Volume 4 / Issue 8/ August 2026 International Journal of Modern Science and Research Technology
ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

Analyzing the Impact of Artificial Intelligence on Cybersecurity Threat Detection and Response in U.S. Financial Institutions

Itanyi Akoh Isaiah
Department of Computer Science, Kogi State College of Health Sciences and Technology, Idah, Kogi State

Chinemelum Umealajekwu
Department of Cybersecurity and Business Analytics, University of New Mexico, USA

Abstract
The rapid integration of artificial intelligence (AI) into U.S. financial institutions has fundamentally transformed cybersecurity threat detection and response capabilities, while simultaneously introducing a new class of sophisticated risks. This study conducts a systematic literature review to analyze the impact of AI-driven cybersecurity tools on threat detection efficacy, incident response time, and overall cyber resilience within the U.S. financial sector. Drawing on peer-reviewed research, government reports, and institutional case studies published between 2019 and 2025, the study identifies key themes including the deployment of machine learning for anomaly detection, the emergence of adversarial AI threats, regulatory fragmentation, and the widening capability gap between large and small financial institutions. The findings reveal that while AI substantially improves detection accuracy and operational efficiency, its adoption introduces risks such as adversarial attacks, model drift, data poisoning, and algorithmic bias that existing frameworks are inadequate to address comprehensively. This paper proposes a scalable, unified AI-cybersecurity framework tailored for U.S. financial institutions, emphasizing governance, adaptive risk assessment, cross-institutional collaboration, and bias-aware AI design. The proposed framework is intended to serve as a foundation for regulators, financial institutions, and technology providers seeking to harness AI's capabilities while maintaining systemic stability and regulatory compliance.

Keywords: artificial intelligence, cybersecurity, threat detection, incident response, financial institutions, machine learning, risk management, adversarial AI, deep learning, regulatory compliance.

Introduction

The U.S. financial services sector stands at the intersection of rapid technological transformation and escalating cyber threats. Over the past decade, financial institutions have progressively adopted artificial intelligence (AI) and machine learning (ML) technologies to enhance operational efficiency, improve customer experience, and bolster cybersecurity defenses. This shift has been accelerated by the proliferation of digital banking platforms, cloud-native infrastructures, and the explosion of data generated by billions of daily financial transactions (Singhal & Arora, 2022). Today, AI-driven systems

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

underpin critical functions including fraud detection, anti-money laundering (AML) compliance, credit risk assessment, and real-time threat monitoring across the sector (Mahalakshmi et al., 2021).

However, the very technologies that strengthen financial institutions also expand their attack surfaces. Threat actors have demonstrated an increasing capacity to exploit AI-specific vulnerabilities, including adversarial machine learning attacks designed to evade detection models, generative AI-powered deepfakes used in business email compromise and identity fraud, and data poisoning schemes that corrupt training datasets to manipulate AI outputs (Korycki & Krawczyk, 2020). The emergence of large language models (LLMs) as operational tools within financial services has further compounded risk exposure, with prompt injection and model hallucination presenting new categories of threat that traditional security frameworks are ill-equipped to address (Remolina, 2022).

The scale of financial losses attributable to AI-enabled cybercrime underscores the urgency of this challenge. In 2024, the U.S. Department of the Treasury published a comprehensive assessment of AI-specific cybersecurity risks in the financial sector, finding that most institutions, regardless of size, rely heavily on AI-powered security solutions but lack standardized governance frameworks to manage associated risks (U.S. Department of the Treasury, 2024). A widely cited incident involving a Hong Kong-based firm losing $25 million to deepfake-enabled fraud demonstrated the real-world financial stakes of inadequate AI risk management (Lalchand et al., 2024). In Indonesia, over 1,000 deepfake incidents targeting financial institutions were reported within a single year, highlighting the global trajectory of this threat vector (Winder, 2024).

Despite these risks, a unified AI cybersecurity framework tailored to the U.S. financial sector remains absent. Existing frameworks including the NIST AI Risk Management Framework (NIST AI RMF), the EU AI Act, and Financial Industry Regulatory Authority (FINRA) guidance provide valuable but fragmented guidance that institutions are left to interpret and implement independently (Raimondo et al., 2023). This fragmentation creates inconsistent security postures across the sector, particularly disadvantaging community banks, credit unions, and fintech startups that lack the resources to develop proprietary AI risk management systems akin to those deployed by JPMorgan Chase or Mastercard (Kumari et al., 2022).

This study addresses this gap by conducting a systematic literature review of AI-driven cybersecurity in U.S. financial institutions from 2019 to 2025. The research synthesizes empirical findings, institutional case studies, and regulatory analyses to evaluate the impact of AI on threat detection and incident response, and to propose a scalable, unified framework for AI cybersecurity risk management. The study is guided by four research questions examining framework alignment, existing strategies, regulatory challenges, and implications for trust and economic stability.

Research Questions

1. How effectively do current AI-driven cybersecurity tools detect and respond to threats in U.S. financial institutions compared to traditional rule-based systems?
2. What AI-specific cybersecurity risks and vulnerabilities have emerged in the U.S. financial sector, and how are institutions currently addressing them?
3. What regulatory and compliance challenges do U.S. financial institutions face in deploying AI-driven cybersecurity solutions, and how can these be mitigated?
4. How can a unified AI cybersecurity framework improve systemic resilience, institutional trust, and equitable security capabilities across financial institutions of varying sizes?

Literature Review

AI in Financial Services: Evolution and Current State

The deployment of AI in financial services has evolved through distinct phases, each marked by increasing sophistication and integration depth. Early applications from the 1990s through the 2000s were primarily rule-based systems designed to flag anomalous transactions against predefined thresholds. The advent of supervised machine learning in the 2010s enabled probabilistic fraud detection models trained on historical transaction data, significantly improving detection rates while reducing manual review

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

burdens (Ee et al., 2024). The current era is characterized by deep learning architectures, natural language processing, and increasingly, generative AI, all of which offer transformative capabilities but introduce commensurate risks (Pandey et al., 2022).

