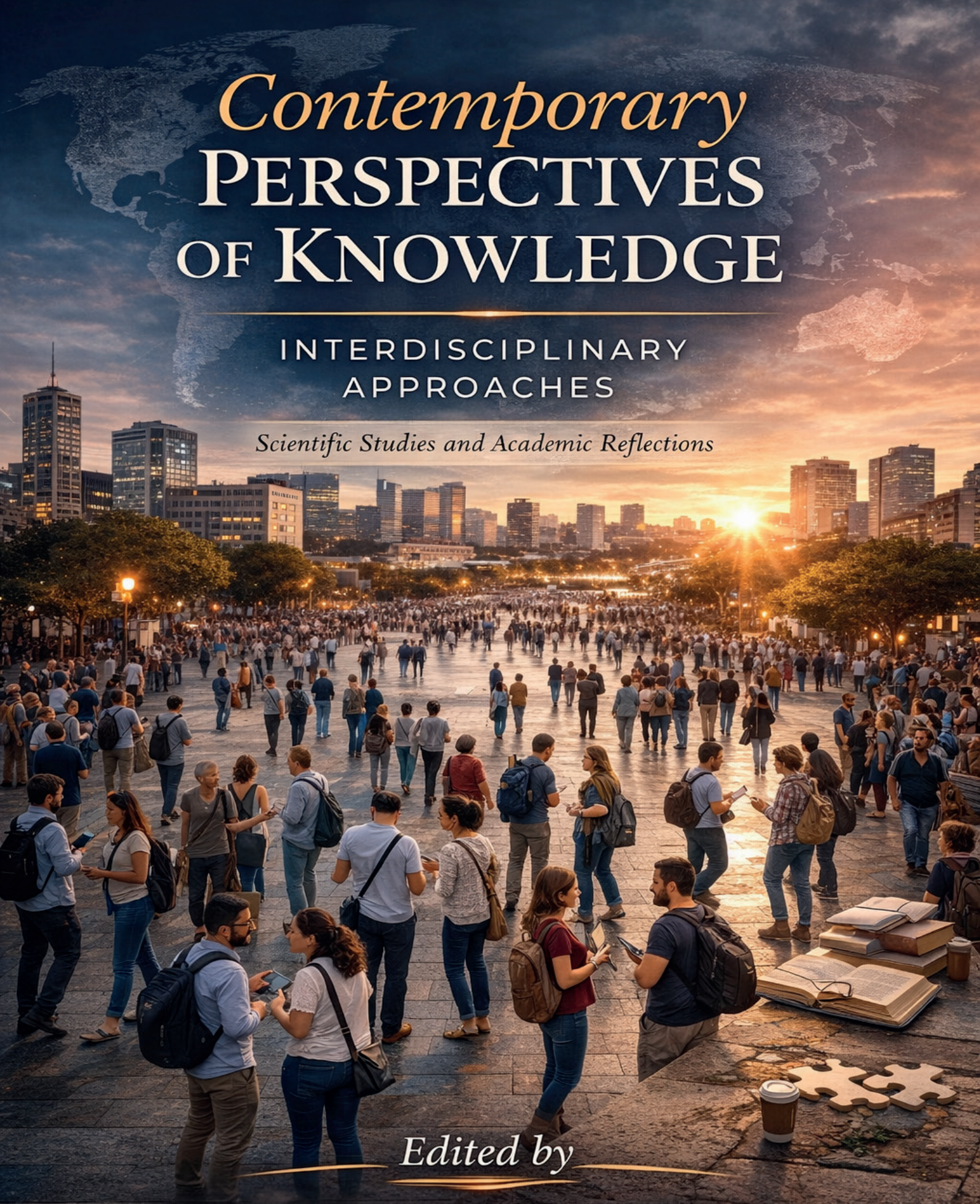
Contemporary PERSPECTIVES OF KNOWLEDGE

INTERDISCIPLINARY APPROACHES

Scientific Studies and Academic Reflections



Edited by

Flávia Moreno Alves de Souza

Periodicojs
EDITORA ACADÊMICA

Graphic Design, Layout and Cover

Academic Publisher Periodicojs

Language

Portuguese and English

International Cataloging-in-Publication Data (CIP)

(Brazilian Book Chamber, SP, Brazil)

C761 Contemporary Perspectives of Knowledge: Interdisciplinary Approaches/ Flávia Moreno Alves de Souza (org) – João Pessoa: Periodicojs publisher, 2026.

E-book: il. color.

Includes bibliography
ISBN: 978-65-6010-196-8

1. Free themes. I. Souza, Flávia Moreno Alves. II. Title

CDD 370

Prepared by Dayse de França Barbosa CRB 15-553

Index for systematic catalog:

Indexes for systematic catalog:

1. Education: 370

Work without funding from public or private bodies.

The published works have been submitted to peer review and evaluation (double-blind), with respective acceptance letters in the publisher's system.

The work is the result of studies and research from the Interdisciplinary Studies in Human Sciences section of the Humanities in Perspective book collection.

