**Issues in Information Systems**
Volume 26, Issue 2, pp. 489-507, 2025

DOI: https://doi.org/10.48009/2_iis_2025_136

# Unifying risk management for AI-driven cybersecurity in finance.

**Simeon, Olaomo,** *Middle Georgia State University, drsimeon@proton.me*

## Abstract

Artificial intelligence is reshaping organizational strategy, with its influence on the financial services sector continually expanding and introducing novel risks. This study examines the impact of AI-driven cybersecurity on the risk management framework in financial services through a systematic literature review, identifying key themes and research gaps. While existing studies primarily focus on AI-related risks, this study extends this by exploring how organizations can leverage unified risk management strategies to manage the rise of AI-driven cybersecurity in financial services. It aims to unify diverse approaches within the financial sector to develop a robust, scalable risk-management framework applicable across organizations, regardless of size or structure. A common theme among researchers is the need for a comprehensive risk-management strategy, resource collaboration, and data-sharing capabilities to enhance threat detection, analysis, and intelligence sharing. The thematic benefits of this unified approach include collaboration among stakeholders, innovation, holistic risk management, the eradication of fragmented approaches, trust and transparency, reduced cost, and increased efficiency.

**Keywords**: risk management framework, artificial intelligence, cybersecurity, financial services, machine learning, critical infrastructure

## Introduction

Over the past few decades, there have been noticeable and profound changes in the financial services industry's operational elements and information technology (IT) infrastructure (Singhal and Arora, 2022). Machine learning (ML) and artificial intelligence (AI) have been implemented by organizations, including all financial service providers, at various stages of their operations to bolster financial management, including automating risk management, fraud detection, trading activities, and providing financial advice to customers. (Mahalakshmi et al., 2021). The new features introduced by ML and AI have changed and strengthened user experience and user acceptance, but have also complicated financial operations, risks, and regulatory frameworks. The financial system is a vital component of any economy that ensures social stability and security, increasing the dangers associated with implementing AI in such a complex environment. The principles of including AI syndicates apply to many facets of financial institutions, particularly when these technologies are deployed in test or production environments where the disclosure and manipulation of sensitive data make the organization more vulnerable to actual attacks (Dańıelsson et al., 2020; Remolina, 2022).

The commercialization of various AI products and services, such as OpenAI ChatGPT, Claude, Microsoft Copilot, Google Gemini, and DeepSeek, along with the recent race to AI dominance, has led to increased awareness and adoption of AI concepts around the world. Although this is a greatly welcomed development, it has sparked concerns about the exponential risks AI creates and their exposure levels. According to

Korycki et al. (2020), there are several risks associated with the adoption and use of AI and machine learning, including data poisoning, adversarial attacks, and model or concept drift. These issues can be challenging to identify and investigate without using appropriate tools. Dańıelsson et al. (2020) concluded from their investigation that financial services can benefit from AI technology when managing small-scale financial chores (exogenous risks), as it increases efficiency and lowers costs when clear rules and patterns are established. However, this cannot be said for situations with significant problems and financial instability. In these situations, AI is challenged by the complexity and unpredictable nature of crises (endogenous risks). This results in AI having difficulty adapting rules and being unable to understand causation comprehensively. While beneficial for microregulation, AI might replace some jobs in risk management. In macro regulation, its limitations could pose significant risks; therefore, its use should be carefully controlled. Thus, researchers assert that relying too heavily on AI in these focus areas could lead to a loss of human oversight. Despite the concerns raised by Danielson et al., (2020), financial institutions have extensively progressed with predictive and generative AI to reduce fraud, increase productivity, profitability, and career movements within the organization, and increase customer trust (Pandey et al., 2022). AI applications in cybersecurity pose several common risks, as presented in Table 1. In this process, banks and other financial service providers are also aware of their risks and must prioritize addressing these risks, as well as the traditional risks embedded within their regulatory obligations.

Table 1. Common AI risks identified by Remolina (2022) and US Treasury Departments (2024)

| Financial Institutions AI-Specific Risks |
|---|
| Adversarial attacks |
| Model and concept drifting |
| Data poisoning |
| Intended and unintended Bias in training data |
| AI Hallucination |
| Complex internal working structures are not easily explainable or auditable (Arun, 2020). |
| Data privacy leakages |
| Data biases and outcomes |
| Complications of financial institutions' systemic or traditional risks |
| Generative AI fraudulent tactics |

**Problem statement**

There is no turning back from the progress made by financial service systems around the world, as they have recognized the advantages and have a growing interest in implementing AI in their cyber operations. In the future, AI will involve significant risk. In 2024, the US Department of Treasury produced a comprehensive white paper on managing AI-related cyber risks in the financial services sector. The study finds that most participating financial institutions use AI-powered solutions extensively and play various roles. However, global banks, including US banks and credit card companies, have been struggling with AI-driven cyber risks. In a 2024 article, the Deloitte Center for Financial Services reported that an employee

of a Hong Kong firm transferred $25 million to a deep-faked fraudster account after cloning the organization's chief financial officer through a video call. (Lalchand, et al., 2024, para. 1). Similarly, as Generative AI increases, a Forbes tech journalist, Winder (2024), noted an increase in AI's ability to bypass biometric banking security and compromise the organization's defense layers, leading to over 1000 deepfake compromises in Indonesia's financial institution. Most interviewees in the US Treasury Department report, regardless of size, aligned their use of AI by adhering to the National Institute of Standards and Technology's AI Risk Management (NIST RMF) recommendations. However, the prevailing characteristics of AI usage in these institutions exist in silos and lack standardization, and many of the current defense-in-depth strategies operate at a slower speed compared to the level of compromise that technology such as generative AI and neural network damage can cause, making it difficult to effectively coordinate the progress and advancements made by individual institutions. The challenges AI faces in the financial sector are not dissimilar to other industries, but circumstantially unique and difficult to assess and evaluate (US Treasury Department, 2024). In 2023, in partnership with Wakefield Research, Deduce. com published a report on financial jeopardy. The results detailed real issues faced by financial institutions with the rise of sophisticated synthetic fraud fueled by AI-powered identity fraud. Despite having traditional synthetic fraud detection systems in place, several organizations are failing to prevent these advanced AI-driven fraud schemes and are losing billions of dollars (Deduce, 2023).