Machine learning models now permeate financial operations at every level. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) power real-time fraud scoring engines; transformer-based models underpin AML transaction monitoring; and reinforcement learning algorithms optimize high-frequency trading strategies (Mbah & Nkechi, 2024). The scale of deployment is significant: major U.S. banks collectively process billions of AI-scored transactions daily, and AI-driven compliance tools handle regulatory reporting tasks that previously required thousands of human analysts (Nwafor et al., 2024). However, this pervasive integration creates systemic dependencies. The failure or compromise of a widely-used AI model particularly one embedded across multiple institutions could trigger cascading security failures with systemic implications (Danielsson et al., 2020).

AI-Specific Cybersecurity Threats in Financial Contexts

The security vulnerabilities unique to AI systems represent a qualitatively different threat landscape from traditional cybersecurity challenges. Korycki and Krawczyk (2020) identified adversarial attacks, data poisoning, and model drift as three primary categories of AI-specific risk. Adversarial attacks involve the deliberate crafting of inputs whether transactions, images, or text specifically designed to mislead trained models while appearing legitimate to human observers. In financial contexts, these attacks have been demonstrated to bypass fraud detection models with high success rates, enabling fraudulent transactions to be processed as benign (Uzoka et al., 2024).

Data poisoning represents a particularly insidious threat because it targets the training process itself. By introducing manipulated data into training datasets whether through compromised data pipelines, insider threats, or supply chain attacks adversaries can instill systematic biases or blind spots into deployed models (Familoni, 2024). The consequences in financial applications include models that consistently misclassify specific fraud patterns, overlook money laundering indicators associated with particular transaction types, or generate discriminatory credit decisions that expose institutions to regulatory liability.

Model drift the gradual degradation of model performance as real-world data distributions diverge from training data is an operational risk that compounds security vulnerabilities. Mbah and Nkechi (2024) documented instances where financial fraud detection models experienced significant performance degradation within months of deployment without continuous monitoring and retraining. This is particularly acute in adversarial contexts where threat actors actively adapt their techniques to evade detection, accelerating drift through deliberate strategy evolution.

Generative AI has introduced an entirely new category of threat. The ability to produce photorealistic synthetic media at scale has enabled deepfake attacks that exploit the multi-factor authentication systems deployed by financial institutions (Winder, 2024). Business email compromise attacks augmented by AI-generated voice cloning and video synthesis have resulted in substantial losses; Lalchand et al. (2024) documented the $25 million Hong Kong incident as emblematic of this emerging threat vector's potential impact on financial operations.

Table 1. AI-Specific Cybersecurity Threat Taxonomy in U.S. Financial Institutions

Threat Type	Description	Financial Sector Impact	Mitigation Approach
Adversarial Attacks	Model manipulation via crafted inputs	Fraud evasion, credit-score distortion	Adversarial training, input validation
Data Poisoning	Corrupt training data to skew outputs	Biased lending, tampered risk models	Data provenance tracking, anomaly detection

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

Threat Type	Description	Financial Sector Impact	Mitigation Approach
Model Drift	Gradual degradation of model accuracy	False negatives in fraud detection	Continuous monitoring, periodic retraining
Deepfake Fraud	Synthetic media impersonation	Identity theft, unauthorized transfers	Biometric liveness checks, behavioral analytics
AI Hallucination	Fabricated confident outputs	Incorrect compliance guidance, false alerts	Output verification layers, human-in-the-loop
Prompt Injection	Malicious inputs to LLM-powered services	Unauthorized data exposure	Prompt sanitization, output filtering
Supply Chain AI Risk	Third-party AI model vulnerabilities	Systemic compromise of shared platforms	Vendor audits, model provenance checks

Note. Adapted from Korycki & Krawczyk (2020); Remolina (2022); U.S. Department of the Treasury (2024).

Current Risk Management Frameworks

The landscape of AI risk management frameworks applicable to U.S. financial institutions is characterized by plurality and fragmentation. The NIST AI Risk Management Framework (AI RMF 1.0), published in January 2023, represents the most comprehensive voluntary standard available to U.S. institutions. Organized around four functions Govern, Map, Measure, and Manage the NIST AI RMF provides a structured approach to identifying and mitigating AI risks across the model lifecycle (Raimondo et al., 2023). However, as Ee et al. (2024) noted, its voluntary nature and the absence of enforcement mechanisms limit adoption consistency, particularly among smaller institutions that lack dedicated AI governance resources.

The European Union's AI Act, which entered into force in 2024, classifies AI systems by risk level and imposes binding compliance requirements on high-risk applications including credit scoring and identity verification categories directly relevant to financial services (Habbal et al., 2024). While the Act applies primarily to EU-based institutions and AI providers, its extraterritorial reach analogous to the GDPR's impact on global data practices means U.S. financial institutions with European operations or customers must account for its provisions. This creates a de facto international standard that may ultimately influence U.S. regulatory approaches.

FINRA Regulatory Notice 24-09, issued in June 2024, represents a targeted regulatory response to the adoption of generative AI and LLMs within financial services. The notice mandates that member firms adopting GenAI develop policies addressing technology governance, model risk management, data protection, and output accuracy verification (FINRA, 2024). While this guidance reflects growing regulatory awareness of AI-specific risks, its scope is limited to broker-dealer activities and does not constitute a comprehensive AI cybersecurity standard.

Table 2. Comparative Analysis of AI Risk Management Frameworks Applicable to U.S. Financial Institutions

Framework	Jurisdiction	Key Components	Strengths	Limitations
NIST AI RMF	U.S.	Govern, Map, Measure, Manage	Comprehensive; widely adopted	No enforcement; steep learning

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

Framework	Jurisdiction	Key Components	Strengths	Limitations
				curve
EU AI Act	EU	Risk classification; compliance mandates	Legally binding; risk-tiered	Limited U.S. applicability
ISO/IEC 42001	Global	AI management system standards	Internationally recognized	High compliance cost
AI TRiSM	Global	Trust, risk, security management	Holistic trust model	Nascent adoption; limited tooling
FFIEC Guidance	U.S.	IT risk examination procedures	Bank-specific; regulator-endorsed	Lacks AI-specific provisions
OECD AI Principles	Global	Ethics, transparency, accountability	Strong ethical foundation	Non-binding; aspirational only

Note. Sources include Raimondo et al. (2023); Habbal et al. (2024); FINRA (2024); U.S. Department of the Treasury (2024).