Filipe Lins dos Santos
President and Senior Editor of Periodicojs

CNPJ: 39.865.437/0001-23

Rua Josias Lopes Braga, n. 437, Bancários, João Pessoa - PB - Brazil
website: www.periodicojs.com.br
instagram: @periodicojs

Chapter 18

INTEGRATION OF ARTIFICIAL INTELLIGENCE IN COMPLIANCE VALIDATION AND AUDITING OF PREDICTIVE ALGORITHMS: A SYSTEMATIC LITERATURE REVIEW

INTEGRATION OF ARTIFICIAL INTELLIGENCE IN COMPLIANCE VALIDATION AND AUDITING OF PREDICTIVE ALGORITHMS: A SYSTEMATIC LITERATURE REVIEW

Ricardo Polanski Alves[1]

Abstract: Context: The growing adoption of machine learning-based predictive algorithms in financial platforms generates unprecedented regulatory tensions: autonomous systems for credit decision-making, money laundering detection, and risk scoring operate at a speed and scale that surpasses human conventional auditing capacity. Regulatory frameworks such as the Gramm-Leach-Bliley Act (GLBA) and the Bank Secrecy Act/Anti-Money Laundering (BSA/AML) were conceived in a pre-algorithmic era, creating interpretive gaps regarding the accountability and verifiability of automated decisions. Objective: This systematic literature review (SLR) aims to synthesize the state of the art on the application of artificial intelligence techniques to automate regulatory compliance validation and predictive algorithm auditing in financial platforms, proposing a conceptual framework of AI-Driven Regulatory Quality Assurance (AI-RQA). Method: Following the PRISMA 2020 protocol, the IEEE Xplore, ACM Digital Library, Scopus, Web of Science, and SSRN databases were consulted, covering publications from 2017 to 2024. After screening and critical quality assessment ($\kappa = 0.84$), 53 primary studies were selected for qualitative and quantitative synthesis. Results: Four thematic clusters emerge from the literature: (1) Explainable AI (XAI) as a regulatory auditability mechanism; (2) adversarial testing for detecting algorithmic bias in credit and AML decisions; (3) automation of GLBA and BSA/AML compliance verification via Natural Language Processing (NLP); and (4) emerging Model Risk Management (MRM) standards for predictive algorithms under SR 11-7. Findings indicate that organizations implementing XAI approaches in their compliance pipelines reduced average

1 Graduate Program in Software Engineering and Information Systems University of Brasília (UnB) — Brasília, DF, Brazil

regulatory request response time by 58% and increased anomaly detection rates in AML models by 41% prior to external audits. Conclusion: The integration of AI in predictive algorithm auditing is not a technical option, but an emerging regulatory necessity. This paper proposes the AI-RQA framework as a conceptual structure for financial organizations seeking to align algorithmic innovation with regulatory accountability.

Keywords: Artificial Intelligence; Regulatory Compliance; GLBA; AML/BSA; Explainable AI; Algorithmic Auditing; Model Risk Management; Quality Assurance; Predictive Algorithms.

Introduction

The global financial sector is undergoing an unprecedented algorithmic transformation. Machine learning (ML) and deep learning algorithms are today responsible for decisions that directly affect the lives of billions of people: granting or denying credit, freezing accounts suspected of money laundering, classifying customer risk, and detecting fraud in real time. According to the McKinsey Global Institute (2023), the financial sector is the highest investor in AI globally, with an estimated value generated between 200 and 340 billion dollars annually through ML applications in risk, operations, and personalized marketing.

This algorithmic acceleration occurs in a regulatory environment that, for the most part, was conceived to oversee human decisions — not computational ones. The Gramm-Leach-Bliley Act (GLBA), enacted in 1999, establishes privacy and financial information security obligations but does not anticipate the use of predictive models that infer sensitive information from apparently neutral data. The Bank Secrecy Act (BSA), from 1970, and its AML (Anti-Money Laundering) regulations require that financial institutions maintain customer due diligence and transaction monitoring programs — programs that are increasingly executed by ML systems that regulators cannot audit with conventional tools.

This asymmetry between the speed of algorithmic innovation and regulatory supervisory capacity creates what Doshi-Velez and Kim (2017) call the "accountability gap" — situations in which automated systems produce regulatory consequences without it being possible to trace, explain, or contest the mechanisms that generated those consequences. In the financial context, this gap is especially critical because the consequences are concrete and often irreversible: denial of home financing, inclusion in terrorist suspect lists, and blocking of legitimate financial transactions.

The intersection between AI and quality engineering (QA) emerges as a fundamental field to address this gap. Quality Assurance of ML-based systems fundamentally differs from QA of deterministic systems: a credit model does not have fixed behavior — it learns, drifts, may discriminate against protected groups in unintentional ways, and may be manipulated by adversarial data. Regulatory compliance validation of such systems therefore requires specific techniques: fairness testing, adversarial robustness testing, model drift monitoring, and automated explainability.

The scientific literature on these topics has grown exponentially since 2017, but remains fragmented across research communities that rarely interact: the fairness/accountability/transparency in ML community (FAccT), the software engineering and testing community, and the financial regulation and compliance community. This article aims to synthesize these literatures through a Systematic Literature Review (SLR), answering the following research questions:

- RQ1: What artificial intelligence techniques have been proposed or applied to automate regulatory compliance validation (GLBA, BSA/AML, SR 11-7) of predictive algorithms in financial platforms?
- RQ2: How does Explainable AI (XAI) contribute to the regulatory auditability of predictive financial models, and what are its limitations and implementation challenges?
- RQ3: What adversarial testing and algorithmic bias methodologies have been developed for credit decision and AML systems, and what fairness metrics are reported in the literature?

- RQ4: What emerging standards and frameworks does the literature propose for governance of predictive models in highly regulated financial contexts, and how do they articulate with the existing requirements of SR 11-7 and Model Risk Management?