This study aims to unify the available solutions by providing a comprehensive framework to address AI-specific cyber risks and assist financial institutions in overcoming these challenges. The findings of this study will bridge existing gaps by offering standardized guidelines for AI risk management, particularly focusing on the integration of AI models, improving interoperability across financial institutions, and addressing the disparity in AI expertise between small and large organizations. Moreover, by proposing a common risk management framework for the financial sector, this study seeks to reduce the inconsistencies that lead to vulnerabilities, especially in the areas of fraud detection and cyber resilience. This study also addresses the unique challenges faced by institutions dependent on third-party AI systems or those with limited in-house AI capabilities, thereby ensuring a more cohesive approach to AI risk management in the industry (Kumari et al., 2022).

**Purpose of the study**

The purpose of this study is to develop a unified risk-management framework tailored for AI-driven cybersecurity systems within financial systems by conducting a systematic literature review. Considering the increased use of AI systems to secure critical infrastructure, such as power grids, medical records, and financial institutions, it is vital to understand the risk factors that result from implementing such systems. The findings of this study will directly benefit financial regulators and institutions by providing standardized guidelines for managing AI-specific cybersecurity risks, improving their ability to assess vulnerabilities, and adopting AI technologies safely. By ensuring better control of these risks, this study can help regulators develop more effective policies and aid financial institutions in making informed decisions regarding AI adoption and integration. The specific objectives of this study are to establish, identify, and control the risks associated with AI utilization by meshing AI risk findings with the policies and methodologies currently in use. This supports the safe adoption of AI systems in financial cybersecurity, particularly focusing on building trust and reducing vulnerabilities in AI-powered financial services.

**Research Questions**

Based on the purpose of the study, the following questions will be answered

**RQ1**: How does the current risk management framework align with adopting AI-driven cybersecurity solutions in financial systems, and what gaps exist in its current application?

**RQ2:** What existing risk-management strategies in AI-driven cybersecurity solutions are being utilized by financial institutions, and how can these be adapted or expanded to address emerging risks and challenges?

**RQ3:** What regulatory and compliance challenges do financial institutions face when implementing AI-driven cybersecurity solutions, and how can these challenges be mitigated through improved frameworks?

**RQ4:** How does implementing a new AI-driven risk management framework influence trust among customers and financial service providers, and impact overall economic stability?

### The Objectives of The Research

The objective of this research is to advance the field of AI-driven cybersecurity by exploring how organizations can leverage unified risk-management strategies to manage the rise of AI-driven cybersecurity in financial services through a comprehensive risk-management framework tailored to the evolving needs of financial institutions and other industries. This study seeks to identify and catalog AI-specific risks, including new and existing vulnerabilities introduced by AI adoption, and design a scalable risk assessment framework adaptable to the stages of AI development. Additionally, the research aims to propose innovative mitigation techniques, such as adversarial training, enhanced human-AI collaboration, and bias detection strategies, to address AI vulnerabilities. By investigating the impact of model drift on cybersecurity and evaluating advanced monitoring approaches, this study strengthens AI resilience in dynamic threat landscapes. A case study involving a U.S. bank is used to simulate and validate the effectiveness of the proposed framework, ensuring practical applicability and industry relevance.

## Literature Review

### Current Financial Services Risk Management Framework

The financial services industry is highly complex, making effective risk management a critical component of institutional operations and governance. Traditionally, these institutions worked with the US Government, risk management professionals, and in-house teams to address financial risks based on established principles (Fabozzi & Drake, 2010). In 2024, the US Treasury reported that financial institutions across the US largely depend on the NIST Risk Management Framework (RMF) to govern their risk management approaches. According to data on the NIST website (https://csrc.nist.gov/projects/risk-management/about-rmf), the NIST RMF covers new and legacy systems, including IoT and control systems. However, the adoption of AI models in financial services has constantly evolved from rule-based systems in the 1990s and the early 2000s to an enormous data-driven approach (Ee et al., 2024), introducing new risks and challenges that require enhancements in governance, compliance, policies, and risk management frameworks (Souza, 2023).

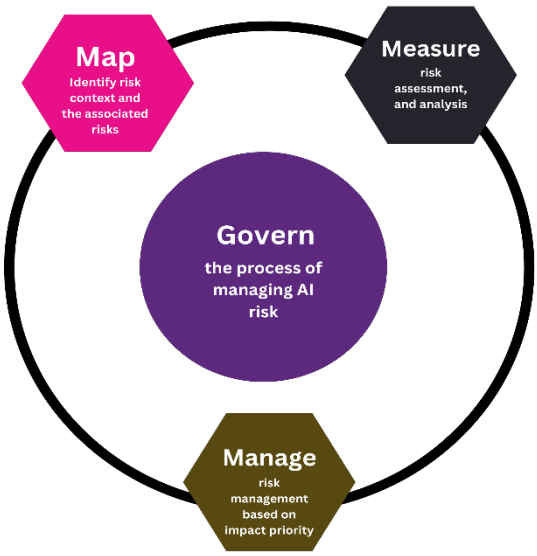
Table 2: Global Overview of AI Risk Management Frameworks for Financial Institutions

| AI Risk Management Frameworks | Institutions | Focus Area |
|---|---|---|

| | | |
|---|---|---|
| NIST AI RMF | U.S. financial institutions, such as JPMorgan Chase and the Bank of America. | Risk management, trustworthiness, financial algorithms, fraud detection, credit scoring, and data privacy. |
| ISO/IEC 42001 | Investment firms such as Goldman Sachs and Wells Fargo. | control and monitor AI systems' behavior in finance |
| ISO/IEC JTC 1/SC 42 | Global banks such as HSBC and Citibank. | fraud detection, algorithm development, AI system lifecycle management, and decision-making processes. |
| EU AI Act | European banks such as Deutsche Bank and BNP Paribas. | classified AI systems based on risk levels and compliance mandates for high-risk systems (e.g., credit scoring and trading algorithms). |
| OECD AI Principles | Linked to global financial institutions, such as UBS. | AI guidelines for ethics, transparency, and accountability. (Habbal et al., 2024) |
| Artificial Intelligence Trust, Risk, and Security Management (AI TRiSM) Frameworks | Finance, Healthcare, and the Metaverse | Framework for reliability and trustworthiness of AI systems |

(Refer to organizational websites: (*AI Risk Management Framework*, n.d.))