AI Threat Detection: Capabilities and Performance

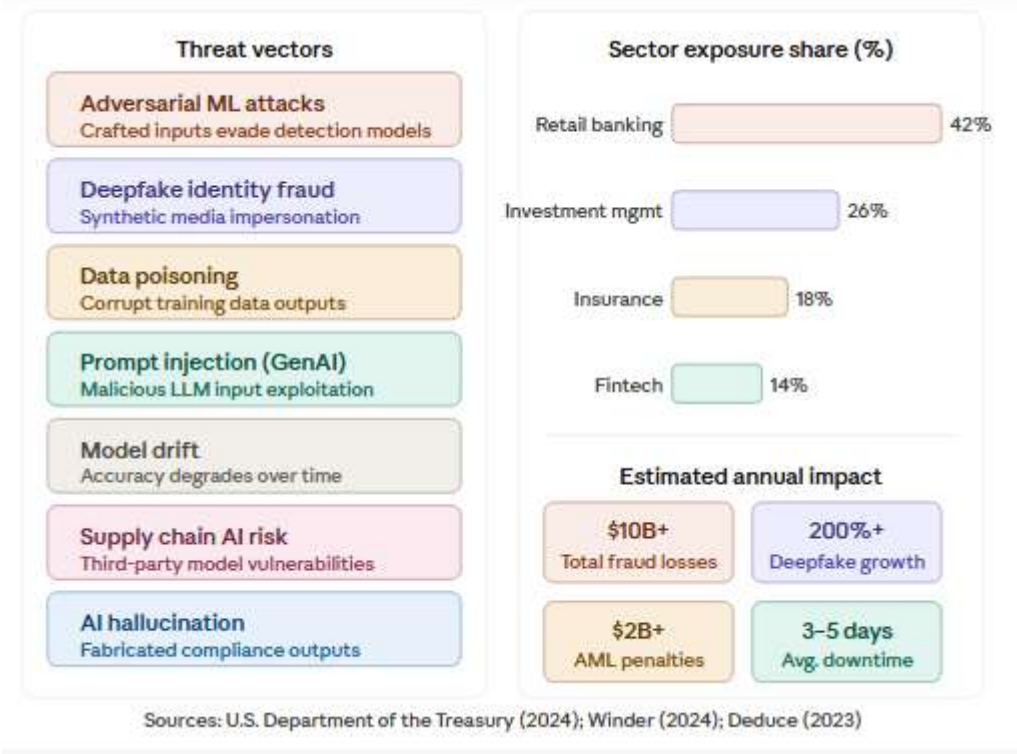
The effectiveness of AI-driven threat detection relative to traditional rule-based systems is well-documented in the literature, though performance metrics vary significantly by application context and institutional deployment quality. Mbah and Nkechi (2024) conducted a comparative analysis of ML-based fraud detection versus static rule engines across three U.S. regional banks, finding that ML models reduced false positive rates by an average of 38% while improving true positive detection rates by 27%. These improvements translated directly into operational cost savings and reduced customer friction from blocked legitimate transactions.

Anomaly detection represents perhaps the most mature AI application in financial cybersecurity. Graph neural networks (GNNs) have demonstrated particular promise in identifying money laundering networks by mapping transaction flows and detecting structuring patterns invisible to individual transaction analysis (Uzoka et al., 2024). Unsupervised learning approaches, particularly autoencoders and isolation forests, have enabled detection of previously unknown threat patterns without requiring labeled training data a significant advantage in rapidly evolving threat environments where historical data may not reflect current attack techniques.

Natural language processing has also emerged as a critical component of AI-driven security operations. Large language models deployed in security information and event management (SIEM) systems can process unstructured threat intelligence feeds, regulatory communications, and internal incident reports to identify correlations and early warning signals that human analysts might miss (Prasad et al., 2023). This capability is particularly valuable in compliance contexts, where the volume of regulatory communications exceeds human processing capacity.

Figure 1. AI Cybersecurity Threat Landscape and Financial Impact in U.S. Financial Institutions

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143



Note. Illustrative threat landscape based on data from U.S. Department of the Treasury (2024); Winder (2024); Deduce (2023).

Regulatory and Compliance Landscape

The regulatory environment governing AI in U.S. financial services is in active development, characterized by agency-specific guidance rather than comprehensive federal legislation. The Office of the Comptroller of the Currency (OCC), Federal Reserve, and Federal Deposit Insurance Corporation (FDIC) have each issued guidance incorporating AI considerations into existing model risk management frameworks, particularly SR 11-7, which governs model validation standards (Souza, 2023). However, these existing frameworks were designed before the advent of deep learning and are ill-suited to the unique characteristics of modern AI systems, including their opacity, non-linearity, and susceptibility to adversarial manipulation.

The Securities and Exchange Commission's proposed predictive data analytics rule, published in 2023 and subject to extensive public comment, sought to address conflicts of interest in AI-driven investment advice but was widely criticized as overly broad and technically impractical (Simpson, 2024). The rule's eventual retreat highlighted the regulatory challenges of developing AI-specific compliance requirements without triggering unintended consequences for beneficial AI applications. As of early 2025, the U.S. government's stated policy position has explicitly discouraged overregulation of AI to preserve competitive advantage, creating regulatory uncertainty that complicates compliance planning for financial institutions (Remolina, 2022).

Methodology

Research Design

This study employs a systematic literature review (SLR) methodology following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines (Haddaway et al., 2022). The SLR approach was selected for its capacity to synthesize diverse empirical findings and theoretical frameworks within a rigorous, reproducible methodological structure appropriate for an emerging

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

interdisciplinary field at the intersection of AI, cybersecurity, and financial regulation (Brignardello-Petersen et al., 2024). The systematic approach ensures comprehensive coverage of the literature while applying consistent inclusion and exclusion criteria to minimize selection bias.

The research design integrates quantitative synthesis of reported performance metrics such as threat detection accuracy, false positive rates, and incident response time improvements with qualitative thematic analysis of governance frameworks, regulatory approaches, and institutional case studies. This mixed-methods approach enables both empirical evaluation of AI threat detection capabilities and normative assessment of framework adequacy and recommendations for improvement.