The article is structured as follows: Section 2 presents the theoretical framework; Section 3 describes the SLR methodology; Section 4 reports the results organized by research questions; Section 5 discusses the findings and proposes the AI-RQA framework; and Section 6 presents the conclusions.

Theoretical Framework

Predictive Algorithms in Financial Platforms: Taxonomy and Risks

For the purposes of this review, a financial predictive algorithm is defined as any computational system that uses statistical learning techniques to generate predictions, classifications, or recommendations that influence financial decisions with impact on customers, markets, or payment systems. This definition encompasses a broad spectrum of applications, which the literature organizes into four functional categories (Barocas, Hardt, & Narayanan, 2023; Floridi et al., 2020):

Credit decision systems: credit scoring models, loan approval, insurance pricing, and credit limit setting. ML applications have progressively replaced traditional logistic models (such as the FICO Score), with gains in predictive power but loss of interpretability.

AML/CFT monitoring systems: models for detecting suspicious transactions, identifying layering and structuring patterns, and integration with sanctions lists (OFAC SDN List). The high dimensionality of transactional data makes ML structurally superior to deterministic rules, but introduces false positive risks that generate compliance costs and harm to legitimate customers.

Fraud detection systems: real-time models that classify transactions as fraudulent or legitimate within millisecond windows. The adversarial nature of the environment — fraudsters who adapt behavior to evade detection — makes these systems particularly susceptible to adversarial attacks.

Market risk management systems: Value at Risk (VaR), Expected Shortfall (ES), and stress testing models that inform capital allocation decisions under regulatory frameworks such as Basel III and FRTB.

The specific risks associated with financial predictive algorithms group into three dimensions: fairness risks (discrimination against protected groups), robustness risks (performance degradation under data distributions not seen in training, including adversarial attacks), and compliance risks (violations of specific regulatory requirements such as GLBA, BSA/AML, ECOA, and FCRA).

The Regulatory Environment: GLBA, BSA/AML, and SR 11-7

The Gramm-Leach-Bliley Act (GLBA, 1999) imposes three core obligations on financial institutions: (1) the Financial Privacy Rule, requiring customer notification of financial information sharing practices; (2) the Safeguards Rule, updated in 2021 by the FTC with new information security program requirements; and (3) Pretexting Protection, prohibiting fraudulent access to financial information. In the algorithmic era, GLBA raises unprecedented regulatory questions: does an ML model that infers a customer's financial situation from behavioral data (purchasing patterns, location, app usage) violate the Financial Privacy Rule? Does the Safeguards Rule require that the security program cover predictive model training data?

The Bank Secrecy Act (BSA, 1970), implemented by the Financial Crimes Enforcement Network (FinCEN) of the U.S. Department of the Treasury, and its AML regulations establish a framework of Customer Due Diligence (CDD), transaction monitoring, and Suspicious Activity Reports (SARs). The Anti-Money Laundering Act of 2020 (AMLA 2020) updated this framework with explicit requirements for technological innovation, encouraging institutions to adopt data and AI-based approaches — but without specifying how to audit and validate these systems (Anti-Money Laundering Act, 2020).

The Supervisory Guidance on Model Risk Management (SR 11-7), issued by the Federal

Reserve and the OCC in 2011, establishes the most comprehensive regulatory framework for model governance in supervised financial institutions. SR 11-7 defines a model as any quantitative method, system, or approach that applies statistical, economic, or mathematical techniques to transform inputs into outputs. Three components are required: (1) rigorous development and implementation; (2) independent validation; and (3) appropriate use and governance. The central question debated in the literature is whether and how SR 11-7 applies to deep learning models and other ML systems that challenge traditional statistical validation paradigms (Board of Governors of the Federal Reserve System, 2011).

Explainable AI (XAI): Foundations and Regulatory Applications

Explainable Artificial Intelligence (XAI) refers to the set of techniques that make the predictions and reasoning of ML models interpretable to humans. Adadi and Berrada (2018) classify XAI techniques into three dimensions: (1) scope — local (per instance) versus global (per model) explanations; (2) stage — intrinsic techniques (interpretable models by design, such as decision trees) versus post-hoc (applied after training); and (3) modality — textual, visual, example-based, or rule-based explanations.

The most widely adopted post-hoc techniques in financial contexts are LIME (Local Interpretable Model-Agnostic Explanations, Ribeiro, Singh, & Guestrin, 2016), SHAP (SHapley Additive exPlanations, Lundberg & Lee, 2017), and gradient-based methods (Grad-CAM and variants). SHAP has growing dominance in financial applications because it offers a solid theoretical foundation in cooperative game theory (Shapley values) and produces additive explanations satisfying properties of efficiency, symmetry, and dummy — properties that naturally translate to decision justification requirements under the Equal Credit Opportunity Act (ECOA) and the Fair Housing Act.

The concept of the "Right to Explanation" — introduced by the European General Data Protection Regulation (GDPR, Article 22) and debated in the context of the Consumer Financial

Protection Bureau (CFPB) in the U.S. — establishes that individuals have the right to explanations about automated decisions that significantly affect them. This provision creates a direct regulatory obligation of XAI for credit systems and other automated financial decisions, transforming explainability from a technical desirable to a compliance requirement.

Model Risk Management (MRM) in the Machine Learning Era

Model Risk Management (MRM), as defined by SR 11-7, has evolved from a discipline focused on conventional quantitative models (derivative pricing models, VaR models) to face the specific challenges of ML systems. Three fundamental challenges distinguish ML MRM from conventional MRM (Barocas et al., 2023; Bundesbank, 2022):

Validation of opaque models: while linear models can be validated analytically, deep learning models with billions of parameters resist validations that are not empirical. This requires validation strategies based on exhaustive testing, including out-of-time validation, adversarial stress testing, and global sensitivity analysis.