NIST AI RMF is among the existing risk management frameworks governing both the financial sector and other industries. A Palo Alto report indicated that this framework is a NIST collaborative effort with public and private organizations to respond to the vast complexity of AI implementation at different stages of



operations. The main emphasis is on four principles: govern, map, measure, and management. This framework is an advancement to the traditional RMF but also carries limitations such as a lack of enforcement capacity, and a steep learning curve, and may be untenable for small organizations.AI Risk Management Frameworks Core (Raimondo et al., 2023)

**Figure 1: AI RMF Core (NIST, 2023)**

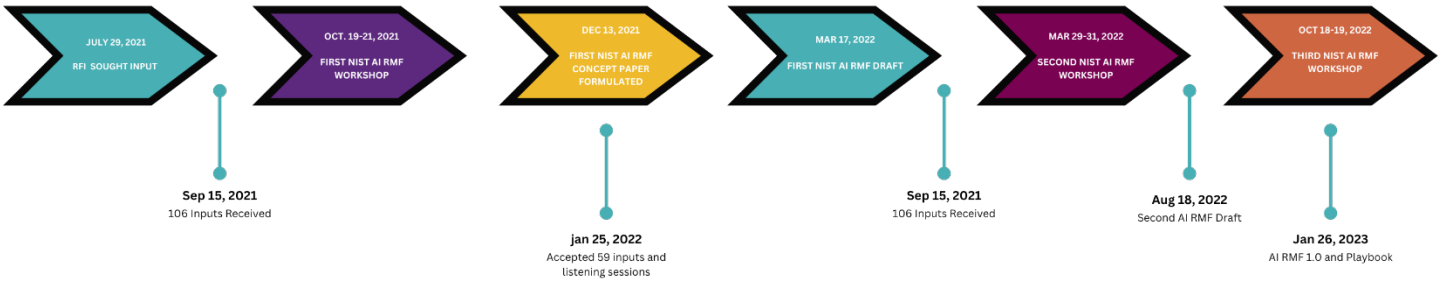



**Figure 2: AI RMF Timeline (NIST, 2024 publication)**

### AI-Driven Cybersecurity and Risk Management

Financial institutions have the potential for operational excellence by leveraging machine learning and predictive analytics for threat detection, incident response, and anomaly detection in transactions and operations (Mbah & Nkechi, 2024; Nwafor et al., 2024). To fully harness these capabilities, institutions must balance technological innovation with ethical considerations and regulatory compliance (Lee et al., 2021; Mbah and Nkechi 2024). As AI technologies advance, their role in combating sophisticated cyberattacks has become increasingly critical. Consequently, financial institutions must continuously develop risk management strategies to account for AI's evolving risks (Uzoka et al., 2024). The risks associated with deploying AI to combat micro- and macroeconomic challenges unfold new attack doors for threat actors to target the AI model and the associated LLM. This highlights the need for AI-specific frameworks to manage new and dynamic risks.

### JPMorgan Chase AI Innovation Case

According to reports from the JPMorgan Chase technology website, JPMorgan Chase is all in on AI adoption, using it to reshape how it operates and serves clients. They are particularly focused on large language models (LLMs), which help to analyze massive datasets and generate natural language insights. With such powerful technology, the firm places strong emphasis on governance, risk management, and ethical considerations, especially in highly regulated industries such as finance and healthcare. They are

committed to using responsible, explainable, and ethical AI principles to prevent bias and ensure that their systems are transparent and safe, with dedicated teams working to assess and mitigate any risks involved. Beyond the technology itself, JPMorgan Chase is the leader in AI innovation and integration. They have been recognized as top performers in the Evident AI Index (EAI), reflecting their ongoing investment and commitment to AI excellence. The company uses AI not only to streamline internal processes but also to enhance client services, making work more efficient and responsive to customer needs. They have made AI a core part of their business strategy, pairing senior leaders with AI experts to spot new opportunities and embed AI across their platforms and operations. With their ongoing focus on responsible AI, JPMorgan Chase is ensuring that they stay ahead of the curve while keeping client interests and regulatory standards at the front and center. JPMorgan Chase has also created the AI Maker Space in collaboration with Carnegie Mellon University and extended partnerships with the D^3 group, a leading AI research organization, to further invest in AI research and innovation. The group also has in-house AI researchers and partners with both undergraduate and graduate students across the world to bolster its financial services AI research. (jpmorganchase.com)

**MasterCard AI Account Intelligence for Fraud and Risk Management Case**

Mastercard, a global financial service provider, uses AI-powered solutions to enhance fraud detection and manage risks at the account level through its AI Account Intelligence platform. This platform uses adaptive AI models that analyze global anonymized data to identify patterns and trends that traditional data analytics may miss. It provides actionable insights across three main categories: account value and risk scores, account merchant activity scores, and account charge-back risk scores. These scores help predict risks such as fraud, account re-issuance, and chargebacks, with the model being able to forecast chargebacks up to 83% accurately and predict the chargeback amount within 15% of the actual value.To further scale its AI efforts, Mastercard's AI Garage focuses on developing machine-learning algorithms that are scalable, low-latency, and secure. These algorithms ensure safer transactions across millions of merchants and billions of payments. Their commitment to responsible AI means fairness, interpretability, and security in all AI developments. AI Garage also fosters professional growth by mentoring employees and collaborating across teams to generate new ideas and innovative solutions, positioning Mastercard as the leader in AI-driven financial services (Mastercard.com).

**Adapting existing risk management frameworks for AI Challenges**

AI technologies will continue to advance and become vital in combating sophisticated cyberattacks. This requires ongoing development of risk management strategies (Uzoka et al., 2024). Current RMFs, such as those outlined by Arogundade (2023), include steps such as preparation, categorization, selection, implementation, assessment, authorization, and monitoring. These have been applied to new technologies such as cloud systems, risk identification, and mitigation techniques. Although these frameworks have benefits, they must be revised to address the gaps in functionality, lifecycle management, and threat approaches related to AI-driven cybersecurity (Ee et al., 2024). Incorporating privacy-by-design principles and a multi-layered approach with human expertise for governance and continuous monitoring is essential for AI security solutions (Familoni, 2024; Mbah & Nkechi, 2024).