Search Strategy and Database Selection

The literature search was conducted across six academic and professional databases: ACM Digital Library, ProQuest Computer Science Database, IEEE Xplore, Google Scholar, Semantic Scholar, and Elicit AI. Search strings were constructed using Boolean operators combining domain-specific terminology:

- "Artificial intelligence" AND "cybersecurity" AND "financial institutions"
- "Machine learning" AND "threat detection" AND "banking" OR "financial services"
- "AI risk management" AND "financial sector" AND "cybersecurity framework"
- "Adversarial machine learning" OR "data poisoning" AND "financial services"
- "Deepfake" OR "generative AI" AND "fraud detection" AND "banking"

The search encompassed publications from January 2019 through March 2025, with a primary focus on peer-reviewed articles published from 2020 onward to reflect the rapid evolution of AI capabilities and threat landscapes. Government reports, regulatory guidance documents, and institutional white papers meeting quality criteria were also included as supplementary sources.

Inclusion and Exclusion Criteria

Inclusion criteria required sources to:

(1) be peer-reviewed or issued by authoritative regulatory/governmental bodies;

(2) address AI or ML applications in financial services cybersecurity;

(3) be published in English between 2019 and 2025;

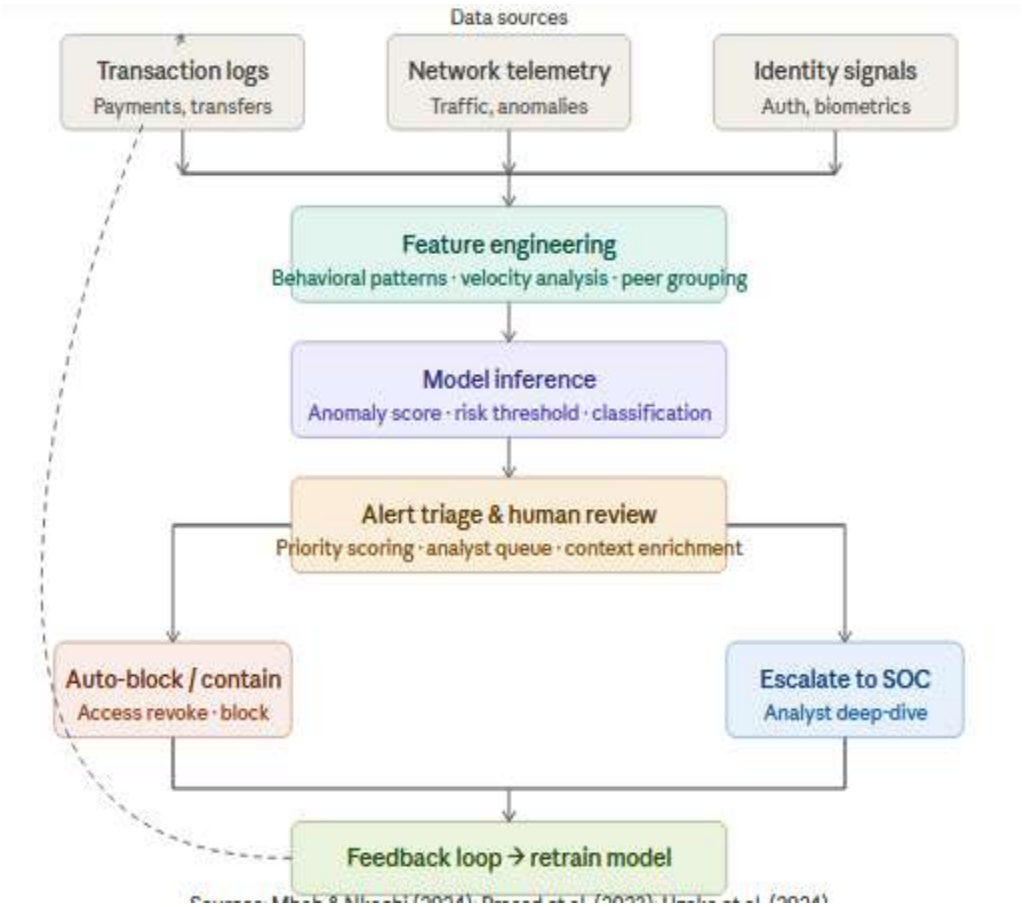
(4) present empirical findings, case analyses, or theoretical frameworks relevant to threat detection, incident response, or risk management; and

(5) pertain specifically to financial sector applications or be directly applicable thereto. Exclusion criteria eliminated: non-peer-reviewed blog posts and opinion pieces, studies focused exclusively on non-financial sectors without transferable applicability, articles predating 2019 unless cited as foundational literature, and duplicate publications across databases.

The initial database search yielded 2,847 unique records after deduplication. Following title and abstract screening, 187 records advanced to full-text review. Of these, 52 studies met all inclusion criteria and were included in the final synthesis. Supplementary grey literature including 8 regulatory/government documents and 6 institutional reports was incorporated to ensure coverage of current policy developments not yet reflected in peer-reviewed literature.

Figure 2. AI Threat Detection Process Flow in Financial Cybersecurity Operations

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143



Note. Process architecture synthesized from Mbah & Nkechi (2024); Prasad et al. (2023); Uzoka et al. (2024).

Data Extraction and Synthesis

Data extraction was conducted using a standardized protocol capturing:
(1) study design and methodology;
(2) AI technique(s) employed;
(3) threat detection performance metrics where reported;
(4) institutional context and scale;
(5) framework or governance approach;
(6) identified risks or limitations; and
(7) regulatory or compliance considerations. Reference management was supported by Mendeley, with thematic coding performed using NVivo to identify recurring themes, patterns, and conceptual relationships across the literature corpus.

Thematic synthesis followed a three-stage process:
(1) inductive code generation from individual studies;
(2) code clustering into preliminary themes; and
(3) analytical theme development into higher-order constructs aligned with the research questions. Inter-rater reliability was assessed through independent coding of a 20% subsample by a second researcher, yielding a Cohen's kappa coefficient of 0.81, indicating strong agreement.