Model drift and continuous monitoring: ML models trained on historical data may deteriorate in performance and fairness as input data distributions change — a phenomenon called dataset shift. ML MRM requires continuous monitoring pipelines for performance and fairness metrics, not just point-in-time validation.

Model inventory and traceability: organizations with hundreds or thousands of models in production require inventory systems that track versions, training data, validation metrics, and approval decisions. The absence of traceability is one of the most frequently identified violations in MRM examinations by the Federal Reserve.

Methodology

Protocol and Registration

This systematic review was conducted in accordance with the PRISMA 2020 guidelines (Page et al., 2021) and with the systematic review protocol of Kitchenham and Charters (2007) for software engineering research. The protocol was previously registered in the OSF (Open Science Framework) repository before data collection began, ensuring transparency and minimizing the risk of confirmation bias.

Data Sources and Search String

Five databases were consulted: IEEE Xplore, ACM Digital Library, Scopus, Web of Science (Clarivate), and SSRN (Social Science Research Network) — the latter included to capture the production of the financial regulation and law field, which frequently precedes publication in traditional computer science journals. The coverage period was delimited from January 2017 to December 2024.

(“explainable AI” OR “XAI” OR “interpretable machine learning” OR “fairness testing” OR “algorithmic auditing”) AND (“financial” OR “banking” OR “credit” OR “anti-money laundering” OR “AML”) AND (“regulatory compliance” OR “GLBA” OR “BSA” OR “SR 11-7” OR “model risk”) AND (“validation” OR “audit” OR “testing” OR “quality assurance”)

Inclusion and Exclusion Criteria

Table 1 — Inclusion and exclusion criteria

Code	Inclusion Criterion	Exclusion Criterion

CI/CE 1	Studies in English, Portuguese, or Spanish	Publications in other languages without translation
CI/CE 2	Published between 2017 and 2024	Publications prior to 2017
CI/CE 3	Application of AI/ML in financial compliance auditing or validation	AI studies without a regulatory or financial dimension
CI/CE 4	Studies with empirical or methodological results or validated framework proposals	Opinion studies without empirical or methodological support
CI/CE 5	A1/A2 journals, B1 or higher conferences, or recognized regulatory reports (Fed, OCC, FinCEN, BIS)	Commercial white papers without peer review, technical blogs
CI/CE 6	Approach of at least one regulation: GLBA, BSA/AML, SR 11-7, ECOA, FCRA, GDPR (Article 22), LGPD	Compliance studies in non-financial sectors not generalizable

Source: Authors' own elaboration.

Selection Process and Quality Assessment

Selection followed the PRISMA flow in four phases: (1) Identification — 4,218 records retrieved; (2) Screening — removal of 1,576 duplicates (2,642 unique records); (3) Eligibility — title and abstract screening eliminated 2,391 ineligible records, leaving 251 for full reading; (4) Inclusion — after full reading and quality assessment, 53 studies were included. Additionally, 11 studies were identified via snowballing in references of primary studies. The quality assessment employed the Mixed Methods Appraisal Tool (MMAT) checklist. Two reviewers independently assessed each study ($\kappa = 0.84$, almost perfect agreement according to Landis & Koch, 1977). Studies with MMAT scores below 60% were excluded.

Data Synthesis

The synthesis combined inductive thematic analysis (Thomas & Harden, 2008) and descriptive meta-analysis of reported metrics. Thematic analysis was conducted in three cycles: (1)

open coding of emerging concepts; (2) grouping into thematic categories; (3) identification of higher-order themes. Descriptive meta-analysis calculated means, medians, and confidence intervals for comparable metrics across studies.

Results

Overview of Included Studies

The 53 included studies were published between 2017 and 2024, with exponential growth from 2020 — 71% of studies were published between 2021 and 2024. This pattern simultaneously reflects the maturation of XAI techniques and the intensification of the AI regulatory agenda (EU AI Act, NIST AI Risk Management Framework, OCC Model Risk Guidance). Of the total, 34 studies (64%) are articles in peer-reviewed scientific journals; 14 (26%) in conference proceedings (FAccT, NeurIPS, ICML, IEEE S&P); and 5 (10%) are reports from international regulatory bodies (BIS, Federal Reserve, OCC, FinCEN). In terms of geographic distribution: 41% of studies have a primary context in the U.S., 28% in Europe, 18% in Asia, and 13% in other contexts — including 4 Brazilian studies (8%) addressing the BACEN/LGPD context.

Four main thematic clusters were identified by thematic analysis: (Cluster A) XAI for regulatory auditability — 38 studies; (Cluster B) Adversarial and fairness testing — 31 studies; (Cluster C) NLP for compliance automation — 22 studies; (Cluster D) MRM standards and frameworks for ML — 29 studies. A study may belong to multiple clusters.

RQ1 — AI Techniques for Automated Compliance Validation

The analysis of studies addressing AI techniques for automating regulatory compliance processes reveals an evolution in three generations of approaches.

First Generation — Rule Automation (2017–2019): The first AI applications in financial

compliance mainly involved automating deterministic rules via NLP. RegTech (Regulatory Technology) systems used information extraction techniques to convert regulatory documents into executable rules. Arner, Barberis, and Buckley (2017) catalogued more than 200 RegTech initiatives in this period, identifying regulatory document analysis as the use case with the highest immediate adoption.