**Regulatory Challenges and Compliance in AI Integration**

As AI becomes increasingly embedded in financial systems, regulatory and compliance challenges arise, particularly in terms of data privacy, transparency, and ethical considerations. Financial institutions face significant difficulties in navigating these complexities, especially with AI models' potential biases and the opacity of decision-making processes. A regulatory approach that ensures compliance while promoting

innovation is crucial in fostering trust in AI-driven systems. The US Treasury (2024) highlighted the necessity of regulatory frameworks to adapt to AI's challenges while ensuring financial stability. In response, researchers such as Thapaliya and Bokani (2024) emphasize that traditional regulatory models must be updated to better accommodate the unique risks of AI-driven financial systems. Currently, there is no widespread regulation of AI in the United States. In February 2025, the current administration declared that there would be no overregulation of AI activities as an incentive for the United States to take the lead as an AI superpower. However, many organizations and government bodies (federal, state, and municipal) have regulatory planning underway. Global attention has also been paid to AI regulations, as seen by the Bletchley Declaration in 2023 and the Paris AI Action Summit in February 2025. Although there is considerable support for the advancement of AI, there are differences in regulatory propositions. A report by Simpson (2024) highlighted advancements in AI regulation by the Securities and Exchange Commission (SEC) regarding the control of predictive data, but was largely criticized for its strict and untenable proposals. In June 2024, FINRA issued a memo reminding its members to adhere to regulatory requirements when adopting generative artificial intelligence and large-language models (LLMS). For example, FINRA recommends that firms adopting Gen-AI develop policies and procedures that address aspects such as technology governance, management of model-related risks, data protection and accuracy, and ensuring the dependability and correctness of the AI model (FINRA, 2024).

**Risk Management Frameworks for Critical Infrastructure**

Critical infrastructure (CI) frameworks are essential assets in cybersecurity risk management, as they sustain national security and the economy. Kure and Islam (2019) introduced an asset-focused risk-management framework that focuses on vulnerability identification in interconnected assets, such as financial services. Julia et al. (2023) argue that an IT risk framework in financial services requires a comprehensive approach that includes governance, culture, communication, and risk monitoring. This broader framework is crucial for addressing systemic risks, particularly as AI technologies introduce new vulnerabilities.

**Collaborative Cybersecurity Frameworks and Addressing Systemic Risks in Financial Systems**

The complexity and interdependencies within financial systems necessitate coordinated platforms for effective cyber-security management. However, a 2024 report from the U.S. Department of the Treasury found that U.S. financial institutions have not fully embraced this collaborative approach. El Amin et al. (2024) proposed the use of blockchain technology for a decentralized and transparent cybersecurity framework that enhances collaboration and trust among stakeholders. Blockchain offers an avenue for managing risks across organizational boundaries, thus improving financial security. Traditional financial systems face inherent systemic risks, and the failure of one entity can trigger widespread consequences on the network (U.S. Department of the Treasury, 2024). Hoffmann (2020) argues that existing banking frameworks have been overly rigid, limiting their effectiveness in addressing systemic risks. By incorporating systems thinking and causal loop models, a more holistic approach to risk Management can tackle the interdependencies and complexities of financial markets. This is essential. because AI introduces new dynamics and amplifies systemic risks.

**AI-Driven Risk Management Frameworks and Toward a Unified Risk Management Framework**

Small and large financial institutions recognize the benefits of developing AI-driven risk management frameworks, which can enhance market stability and trust among stakeholders. AI improves decision-making and reduces volatility, but introduces concerns about the stability of AI models, including issues of bias and ethics (Kuzior, 2024). Models such as Type-2 Fuzzy Logic and the AI Trust, Risk, and

Security Management (AI TRiSM) frameworks have shown promise in promoting trust and security in AI systems (Adams & Hagras, 2020; Habbal et al., 2024). However, the integration of AI also brings challenges related to bias, economic inequality, and job displacement, which need to be addressed collaboratively through human-centered approaches in AI design (Tjondronegoro et al., 2022). The unification of risk-management frameworks for AI-driven cybersecurity systems in financial institutions is both urgent and important. As noted by the US Department of Treasury (2024), synthesizing various approaches is essential for managing the dynamic risks posed by AI. Ahmed et al. (2023) suggest integrating attack trees and residual risk management frameworks, which could provide a structured approach to managing threats in financial systems. The lack of explainability, AI bias, and siloed operations in AI models for fraud detection present ongoing challenges, underscoring the need for a cohesive, unified risk management framework that supports continuous adaptation and threat monitoring.

### Critical Analysis of Limitations in Existing Studies

While the literature provides valuable insights into the evolution of risk management frameworks, several limitations have emerged. First, many studies, including those by Mbah and Nkechi (2024) and Nwafor et al. (2024), highlight the potential of AI-driven cybersecurity but do not fully address the practical implementation challenges that institutions face in adopting these frameworks. Additionally, while frameworks such as the NIST RMF and AI Trust models have been proposed, their real-world applicability remains limited owing to a lack of empirical evidence on their effectiveness in diverse financial environments. Furthermore, existing studies often overlook the complexity of integrating AI models into legacy financial systems as well as the challenges related to ethical considerations, such as AI bias and fairness.

## Methodology

This study employs a systematic literature review (SLR) to examine existing AI risk-management frameworks for AI-driven cybersecurity in financial systems, as well as the importance of unifying these frameworks. The central point includes understanding the existing frameworks, advocating for a common AI framework, addressing capability gaps among financial institutions and human capital deficiencies, and expanding NIST AI frameworks (Dhir et al., 2020; Singh et al., 2023). To achieve this, the researcher conducts a rigorous literature review, first examining a variety of articles, publications, newsletters, government releases, private organization reports, and article databases, including ProQuest, Computer Source, ACM Digital Library, Google Scholar, Elicit, and Semantic Scholar. This also includes developing research questions, describing and refining literature searches based on specific search criteria and their respective sources, which form the inclusion and exclusion criteria (Krnic Martinic et al., 2019).

### Search Strategy and Database Selection

The literature search was conducted using specific keywords and Boolean search strings, ensuring comprehensive retrieval of relevant studies. The search terms were refined iteratively, focusing on peer-reviewed articles published within the last five years to maintain relevance, given the rapidly evolving nature of AI and cybersecurity.