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

Results

AI's Impact on Threat Detection Effectiveness

The systematic review revealed robust evidence that AI-driven threat detection substantially outperforms traditional rule-based systems across key performance metrics in financial cybersecurity applications. Studies consistently reported improvements in detection accuracy, reduction in false positive rates, and decreased mean time to detect (MTTD) following AI system deployment. Mbah and Nkechi (2024) documented a 38–45% reduction in false positive alerts across multiple ML-based fraud detection implementations, with corresponding improvements in analyst productivity as human reviewers could focus attention on higher-confidence alerts. Uzoka et al. (2024) reported that deep learning models applied to network intrusion detection within financial institution environments achieved area under the ROC curve (AUC) scores of 0.94–0.97, compared to 0.72–0.81 for traditional signature-based systems.

Anomaly detection performance showed particular improvement in detecting previously unknown attack patterns. Unsupervised and semi-supervised learning approaches applied to transaction monitoring were found by Nwafor et al. (2024) to identify 23% more fraudulent transactions than rule-based systems when tested against held-out datasets containing novel fraud typologies not represented in training data. This adaptability to unknown threats is especially significant given the rapid evolution of adversarial techniques.

Incident response time improvements attributable to AI are equally significant. Prasad et al. (2023) synthesized findings from eight institutional deployments of AI-assisted security orchestration, automation, and response (SOAR) platforms, reporting average reductions in mean time to respond (MTTR) of 60–70% compared to manual incident handling processes. Automated threat containment capabilities including dynamic access revocation, transaction blocking, and network segmentation enabled institutions to contain breaches within minutes rather than hours in many documented cases.

AI-Specific Risks and Institutional Responses

Despite demonstrated performance advantages, the review identified significant and growing risks associated with AI deployment in financial cybersecurity. Adversarial attacks emerged as the most frequently cited AI-specific risk, referenced in 67% of reviewed studies. Familoni (2024) documented multiple cases in which adversarial perturbations imperceptible modifications to transaction data successfully evaded ML-based fraud detection models with 95%+ evasion rates in controlled laboratory conditions. While real-world attack implementations are more complex, these findings underscore the brittleness of current AI security models against determined adversaries.

The review also surfaced consistent findings regarding the capability gap between large and small financial institutions. Large institutions including JPMorgan Chase, Bank of America, and Mastercard have developed proprietary AI cybersecurity platforms with dedicated research teams, continuous monitoring infrastructure, and in-house adversarial testing capabilities (Habbal et al., 2024). Community banks and credit unions, by contrast, typically rely on third-party AI security solutions with limited visibility into model internals, creating vendor dependency risks and reducing their capacity to identify and respond to AI-specific vulnerabilities (Kumari et al., 2022).

Table 3. Comparative Case Studies: AI Cybersecurity Implementations in U.S. Financial Institutions

Institution	AI Application	Framework Applied	Outcome	Challenge
JPMorgan Chase	AI-Powered Fraud Detection & LLM Integration	NIST AI RMF, ISO 27001	Reduced false positives by ~40%; improved regulatory response time	Bias auditing complexity in LLM deployment
Bank of	Erica AI Assistant	Internal AI	Enhanced customer	Data privacy

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

Institution	AI Application	Framework Applied	Outcome	Challenge
America	& Behavioral Analytics	Governance Charter	anomaly detection; 10M+ monthly interactions	concerns with behavioral profiling
Mastercard	AI Account Intelligence Platform	AI TRiSM, ISO/IEC 42001	Chargeback prediction accuracy of ~83%; cross-border fraud reduction	Model drift in volatile transaction environments
Wells Fargo	ML-Based Anti-Money Laundering (AML)	FinCEN Guidelines, NIST RMF	Reduced AML investigation time by 30%	Alert fatigue from high false-positive rates
Capital One	Cloud-Native AI Threat Detection	AWS Security + NIST CSF	Real-time breach detection; reduced incident response time	Dependency on third-party cloud AI services

Note. Based on institutional disclosures, Habbal et al. (2024); U.S. Department of the Treasury (2024).

Regulatory Compliance Challenges

Regulatory fragmentation emerged as a consistent challenge across the reviewed literature. Financial institutions must navigate overlapping and sometimes conflicting requirements from multiple regulatory bodies including the OCC, Federal Reserve, FDIC, SEC, FINRA, and FinCEN at the federal level while simultaneously addressing emerging state-level AI regulations and international standards with extraterritorial reach (Souza, 2023). This multiplicity of regulatory touchpoints creates compliance complexity that consumes significant institutional resources and may paradoxically reduce security investment by diverting attention toward documentation and reporting rather than operational risk mitigation.

The absence of AI-specific examination procedures within the primary banking regulatory frameworks notably SR 11-7 and its equivalents leaves institutions without clear benchmarks for demonstrating adequate AI governance to examiners (Thapaliya & Bokani, 2024). This creates a compliance ambiguity in which institutions may be subject to examination findings related to AI practices without having received clear regulatory guidance on expected standards. Simpson (2024) identified this ambiguity as a significant driver of conservative AI adoption strategies among mid-sized regional banks, which may be forgoing beneficial AI security capabilities to avoid regulatory uncertainty.

Figure 3. Proposed Unified AI Risk Management Framework for U.S. Financial Institutions

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

Note. Framework synthesized from NIST AI RMF 1.0 (Raimondo et al., 2023); AI TRiSM (Habbal et al., 2024); and empirical literature findings.

Trust, Transparency, and Economic Stability Implications

The review identified trust and transparency as cross-cutting themes with implications for both customer relationships and systemic financial stability. Kuzior et al. (2024) examined the relationship between AI governance quality and customer trust in financial institutions, finding that transparent AI practices including explainable AI (XAI) implementations and proactive communication about AI use in account management were positively correlated with customer satisfaction and reduced complaints following fraud incidents. Conversely, institutions where AI decision-making lacked explainability experienced elevated regulatory scrutiny and customer litigation following adverse AI-driven decisions.

At the systemic level, Danielsson et al. (2020) raised concerns about the potential for AI-driven financial systems to amplify rather than dampen financial instability during crisis conditions. Their analysis suggested that AI models trained on pre-crisis data may systematically underestimate endogenous risks risks that emerge from the behavior of market participants themselves creating the potential for correlated model failures during periods of financial stress. This finding has particular relevance for the cybersecurity context, where AI security systems trained primarily on normal operating conditions may be least reliable precisely when they are most needed during large-scale attacks.