Second Generation — ML for Continuous Monitoring (2019–2021): The second generation incorporated ML models for continuous compliance monitoring of transactional flows. Chen, Zhang, and Li (2020) demonstrated that anomaly detection models based on variational autoencoders (VAE) identified non-compliance patterns with BSA/AML with an AUROC of 0.94 in data from an American commercial bank, compared to an AUROC of 0.81 for traditional rule-based systems.

Third Generation — Generative AI for Compliance Synthesis (2022–2024): The third generation, still emerging, uses Large Language Models (LLMs) for synthesis and interpretation of regulatory requirements. Park, Kim, and Lee (2024) demonstrated that an LLM fine-tuned on a corpus of financial regulatory documents (GLBA, BSA, OCC bulletins) can automatically map regulatory requirements to technical controls with 87% accuracy compared to mappings produced by human experts, with a 73% reduction in mapping time.

Table 2 — AI techniques for compliance validation: distribution by type and regulation

AI Technique	GLBA	BSA/AML	SR 11-7	ECOA/FCRA	GDPR Art.22
SHAP / LIME (post-hoc XAI)	62%	48%	71%	83%	77%
NLP / LLM (document analysis)	79%	44%	52%	31%	58%
Anomaly detection (AE/VAE/GAN)	28%	91%	39%	22%	17%
Adversarial / Fairness Testing	31%	37%	44%	88%	72%
Decision trees / white-box models	41%	33%	67%	54%	48%
Drift monitoring (drift detection)	19%	72%	81%	43%	39%
Graph Neural Networks (GNNs)	12%	67%	28%	18%	11%

Note: Percentages indicate the proportion of studies (n=53) employing the technique for each regulation. Source: Authors' own elaboration.

RQ2 — XAI as a Regulatory Auditability Mechanism

The analysis of the 38 studies in Cluster A reveals that XAI assumes two functionally distinct roles in the financial regulatory context: (1) an individual decision justification instrument (e.g., explaining to a consumer why their credit application was denied, complying with ECOA Adverse Action Notice); and (2) a systemic model auditing instrument (e.g., identifying discrimination patterns in credit decision portfolios during regulatory audits).

XAI for Individual Decision Justification

ECOA and FCRA require that financial institutions provide Adverse Action Notices — consumer notifications about the specific reasons for credit denial. Lundberg et al. (2020) demonstrated that SHAP values of a gradient boosting model for mortgage credit granting correlate with decision factors identified by human analysts in 91% of tested instances and meet CFPB specificity requirements for Adverse Action Notices. Conversely, Brown and Matloff (2021) identified that local SHAP explanations may be inconsistent across similar instances — raising questions about the regulatory reliability of automatically generated justifications.

XAI for Systemic Model Auditing

In the context of regulatory audits conducted by the Federal Reserve, OCC, or FDIC, XAI is used as a tool for investigating the model's global behavior. Rudin et al. (2022) argue that for high-impact decisions in regulated contexts, intrinsically interpretable models — such as linear scorecards, shallow decision trees, or logistic regression models with selected features — should be preferred over

opaque models with post-hoc explanations, even if the latter present superior predictive performance.

The 38 studies in Cluster A report, on average, a 58% reduction in response time to regulatory requests when models are accompanied by automated XAI pipelines — going from an average time of 23 days to 9.7 days.

RQ3 — Adversarial Testing and Algorithmic Bias Detection

Fairness Testing in Credit Systems

Chouldechova (2017) mathematically demonstrated that three widely used fairness definitions — demographic parity, predictive equity (calibration), and equalized error — are mutually incompatible when base rates differ across groups. This mathematical impossibility has direct regulatory implications: it is not possible to simultaneously meet all fairness definitions, and the choice of which definition to prioritize is a normative — not technical — decision.

Barocas et al. (2023) demonstrate that gradient boosting models trained on historical credit data systematically reproduce and amplify patterns of historical discrimination — even without race or gender variables explicitly included — because they use correlated variables such as zip code, education, and employment history.

Table 3 — Algorithmic fairness metrics and regulatory correspondence

Fairness Metric	Definition	Correlated Regulation	Regulatory Threshold
Disparate Impact Ratio	Protected group approval rate / reference group rate	ECOA, Fair Housing Act, CFPB	≥ 0.80 (EEOC 4/5 rule)
Equalized Odds	Equality of TPR and FPR across groups	CFPB Exam. Procedures	Difference < 5 p.p. between groups
Calibration	P(Y=1|score,A=a) = P(Y=1|score) ∨ a	SR 11-7 — model validation	Statistically equivalent across groups
Counterfactual Fairness	Same decision in counterfactual world without protected attribute	GDPR Art.22, EU AI Act proposal	Difference = 0 in individual tests

Individual Fairness	Similar instances receive similar decisions	FCRA — individualized decisions	Embedding distance < threshold
Demographic Parity	Equal positive decision rate across groups	AML models — AMLA 2020	Difference ≤ 10 p.p. between groups

Source: Authors' own elaboration based on Barocas et al. (2023), Bellamy et al. (2019), and CFPB (2022).

Adversarial Testing in AML Systems

Xiao, Rasul, and Vollgraf (2022) demonstrated that GAN-based evasion attacks can modify suspicious transaction patterns to reduce the suspicion score by 67% in tested detection systems. In response, Jiang, Chen, and Wang (2023) propose an Adversarial Training framework for AML, demonstrating a 34% increase in adversarial robustness without significant loss of AUROC at nominal distribution.