#### Research Search Strings

"Artificial intelligence (AI)" OR "risk management"
"AI and machine learning in financial services"
"Risk management AND AI cybersecurity in financial services"

"AI-driven Risk Management Framework AND critical Infrastructure"
"Risk management and cybersecurity" OR “Unifying Risk Management Framework”

**Systematic Review Flow Diagram**

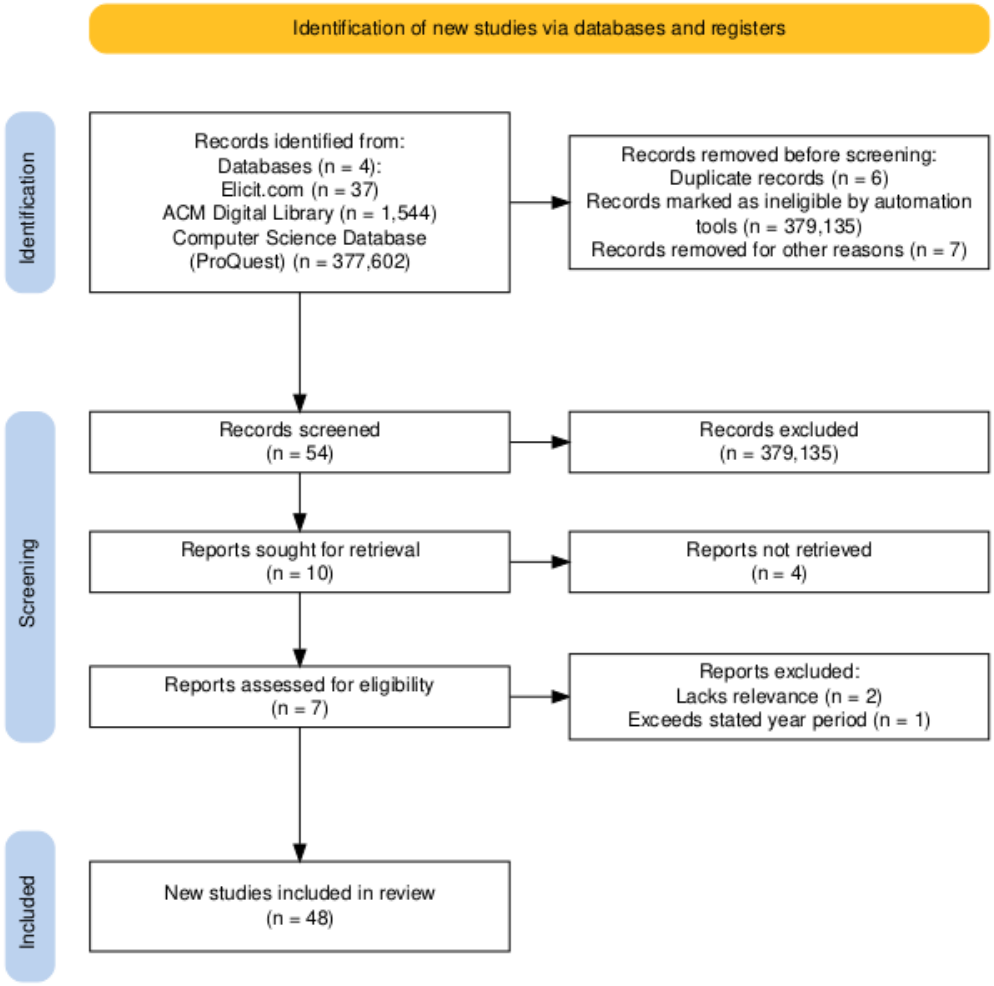



**Figure 3: PRISMA Flow Diagram**

**Inclusion and Exclusion Criteria**

Table 3: Inclusion and Exclusion criteria

| Criteria Type | Description |
|---|---|
| Inclusion Criteria | • Peer-reviewed articles (2020 onwards),<br>• English-language publications,<br>• Focus on AI-driven cybersecurity<br>• Focus on AI-driven cybersecurity risk management, |

| | |
|---|---|
| | • Relevance to research questions.<br>• Financial services AI risks and management |
| Exclusion Criteria | • Non-peer-reviewed sources,<br>• Non-English articles,<br>• studies unrelated to AI risk management,<br>• Outdated methodologies. |

**Data Extraction and Synthesis**

Each selected study was systematically analyzed, and key findings were categorized into thematic areas. The research employed software tools like AI-based engines, namely Elicit, Tera, Catchii, Excel, and Mendeley, for reference management, data organization, and citation tracking, enhancing accuracy and efficiency in data handling.