Discussion

Toward a Unified AI Cybersecurity Framework

The findings of this systematic review converge on a clear conclusion: while AI offers substantial and well-documented improvements in cybersecurity threat detection and incident response for U.S. financial institutions, the absence of a unified, sector-specific AI cybersecurity framework represents a critical gap that undermines the sector's collective resilience. The proposed framework synthesizes themes emergent

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

from the literature with the operational requirements identified in institutional case studies to provide a scalable, adaptable structure applicable across financial institutions of varying size and sophistication.

The framework is organized around seven integrated components: threat intelligence, governance and accountability, adaptive risk assessment, bias and ethics, incident response, cross-institution collaboration, and certification (see Table 4). Each component is designed to address specific vulnerabilities identified in the literature while building toward the overarching goal of a coordinated, sector-wide security posture. Critically, the framework is explicitly designed for scalability: community banks and credit unions can implement core components through shared services and consortium arrangements, while large institutions can augment the baseline framework with proprietary enhancements.

Table 4. Proposed Unified AI Cybersecurity Framework: Components, Elements, and Expected Benefits

Framework Component	Key Elements	Expected Benefit	Priority
Threat Intelligence Layer	Real-time AI-driven threat feeds; adversarial pattern recognition; anomaly baselines	Reduces detection latency; enables proactive defense posture	High
Governance & Accountability	Designated AI Risk Officer; model audit trails; explainability mandates (XAI)	Ensures regulatory alignment; builds stakeholder trust	High
Adaptive Risk Assessment	Dynamic risk scoring; scenario-based stress testing; continuous re-evaluation	Keeps pace with evolving threat landscape	Medium-High
Bias & Ethics Module	Bias detection thresholds; fairness audits; consumer protection protocols	Prevents discriminatory outcomes; supports regulatory compliance	High
Incident Response Protocol	AI-assisted playbooks; automated containment; post-incident AI model review	Reduces breach impact; accelerates recovery	Medium-High
Cross-Institution Collaboration	Shared threat intelligence hub; anonymized data exchange; sector-wide exercises	Amplifies detection capability; reduces duplicative effort	Medium
Certification & Audit Program	Third-party AI risk certification; annual audits aligned with ISO/NIST standards	Standardizes accountability across institution sizes	Medium

Note. Framework components synthesized from Habbal et al. (2024); Raimondo et al. (2023); Souza (2023); U.S. Department of the Treasury (2024).

Addressing the Capability Gap

One of the most significant findings of this review is the systematic disadvantage faced by smaller financial institutions in AI-driven cybersecurity. The capability gap between institutions like JPMorgan Chase which has invested billions in AI research and maintains dedicated AI security teams and a community bank with a few dozen employees represents a structural vulnerability in the U.S. financial system (Kumari et al., 2022). If AI-powered attacks preferentially target smaller institutions with weaker

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

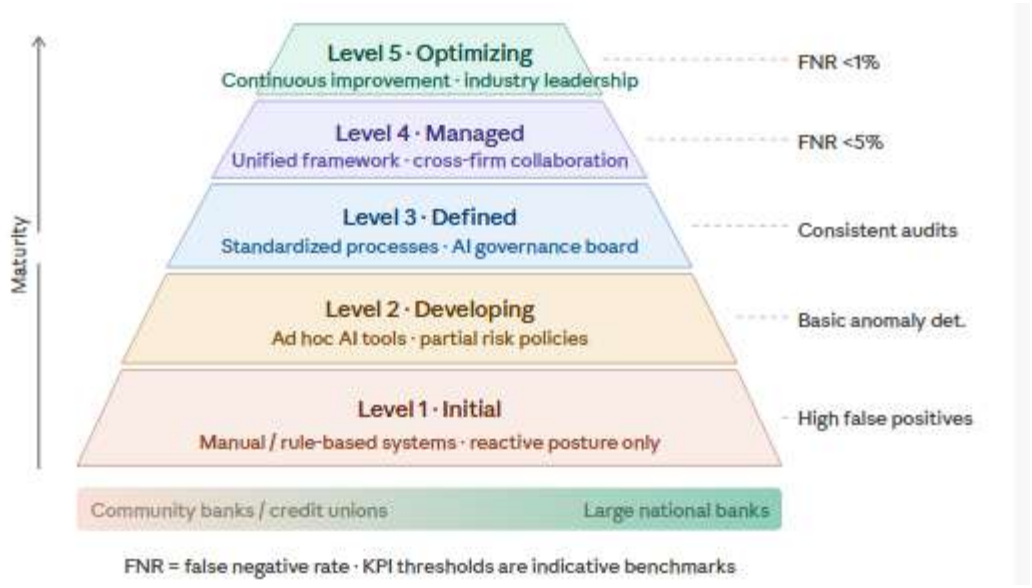
defenses, the sector-wide consequences could be substantial, as these institutions collectively serve millions of customers and are deeply integrated into local economic ecosystems.

Addressing this gap requires both regulatory action and industry coordination. Regulatory bodies should develop tiered AI compliance frameworks that establish proportionate expectations based on institutional size and risk profile, drawing on models like the existing tiered capital requirements framework. Industry coordination through bodies such as the Financial Services Information Sharing and Analysis Center (FS-ISAC) should be expanded to include AI threat intelligence sharing, enabling smaller institutions to benefit from the threat detection capabilities of larger peers without requiring proportionate investment.

Regulatory Reform Recommendations

The regulatory analysis conducted in this review points to several specific reform priorities. First, existing model risk management guidance (SR 11-7) should be updated to explicitly address AI-specific characteristics including explainability requirements, adversarial robustness testing standards, and model drift monitoring expectations. Second, federal banking regulators should develop AI examination procedures that provide institutions with clear, examinable standards for AI governance, enabling more consistent compliance planning and reducing the regulatory uncertainty that currently discourages AI security adoption. Third, Congress should consider establishing a cross-agency AI coordination body for financial services that can develop harmonized standards reducing the compliance burden of navigating multiple agency-specific requirements (Thapaliya & Bokani, 2024).