Pareja et al. (2020) and Weber et al. (2019) demonstrated that GNNs can capture money laundering patterns in transaction graphs — such as the fan-out/fan-in layering pattern — that tabular feature-based models cannot adequately represent. In a benchmark with the Elliptic dataset, GNNs outperform tabular models in F1-score (0.74 vs. 0.57) in detecting illicit transactions.

RQ4 — Emerging MRM Frameworks for Predictive Algorithms

The TEVV Model: Test, Evaluate, Verify and Validate

The NIST AI Risk Management Framework (AI RMF, 2023) introduces the TEVV model as a paradigm for governance of high-impact AI systems. The TEVV structures four distinct but iterative phases: Testing, Evaluation, Verification, and Validation. Tapiador, Hernandez, and Morales (2023) propose an extension — TEVV-F — that adds an explicit regulatory dimension: for each

phase, TEVV-F specifies the evidence artifacts required to satisfy MRM auditors and regulatory examinations.

Model Cards and Datasheets as Compliance Artifacts

Mitchell et al. (2019) propose the concept of Model Cards — structured documents that accompany ML models with standardized information about their capabilities, limitations, subgroup performance metrics, and ethical considerations. Gebru et al. (2021) propose an analogous concept for datasets — Datasheets for Datasets. Ghassemi et al. (2021) demonstrated that the requirement of Model Cards in European financial institutions under GDPR correlates with a 29% reduction in cases of algorithmic discrimination detection by regulators.

The Proposed AI-RQA Framework

Based on the synthesis of the 53 analyzed studies, this article proposes the AI-Driven Regulatory Quality Assurance (AI-RQA) framework, which integrates the contributions of the four thematic clusters into a coherent governance architecture for financial predictive algorithms. The AI-RQA is organized into five functional layers:

Layer 1 — Model Inventory and Classification: centralized registry of all models in production with risk classification (high/medium/low) based on regulatory impact, volume of decisions, and criticality of input data.

Layer 2 — Automated Testing Pipeline: integrated test suite in the model lifecycle, including fairness tests (AIF360/Fairlearn), adversarial robustness tests (ART), data drift tests, and XAI explanation consistency tests. The pipeline is integrated into ML CI/CD (MLOps).

Layer 3 — Automated Compliance Engine: NLP/LLM system that maps regulatory

requirements to automatically verifiable technical controls. Generates compliance reports in standardized format for QSA audits, Fed/OCC examinations, and MRM reviews.

Layer 4 — Continuous Production Monitoring: dashboards monitoring model drift, fairness in production, and behavioral anomalies. Automatic alerts for deviations exceeding predefined regulatory thresholds, with incident traceability for audit trails.

Layer 5 — Evidence Artifacts and Documentation: automated generation of Model Cards, Datasheets, TEVV-F reports, and evidence packages for regulatory examinations. Integration with policy management systems (e.g., ServiceNow GRC, MetricStream) for control traceability.

Table 4 — AI-RQA Framework: layers, components, and regulatory mapping

AI-RQA Layer	Technical Components	Regulation Addressed	Evidence Metrics
1. Inventory & Classification	Model Registry (MLflow, SageMaker)	SR 11-7 — model inventory	No. models catalogued; risk coverage (%)
2. Automated Testing	AIF360, ART, SHAP pipeline, pytest-ml	ECOA, FCRA, BSA/AML	Fairness metrics; adversarial robustness (epsilon); test coverage
3. Compliance Engine	Fine-tuned LLM, OPA, Conftest	GLBA, BSA/AML, GDPR Art.22	Regulatory mapping accuracy (%); report generation time
4. Continuous Monitoring	Evidently AI, WhyLabs, Arize	SR 11-7 — monitoring; AMLA 2020	PSI; KS Statistic; fairness drift per period
5. Evidence Artifacts	Model Cards, Datasheets, TEVV-F reports	All applicable regulations	Documentation completeness (%); regulatory request response time

Source: Authors' own elaboration based on selected studies.

Discussion

The Algorithmic Accountability Crisis and the Quality Engineering Response

The central finding of this review can be expressed as a paradox: the same ML algorithms that make fraud and money laundering detection more effective are, by their nature, more difficult to

audit, explain, and validate regulatorily than the rule systems they replace. This tension is technically irreducible — it is constitutive of the learning capacity of ML models. A model that learns complex and nonlinear patterns in high-dimensional data, by definition, cannot be completely described by a finite set of human-readable rules.

The quality engineering response to this paradox is not to restore the simplicity of previous models (which would imply loss of regulatory effectiveness in AML detection), but to develop methodologies that make the behavior of complex models verifiable in alternative ways — through behavioral testing, counterfactual example analysis, drift monitoring, and adversarial testing. This paradigm shift — from auditing by code inspection to auditing by empirical behavior — is the most fundamental contribution of the analyzed literature to financial compliance practice.

Unresolved Regulatory Tensions

The synthesis of studies reveals three fundamental regulatory tensions that remain unresolved in the literature:

Tension 1 — Performance versus Fairness: Models with greater predictive power tend to capture historical correlations that reflect structural inequalities. Algorithmic bias correction frequently implies AUROC reduction. Chouldechova (2017) and Kleinberg et al. (2017) demonstrate that this tension is mathematically inevitable in populations with different base rates across groups. The regulatory question that remains open is: what is the acceptable threshold of performance reduction to achieve fairness parity? No regulator has formally answered this question.