**Limitations and Challenges**

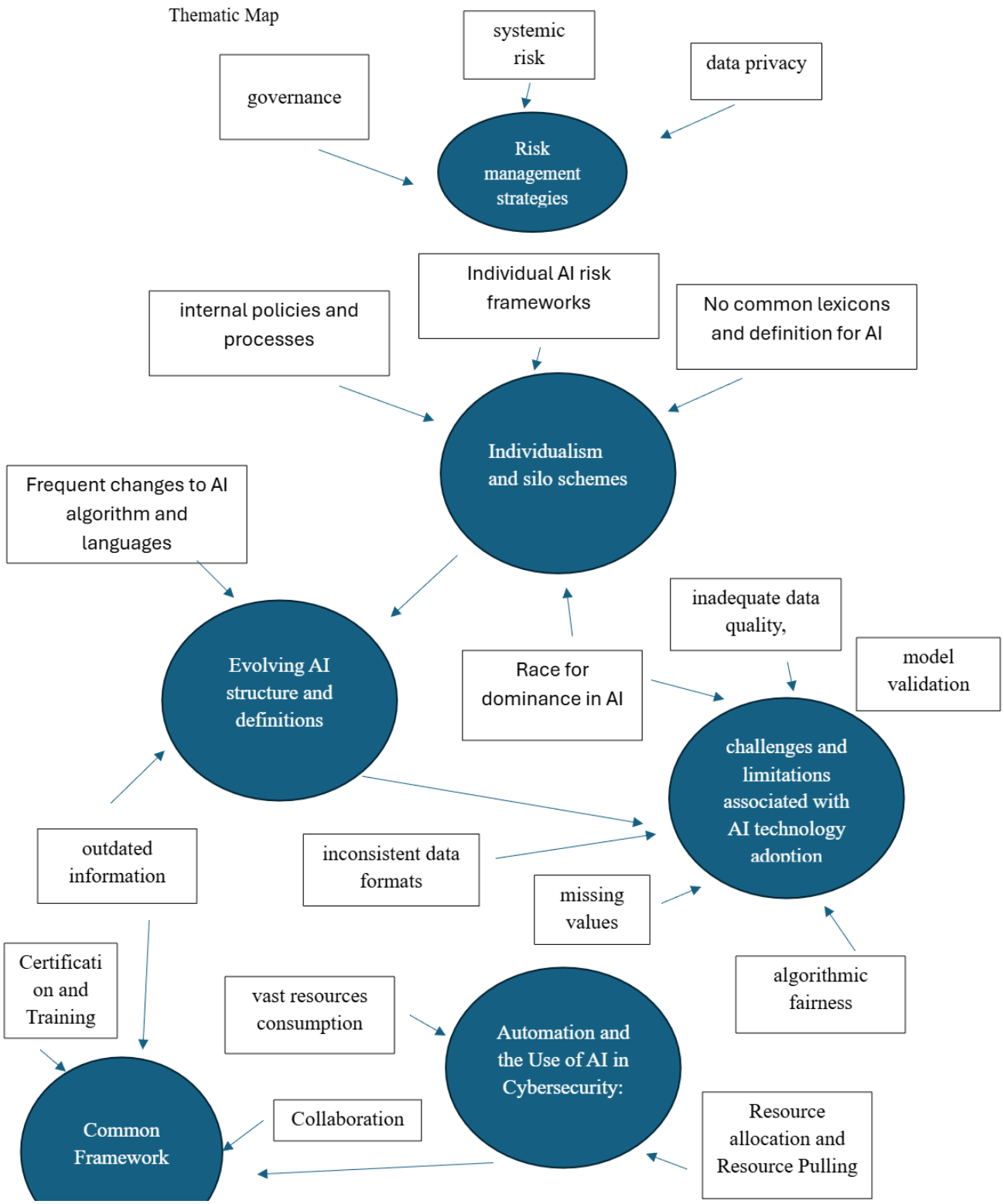
Despite the rigorous approach, certain limitations were encountered:

Table 5: Limitations

| Limitation | Description |
|---|---|
| Potential publication bias | Limited access to non-English or non-indexed studies. |
| Rapidly evolving field | Some findings may become outdated quickly due to changing government and political influence of AI, |
| Database restrictions | Variability in search functionalities across platforms. |

**Figure 3: Thematic Map**

**Thematic Focus and Map Diagram**



## Result

Developing a robust AI-driven cybersecurity framework allows an organization to scale its responses to emerging threats while reducing overhead without sacrificing excellent services to clients. Prasad et al. (2023) stressed in their study that an AI-driven cybersecurity framework would increase the efficiency of cyber defense, especially when using XAI, a term for the explainability of AI. While AI increases confidence for institutions to respond in real-time and proactively, many organizations are also looking at a retroactive approach to AI, focusing on driving trust among stakeholders.

Stemming from the research articles, developing a unified AI risk framework for cybersecurity would require a multifaceted, rigorous approach and be crucial to the thriving of financial institutions. The thematic focus and benefits of this unified approach include collaboration among stakeholders, innovation, holistic risk management, eradication of fragmented approaches, trust and transparency, reduced cost, and increased efficiency. To develop a unified framework, considering the various frameworks that already exist and the dominant themes from the research, the following steps can be taken:

Building a Unified AI Risk Management Framework

1. Identify overlapping AI principles across the existing risk frameworks (Souza, 2023). Through a thorough analysis of existing frameworks, we identify common objectives and guidelines, such as AI accountability, explainability, fairness, security, and transparency, and document them in detail (Kuzior, 2024; Mbah et al; 2024; Souza, 2023; Zhang et al., 2024).
2. Unified governance development: Research studies found this theme to be the topmost concern and recommendation toward a robust risk management framework for AI-driven cybersecurity (Kuzior et al., 2024; Souza, 2023; Mbah et al., 2024).
3. Develop a rigorous, consistent, repeatable, and scalable risk assessment program and make it easily accessible. This includes risk identification, risk impact, likelihood assessment, in-depth monitoring, and a detailed reporting mechanism (Habbal et al., 2024; Mbah et al., 2024; U.S. Department of the Treasury, 2024).
4. Develop inbuilt ethical and societal impact models: The framework should entail easy-to-follow details about ethical implications like bias detection mechanisms, bias measurement thresholds, bias mitigation techniques, and consumer protection rules (Habbal et al., 2024; Kuzior, 2024; Souza, 2023; Xu, et al., 2024).
5. Design adaptation flexibility for industry-specific modification: The framework should serve different financial services purposes, making it suitable for adaptation by industry stakeholders (banks, credit unions, insurance, asset management, fintech startups, and regulators) (U.S. Department of the Treasury, 2024).
6. Formulate robust security standards for AI risk management. This can be done by leveraging existing security standards and improving security where appropriate. 360 cybersecurity checks for interwoven security checks and balances. And include detailed safety protocols and guidelines (Korycki, 2020; Saleem Sultan & Shahid Sultan, 2024; Souza, 2023; Thapaliya, 2024; Xu et al., 2024).
7. Establishment of an AI-driven risk certification rating program: The program should consist of audit and reporting requirements and create certification around the unified risk framework (following ISO certification).
8. Build consensus, collaboration, and stakeholder engagement to facilitate the adoption and commitment of the unification program for risk management. This should include regulators and consumer protection advocates, through workshops and consultation, and continuous improvement approaches (Adeyeri, 2024; Kuzior, 2024; Simpson, 2024; Xu et al., 2024).

9. Invest in automation research and development for threat intelligence and collaboration (Adeyeri, 2024, Xu, et al., 2024).