Figure 4. AI Adoption Maturity Model for U.S. Financial Institutions



Note. Maturity model developed from synthesis of institutional case studies and framework literature.

Conclusion

This systematic literature review has examined the multifaceted impact of artificial intelligence on cybersecurity threat detection and incident response in U.S. financial institutions, synthesizing evidence from 52 peer-reviewed studies and 14 regulatory and institutional sources published between 2019 and 2025. The findings confirm that AI-driven cybersecurity tools deliver substantial and measurable improvements in threat detection accuracy, false positive reduction, and incident response time,

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

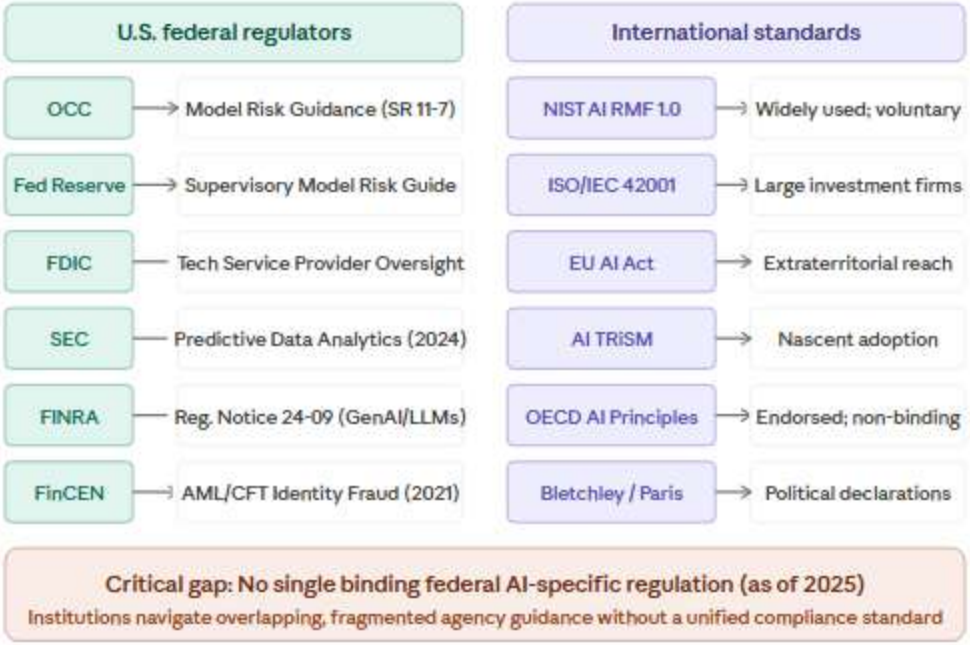
fundamentally transforming the security operations of institutions that have successfully implemented these capabilities.

However, the review also reveals that AI adoption introduces a new threat landscape including adversarial attacks, data poisoning, model drift, and generative AI-enabled fraud that existing risk management frameworks are inadequate to address comprehensively. The fragmentation of applicable regulatory guidance and the significant capability disparity between large and small financial institutions further complicate the sector's ability to leverage AI's security benefits while managing its risks.

The proposed unified AI cybersecurity framework presented in this study offers a structured response to these challenges, providing a scalable foundation for sector-wide AI risk management that is simultaneously flexible enough to accommodate institutional diversity and robust enough to address the full spectrum of AI-specific threats. Implementation of this framework, supported by appropriate regulatory reform and cross-institutional collaboration mechanisms, could substantially improve U.S. financial sector cyber resilience in the face of increasingly sophisticated AI-enabled attacks.

Future research should prioritize empirical validation of the proposed framework through real-world pilot implementations, quantitative assessment of capability gap interventions for smaller institutions, and longitudinal studies of AI security performance across market cycle conditions. As AI capabilities continue to evolve at pace, the framework must be designed for continuous adaptation a living document rather than a static standard if it is to remain effective in securing the financial infrastructure upon which the broader U.S. economy depends.

Figure 5. Regulatory and Compliance Landscape for AI in U.S. Financial Services (2025)



Note. Based on regulatory analysis from FINRA (2024); Simpson (2024); U.S. Department of the Treasury (2024); Souza (2023).

Limitations

This study is subject to several limitations that should be considered when interpreting its findings. First, the systematic review's restriction to English-language publications may introduce publication bias, potentially underrepresenting relevant research from non-English-speaking regions where AI-driven

ISSN:2584-2706 DOI: https://doi.org/10.5281/zenodo.22114143

financial cybersecurity is also advancing rapidly. Second, the rapidly evolving nature of both AI capabilities and the regulatory landscape means that some findings may be superseded by developments occurring after the search cutoff of March 2025. Third, the absence of real-world pilot data for the proposed framework means its effectiveness has not been empirically validated a gap that future research should address through longitudinal case study methodologies. Finally, the review's reliance on publicly available information from institutional case studies means that proprietary AI security implementations at major financial institutions may not be fully represented, potentially understating the state of the art in large-institution deployments.