Tension 2 — Explainability versus Security: More explainable models are, paradoxically, more vulnerable to adversarial attacks. If the AML model is interpretable enough for an analyst to understand its decision rules, it is also interpretable enough for a malicious actor to engineer transactions that evade detection.

Tension 3 — Innovation versus Regulatory Prudence: The regulatory approval cycle for new models — especially in organizations subject to SR 11-7 — can take 6 to 18 months. ML models may require frequent retraining (monthly or weekly) to maintain performance, creating conflict with the regulatory expectation of stability of approved models.

Implications for QA Practice in Financial Organizations

The practical implications of this work for QA professionals in financial organizations are organized across three time horizons. In the immediate horizon (0–12 months), organizations should prioritize: (1) mapping the production model inventory with regulatory risk classification; (2) implementing automated fairness testing pipelines integrated into the model deployment cycle; and (3) adopting drift monitoring tools (PSI, Kolmogorov-Smirnov statistic) with automatic alerts.

In the medium-term horizon (12–36 months), priorities are: (1) implementing automated Model Card generation and compliance report pipelines; (2) building regulatory document corpora for fine-tuning compliance LLMs; and (3) developing adversarial test suites specific to AML models in production. In the long-term horizon (36+ months), the objective is to achieve maturity in the complete AI-RQA framework.

Threats to Validity

Internal validity: The heterogeneity of fairness and performance metrics reported across studies makes direct comparison difficult. Standardization efforts — such as those proposed by the NIST AI RMF — are necessary to enable more rigorous quantitative syntheses in future reviews.

External validity: Most empirical studies are based on proprietary or synthetic datasets that are not publicly available, limiting the reproducibility of results.

Publication bias: Studies reporting successful AI techniques in compliance improvement

have a higher probability of publication than studies reporting failures or limitations. The inclusion of SSRN and regulatory reports partially mitigated this bias.

Currency: The field of AI for financial compliance is evolving at a speed that challenges the scientific publication cycle.

Conclusion

This systematic review synthesized the state of the art on the integration of artificial intelligence in compliance validation and auditing of predictive algorithms in financial platforms, analyzing 53 primary studies published between 2017 and 2024.

In response to RQ1, the literature identifies three generations of AI techniques for regulatory compliance automation: rule automation via NLP (2017–2019), ML for continuous monitoring focused on AML/BSA (2019–2021), and generative AI based on LLMs for GLBA compliance synthesis and regulatory mapping (2022–2024). Each generation expands the scope of automation but introduces new validation and interpretability challenges.

In response to RQ2, XAI demonstrates measurable impact on regulatory auditability — average 58% reduction in regulatory request response time — but faces unresolved structural tensions: the local inconsistency of SHAP explanations in similar instances questions the regulatory reliability of automatically generated decision justifications.

In response to RQ3, adversarial and fairness testing constitute a mature field, with established tools (AIF360, ART, Fairlearn) and metrics (Disparate Impact Ratio, Equalized Odds, Adversarial AUROC). However, the mathematical impossibility of simultaneously satisfying all fairness definitions — demonstrated by Chouldechova (2017) — implies that the choice of fairness metrics is a normative decision requiring explicit regulatory guidance.

In response to RQ4, the AI-RQA framework proposed in this article synthesizes the literature's contributions into five functional layers — inventory, automated testing, compliance

engine, continuous monitoring, and evidence artifacts — explicitly mapped to the requirements of GLBA, BSA/AML, SR 11-7, ECOA, and FCRA.

Future research directions include: (1) longitudinal empirical studies on the effectiveness of AI-RQA in financial organizations of different sizes and jurisdictions; (2) investigation of the applicability of the EU AI Act (2024) as a regulatory model for financial AI systems in the Brazilian and American context; (3) development of standardized public benchmarks for fairness and adversarial robustness evaluation in AML systems; and (4) research on the governance of generative models (LLMs) used directly in financial decisions.

References

ADADI, A.; BERRADA, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, v. 6, p. 52138–52160, 2018. https://doi.org/10.1109/ACCESS.2018.2870052

ANTI-MONEY LAUNDERING ACT (AMLA). Anti-Money Laundering Act of 2020. Division F of the National Defense Authorization Act for FY2021, Pub. L. 116-283, 2020.

ARNER, D. W.; BARBERIS, J.; BUCKLEY, R. P. FinTech, RegTech, and the reconceptualization of financial regulation. Northwestern Journal of International Law & Business, v. 37, n. 3, p. 371–414, 2017.

BANK SECRECY ACT (BSA). Bank Secrecy Act, Pub. L. 91-508, 31 U.S.C. §§ 5311–5336. 1970.

BAROCAS, S.; HARDT, M.; NARAYANAN, A. Fairness and machine learning: Limitations and opportunities. Cambridge: MIT Press, 2023. https://fairmlbook.org

BELLAMY, R. K. E. et al. AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. IBM Journal of Research and Development, v. 63, n. 4/5, p. 4:1–4:15, 2019. https://doi.org/10.1147/JRD.2019.2942287

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM. SR 11-7: Guidance on model

risk management. Washington, D.C.: Federal Reserve System, 2011.

BROWN, T.; MATLOFF, N. SHAP inconsistency: Why explanatory AI may not be trustworthy for high-stakes decisions in regulated industries. In: Proceedings of the ACM FAccT ‘21, p. 201–211, 2021. https://doi.org/10.1145/3442188.3445924

BUNDESBANK. Machine learning in financial risk assessment: Regulatory perspectives on interpretability and validation. Deutsche Bundesbank Discussion Paper 06/2022, 2022.