## Discussion of Findings

The findings of this research underscore the limitations of traditional risk management frameworks when addressing AI-driven cybersecurity risks in financial institutions. While AI adoption improves operational efficiency, it introduces challenges like data poisoning, adversarial attacks, and algorithmic biases. These risks are compounded by financial systems' inherent complexities, which traditional frameworks struggle to address effectively. In line with previous studies (e.g., Dańıelsson et al., 2020; Mahalakshmi et al., 2021), the lack of model governance, inadequate data quality for AI development, data availability issues, missing values, inconsistent data formats, outdated information, data silos model validation, data privacy, algorithmic fairness, vast resource consumption, systemic risk, and the siloed nature of AI implementations are persistent issues. However, the findings of this study highlight a broader gap: Many organizations lack standardized policies for managing these AI-driven risks, especially small- to medium-sized enterprises. This aligns with the literature indicating that financial institutions have adopted NIST AI RMF, but face difficulties in ensuring consistent governance across diverse operational environments.

The study reveals that larger institutions, such as JPMorgan Chase and Mastercard, have developed proprietary AI-driven cybersecurity systems and management, showing that a unified framework could offer smaller institutions similar opportunities. This suggests that financial sectors could mitigate the evolving risks by addressing gaps in AI expertise and governance, particularly for smaller institutions. This adds nuance to Remolina's (2022) findings that AI can either enhance or compromise financial stability, depending on how well risk management frameworks are developed and applied**.**

### Implications of Findings

The findings of this study have both theoretical and practical implications for AI-driven cybersecurity in the financial sector. Theoretically, the research advances our understanding of how AI-driven solutions introduce new cybersecurity risks that traditional frameworks cannot sufficiently address. It emphasizes the need for integrating AI-specific risk management models to reduce vulnerabilities and increase cooperation among financial services for a better understanding of their common threats. This indicates a necessity for further exploration of AI trustworthiness, governance, and adaptability across various organizational sizes, echoing calls from prior studies like Souza (2023) for a unified AI risk management framework. Practically, the development of a unified risk management framework has several implications for policy and industry. Financial regulators could benefit from standardized guidelines, which would reduce disparities between large and small institutions. For industry, the framework could enhance trust among stakeholders, improve threat and fraud detection, and reduce operational costs by streamlining AI governance and compliance. Moreover, the findings suggest that such frameworks could be adapted to other industries, such as healthcare, where AI-driven cybersecurity also emerges as a critical need.

## Conclusion

The study demonstrates the urgent need for a unified risk management framework to address the unique challenges posed by AI-driven cybersecurity in financial services. While AI offers significant operational benefits, its deployment introduces complex risks that existing frameworks cannot manage effectively. This research contributes to the field by identifying gaps in current AI risk management strategies, particularly the lack of standardized governance, trust, and coordination among financial institutions, as well as the resource expenditure by individual organizations.

By linking the findings to the study's objectives, it can be implied that unifying risk management frameworks can enhance both security and efficiency. This unification is essential to ensure that financial institutions of all sizes can adopt AI technologies safely, thereby promoting greater trust and

stability in the industry. Future research should focus on developing real-world case studies and quantitative assessments to further validate the framework proposed in this study, ensuring that it remains adaptable to the rapidly evolving landscape of AI technologies.

### Limitations

This study has several limitations, including its reliance on English-language, peer-reviewed sources, which may introduce publication bias and limit the global applicability of its findings, particularly in regions where AI-driven cybersecurity is advancing but not well documented. The rapid evolution of AI technology also challenges this research, as the frameworks and regulatory standards discussed are likely to become outdated quickly, limiting the study's long-term relevance. The lack of real-world case studies further restricts the practical applicability of the proposed unified risk management framework, as theoretical models have not been tested in operational financial institutions. Finally, the study does not fully address the disparity in AI adoption between large and small institutions, underscoring the need for future research to develop tailored solutions for organizations with varying resources and capabilities. To ensure future research remains relevant and aligns with ongoing advancements in artificial intelligence and cybersecurity, a more inclusive, case-driven, and iterative approach should be adopted.

## Recommendations

Future research should explore the evolving regulatory landscape of AI in financial services (Souza, 2023). The EU AI Act and NIST AI RMF are examples of emerging regulatory frameworks that aim to standardize AI governance. Pandey et al. (2022) emphasize the need for harmonizing AI policies across different markets to ensure consistency and security. Investigating how these frameworks can be universally adopted or adapted to fit regional variations will be crucial for enhancing AI risk management on a global scale. This exploration should also include the development of ethical and bias mitigation strategies in AI models, as discussed by Kuzior (2024), to ensure fairness, transparency, and trustworthiness in AI-driven financial services.

Lastly, real-world case studies are essential for bridging the gap between theoretical frameworks and practical implementation. Uzoka et al. (2024) suggest that detailed case studies provide a clearer understanding of how AI-driven cybersecurity frameworks are applied in real-world financial institutions. Practical insights from organizations like JPMorgan Chase and Mastercard, which have successfully integrated AI cybersecurity solutions, can provide valuable lessons for the broader industry. By focusing on case studies, future research could identify the best practices and potential pitfalls in implementing AI risk management systems while offering actionable recommendations for other institutions aiming to follow suit.

## References

Adams, J., & Hagras, H. (2020). A type-2 fuzzy logic approach to explainable AI for regulatory compliance, fair customer outcomes, and market stability in the global financial sector. *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1-8.

Adeyeri, T. B. (2024). Economic impacts of AI-driven automation in financial services. *International Journal of Scientific Research and Management (IJSRM)*, *12*(07), 6779–6791. https://doi.org/10.18535/ijsrm/v12i07.em07

Ahmed, N. K., Bryans, J., Sabaliauskaite, G., & Jadidbonab, H. (2023). Integrated attack tree in residual

risk management framework. *Information*, 14(12), 639. https://doi.org/10.3390/info14120639

*AI Risk Management Framework*. (n.d.). Palo Alto Networks. https://www.paloaltonetworks.com/cyberpedia/ai-risk-management-framework

Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. (2024). https://doi.org/10.6028/nist.ai.600-1

Arun, R (2020). Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science 48, 137 – 141.

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

Culver III, David (2024). A systematic literature review: African American views of artificial intelligence and machine learning in healthcare. Article Published in the MGA Graduate Repository.

Dańıelsson, J., Macrae, R., & Uthemann, A. (2020). Artificial intelligence and systemic risk. *Artificial Intelligence - Law*.