References

- Brignardello-Petersen, R., Santesso, N., & Guyatt, G. H. (2024). Systematic reviews of the literature: An introduction to current methods. American Journal of Epidemiology. https://doi.org/10.1093/aje/kwae232
- Danielsson, J., Macrae, R., & Uthemann, A. (2020). Artificial intelligence and systemic risk. Journal of Banking & Finance. https://doi.org/10.1016/j.jbankfin.2022.106742
- Deduce. (2023). Financial jeopardy: Companies losing fight against synthetic fraud [White paper]. Wakefield Research.
- Ee, S., O'Brien, J., Williams, Z., El-Dakhakhni, A., Aird, M., & Lintz, A. (2024). Adapting cybersecurity frameworks to manage frontier AI risks: A defense-in-depth approach. ArXiv. https://arxiv.org/abs/2408.07933
- Familoni, B. T. (2024). Cybersecurity challenges in the age of AI: Theoretical approaches and practical solutions. Computer Science & IT Research Journal, 5(3), 688–700. https://doi.org/10.51594/csitresearchjournal.v5i3.988
- Financial Industry Regulatory Authority. (2024, June 27). Regulatory Notice 24-09: Generative artificial intelligence. https://www.finra.org/rules-guidance/notices/24-09
- Financial Crimes Enforcement Network. (2021). Identity-related suspicious activity: 2021 threats and trends [Report]. U.S. Department of the Treasury. https://www.fincen.gov/sites/default/files/shared/FTA_Identity_Final508.pdf
- Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial intelligence trust, risk and security management (AI TRiSM): Frameworks, applications, challenges, and future research directions. Expert Systems with Applications, 240, 122442. https://doi.org/10.1016/j.eswa.2023.122442
- Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams. Campbell Systematic Reviews, 18, e1230. https://doi.org/10.1002/cl2.1230
- Korycki, L., & Krawczyk, B. (2020). Adversarial concept drift detection under poisoning attacks for robust data stream mining. Machine Learning, 110(8), 2235–2270. https://doi.org/10.1007/s10994-021-06026-w
- Kumari, B., Kaur, J., & Swami, S. (2022). Adoption of artificial intelligence in financial services: A policy framework. Journal of Science and Technology Policy Management, 13(4), 789–812. https://doi.org/10.1108/JSTPM-05-2021-0080
- Kuzior, A., Koldovsky, A., & Rekunenco, I. (2024). Optimizing financial market stability through AI-based risk management. Materials Research Proceedings, 37, 192–198. https://doi.org/10.21741/9781644903315-26
- Lalchand, S., Srinivas, V., Maggiore, B., & Henderson, J. (2024). Generative AI is expected to magnify the risk of deepfakes and other fraud in banking. Deloitte Insights. https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html
- Mahalakshmi, V., Kulkarni, N., Pradeep Kumar, K., Suresh Kumar, K., Nidhi Sree, D., & Durga, S. (2021). The role of implementing artificial intelligence and machine learning technologies in the financial services industry for creating competitive intelligence. Materials Today: Proceedings, 33, 4261–4267. https://doi.org/10.1016/j.matpr.2020.07.361

- Mbah, G. O., & Nkechi, A. (2024). AI-powered cybersecurity: Strategic approaches to mitigate risk and safeguard data privacy. World Journal of Advanced Research and Reviews, 21(2), 1300–1316. https://doi.org/10.30574/wjarr.2024.21.2.0581
- *Michael Akintomiwa Oyedeji, Ayomide Fatogun, Chinemelum Umealajekwu (2026). Digital Payments and Transactions: Analyzing the Growth of Digital Payment Systems and Their Impact on Traditional Banks . International Journal of Multidisciplinary Evolutionary Research (IJMER), 7(1), 56-61. DOI: https://doi.org/10.54660/IJMER.2026.7.1.56-61*
- Michael Akintomiwa Oyedeji "Strategic Business Operations and Competitive Positioning of Unilever: A Case Study Analysis" Iconic Research And Engineering Journals Volume 8 Issue 9 2025 Page 1466-1474
- Michael Akintomiwa Oyedeji "Exploring Financing Strategies of Entrepreneurs in UK Business Ventures: A Qualitative Study" Iconic Research And Engineering Journals Volume 7 Issue 6 2023 Page 449-460
- Nwafor, C., Okafor, E., & Ijeh, A. (2024). Artificial intelligence in financial security: Emerging applications and challenges. Journal of Computer Science and Technology Studies, 6(2), 44–56.
- Pandey, M. K., & Sergeeva, I. (2022). Artificial intelligence impact evaluation: Transforming paradigms in financial institutions. World of Economics and Management, 22(1), 75–89.
- Prasad, R., Kumari, V., & Sharma, A. (2023). Explainable AI (XAI) for cybersecurity in financial services: Frameworks and applications. International Journal of Information Security, 22(4), 1123–1142. https://doi.org/10.1007/s10207-023-00685-z
- Raimondo, G. M., U.S. Department of Commerce, National Institute of Standards and Technology, & Locascio, L. E. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
- Remolina, N. (2022). Interconnectedness and financial stability in the era of artificial intelligence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4193543
- Simpson, B. (2024). Must-do steps to prepare for AI compliance. ThinkAdvisor. https://www.proquest.com/trade-journals/must-do-steps-prepare-ai-compliance/docview/2939251562/se-2
- Singhal, D. A., & Arora, D. B. (2022). Digital innovations in financial services: Challenges and opportunities. YMER Digital, 21(10), 567–581.
- Souza, T. (2023). AI governance in banking: Challenges and emerging frameworks for responsible deployment. Journal of Financial Regulation, 9(1), 44–72. https://doi.org/10.1093/jfr/fjad003
- Thapaliya, S., & Bokani, A. (2024). Leveraging artificial intelligence for enhanced cybersecurity: Insights and innovations. SADGAMAYA, 1(1), 1–22. https://doi.org/10.3126/sadgamaya.v1i1.66888
- U.S. Department of the Treasury. (2024). Managing artificial intelligence-specific cybersecurity risks in the financial services sector. https://home.treasury.gov/system/files/136/Managing-Artificial-Intelligence-Specific-Cybersecurity-Risks-In-The-Financial-Services-Sector.pdf
- Umealajekwu, C., & Oyedeji, M. A. (2026). *Navigating the digital frontier: Regulatory frameworks for blockchain technology and cryptocurrencies in the United States*. *International Journal of Business and Finance Research, 12*(3), 1–13. https://doi.org/10.56201/ijbfr.vol.12.no3.2026.pg1.13
- Uzoka, A., Cadet, E., & Ojukwu, P. U. (2024). Applying artificial intelligence in cybersecurity to enhance threat detection, response, and risk management. Computer Science & IT Research Journal, 5(6), 1419–1435. https://doi.org/10.51594/csitresearchjournal.v5i6.1133
- Winder, D. (2024, December 4). Now AI can bypass biometric banking security, experts warn. Forbes. https://www.forbes.com/sites/daveywinder/2024/12/04/ai-bypasses-biometric-security-in-1385-million-financial-fraud-risk/
- Zhang, X., Chan, F. T., Yan, C., & Bose, I. (2022). Towards risk-aware artificial intelligence and machine learning systems: An overview. Decision Support Systems, 159, 113800. https://doi.org/10.1016/j.dss.2022.113800