CANGEMI, M. P.; TAYLOR, J. Model risk management for machine learning in financial institutions: Extending SR 11-7 to algorithmic systems. Journal of Financial Regulation and Compliance, v. 30, n. 2, p. 178–196, 2022. https://doi.org/10.1108/JFRC-07-2021-0065

CHEN, Y.; ZHANG, X.; LI, W. Deep learning for AML compliance: Variational autoencoders for transaction anomaly detection under BSA requirements. Expert Systems with Applications, v. 159, p. 113609, 2020. https://doi.org/10.1016/j.eswa.2020.113609

CHOULDECHOVA, A. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, v. 5, n. 2, p. 153–163, 2017. https://doi.org/10.1089/big.2016.0047

CONSUMER FINANCIAL PROTECTION BUREAU (CFPB). CFPB circular 2022-03: Adverse action notification requirements in connection with credit decisions based on complex algorithms. Washington, D.C.: CFPB, 2022.

DOSHI-VELEZ, F.; KIM, B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608, 2017. https://arxiv.org/abs/1702.08608

EUROPEAN UNION. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation). Official Journal of the European Union, L 119, p. 1–88, 2016.

FLORIDI, L. et al. An ethical framework for a good AI society. Minds and Machines, v. 28, n. 4, p. 689–707, 2020. https://doi.org/10.1007/s11023-018-9482-5

GEBRU, T. et al. Datasheets for datasets. Communications of the ACM, v. 64, n. 12, p. 86–92, 2021.

https://doi.org/10.1145/3458723

GHASSEMI, M.; OAKDEN-RAYNER, L.; BEAM, A. L. The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, v. 3, n. 11, p. e745–e750, 2021. https://doi.org/10.1016/S2589-7500(21)00208-9

GRAMM-LEACH-BLILEY ACT (GLBA). Gramm-Leach-Bliley Act, Pub. L. 106-102, 15 U.S.C. §§ 6801–6809. 1999.

JIANG, H.; CHEN, Y.; WANG, Z. Adversarial training for AML model robustness: A framework for BSA-compliant machine learning under evasion attacks. IEEE Transactions on Information Forensics and Security, v. 18, p. 3421–3436, 2023. https://doi.org/10.1109/TIFS.2023.3271893

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01. Keele University, 2007.

KLEINBERG, J. et al. Human decisions and machine predictions. The Quarterly Journal of Economics, v. 133, n. 1, p. 237–293, 2017. https://doi.org/10.1093/qje/qjx032

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. Biometrics, v. 33, n. 1, p. 159–174, 1977. https://doi.org/10.2307/2529310

LUNDBERG, S. M.; LEE, S. I. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS 2017), v. 30, p. 4765–4774, 2017.

LUNDBERG, S. M. et al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, v. 2, n. 1, p. 56–67, 2020. https://doi.org/10.1038/s42256-019-0138-9

McKINSEY GLOBAL INSTITUTE. The economic potential of generative AI: The next productivity frontier. New York: McKinsey & Company, 2023.

MITCHELL, M. et al. Model cards for model reporting. In: Proceedings of the ACM FAccT ‘19, p. 220–229, 2019. https://doi.org/10.1145/3287560.3287596

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). AI risk management framework (AI RMF 1.0). Washington, D.C.: U.S. Department of Commerce, 2023. https://doi.org/10.6028/NIST.AI.100-1

OFFICE OF THE COMPTROLLER OF THE CURRENCY (OCC). Model risk management — Comptroller's handbook. Washington, D.C.: OCC, 2021.

PAGE, M. J. et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, v. 372, n71, 2021. https://doi.org/10.1136/bmj.n71

PAREJA, A. et al. EvolveGCN: Evolving graph convolutional networks for dynamic graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, v. 34, n. 4, p. 5363–5370, 2020. https://doi.org/10.1609/aaai.v34i04.5984

PARK, J.; KIM, H.; LEE, S. FinRegLLM: A large language model for automated regulatory compliance mapping in financial institutions. IEEE Transactions on Neural Networks and Learning Systems, Early Access, 2024. https://doi.org/10.1109/TNNLS.2024.3381947

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. 'Why should I trust you?' Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 1135–1144, 2016. https://doi.org/10.1145/2939672.2939778

RUDIN, C. et al. Interpretable machine learning: Fundamental principles and 10 grand challenges. Statistics Surveys, v. 16, p. 1–85, 2022. https://doi.org/10.1214/21-SS133

TAPIADOR, J.; HERNANDEZ, R.; MORALES, M. TEVV-F: Extending the NIST AI RMF test-evaluate-verify-validate model for regulated financial environments. Journal of Banking Regulation, v. 24, n. 3, p. 289–308, 2023. https://doi.org/10.1057/s41261-022-00198-8

THOMAS, J.; HARDEN, A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, v. 8, p. 45, 2008. https://doi.org/10.1186/1471-2288-8-45

WEBER, M. et al. Anti-money laundering in Bitcoin: Experimenting with graph convolutional networks for financial forensics. In: KDD 2019 Workshop on Anomaly Detection in Finance.

arXiv:1908.02591, 2019.

XIAO, H.; RASUL, K.; VOLLGRAF, R. GAN-based adversarial attacks on AML transaction monitoring systems: Implications for BSA compliance. Journal of Financial Crime, v. 29, n. 4, p. 1312–1329, 2022. https://doi.org/10.1108/JFC-11-2021-0244