Das, R., & Sandhane, R. (2021). Artificial intelligence in cyber security. *Journal of Physics: Conference Series, 1964*.

Deduce, (2023). Financial jeopardy: Companies losing fight against synthetic fraud. White paper of deduce.com. Wakefield Research White Paper for Axiom

Dhir, A., Talwar, S., Kaur, P., & Malibari, A.A. (2020). Food waste in hospitality and food services: A systematic literature review and framework development approach. *Journal of Cleaner Production, 270*, 122861. https://doi.org/10.1016/j.jclepro.2020.122861

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, abs/2408.07933*.

El Amin, H., Oueidat, L., Chamoun, M., Samhat, A. E., & Feghali, A. (2024). Blockchain-based multi-organizational cyber risk management framework for collaborative environments. *International journal of information security*, 23(2), 1231-1249. https://doi.org/10.1007/s10207-023-00788-7

Fabozzi, F. J., & Drake, P. P. (2010). The basics of finance: An introduction to financial markets, business finance, and portfolio management. http://182.160.97.198:8080/xmlui/handle/123456789/560

Familoni, B.T. (2024). cybersecurity challenges in the age of AI: Theoretical approaches and practical solutions. Computer science & it research journal.

FinCEN. (2021). *Identity-related suspicious activity: 2021 Threats and trends* [Report]. https://www.fincen.gov/sites/default/files/shared/FTA_Identity_Final508.pdf

Haakman, M., Cruz, L., Huijgens, H., & van Deursen, A. (2020). AI lifecycle models need to be revised. *Empirical software engineering, 26*.

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 Syst. Appl., 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, with interactivity for optimised digital transparency and Open Synthesis Campbell Systematic Reviews, 18, e1230. https://doi.org/10.1002/cl2.1230

Hoffmann, C. H. (2020). Unpacking the black box of systemic risks in banking: how causal loop modeling helps overcome rigid risk sharing and categorization. *Kybernetes*, 49(6), 1675-1690. https://doi.org/10.1108/K-05-2019-0314

Julia, J., Iyawa, G., & Gamundani, A.M. (2023). Essential components of an IT risk management framework for the financial services industry: A review. *SSRN Electronic Journal*.

Korycki, L., & Krawczyk, B. (2020). Adversarial concept drift detection under poisoning attacks for robust data stream mining. *Machine learning*, 1 - 36.

Krnic Martinic, M., Pieper, D., Glatt, A. & Puljak, L *(2019).* Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. *BMC medical research methodology* 19 (1), 203. https://doi.org/10.1186/s12874-019-0855-0

Kumari, B., Kaur, J., & Swami, S. (2022). Adoption of artificial intelligence in financial services: A policy framework. *Journal of science and technology policy management*.

Kure, H. I., & Islam, S. (2019). Assets focus risk management framework for critical infrastructure cybersecurity risk management. *IET cyber-physical systems: Theory & applications*, 4(4), 332-340. https://doi.org/10.1049/iet-cps.2018.5079

Kurshan, E., Shen, H., & Chen, J. (2020). Towards self-regulating AI: challenges and opportunities of AI model governance in financial services. *Proceedings of the first ACM international conference on ai in finance.*

Kuzior, A. , Koldovsky, A., Rekunenco, I. (2024). Optimizing financial market stability through ai-based risk management. Materials research proceedings. 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

Library, U. (no date) *Research Guides:* Systematic reviews: planning your systematic review, *guides.library.ucla.edu*. Available at:

https://guides.library.ucla.edu/c.php?g=224129&p=6469506.

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*.

Mastercard. (2022, June 28). Mastercard account intelligence for issuers: Fraud & risk [Video]. YouTube. https://www.youtube.com/watch?v=2m8eB8Po0OQ

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*.

NIST AI Risk Management Framework (AI RMF). (n.d.). Palo Alto Networks. https://www.paloaltonetworks.com/cyberpedia/nist-ai-risk-management-framework

Pandey, M.K., & Sergeeva, I. (2022). Artificial intelligence impact evaluation: Transforming paradigms in financial institutions. *World of Economics and Management*.

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). In *NIST AI 100-1*. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Regulatory Notice 24-09 | FINRA.org. (2024, June 27). https://www.finra.org/rules-guidance/notices/24-09

Remolina, N. (2022). Interconnectedness and financial stability in the era of artificial intelligence. *SSRN Electronic Journal*.

Richard Arogundade, O. (2023). Strategic security risk management in cloud computing: a comprehensive examination and application of the risk management framework. *IARJSET*.

Saleem Sultan, M., & Shahid Sultan, M. (2024). Leveraging artificial intelligence for enhanced cybersecurity: A systematic approach [Journal-article]. *International Journal of Science and Research (IJSR)*, *13*(8), 832. https://www.ijsr.net/archive/v13i8/SR24812100704.pdf

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

Singh, B.J., Chakraborty, A., & Sehgal, R. (2023). A systematic review of industrial wastewater management: Evaluating challenges and enablers. *Journal of environmental management, 348*, 119230. https://doi.org/10.1016/j.jenvman.2023.119230

Singhal, D.A., & Arora, D.B. (2022). Digital innovations in financial services: challenges & opportunities. *YMER Digital*.

Suresh, N., Neelam, H.S., Chakrapani, E., Kumar, K.A., & Ali, S.S. (2023). Artificial intelligence advances and their repercussions on the financial system. *2023 International Conference on*

*Computer Communication and Informatics (ICCCI)*, 1-4. Artificial Intelligence Advances and Their Repercussions on the Financial System | IEEE Conference Publication | IEEE Xplore

Thapaliya, S., & Bokani, A. (2024). Leveraging artificial intelligence for enhanced cybersecurity: insights and innovations. *SADGAMAYA*. https://doi.org/10.3126/sadgamaya.v1i1.66888

Tjondronegoro, D., Yuwono, E.I., Richards, B., Green, D., & Hatakka, S. (2022). Responsible AI implementation: a human-centered framework for accelerating the innovation process. *ArXiv, abs/2209.07076*. https://doi.org/10.48550/arXiv.2209.07076

U.S. Department of the Treasury (2024). Managing artificial intelligence-specific cybersecurity risks in the financial services sector. Available at: https://home.treasury.gov/system/files/136/Managing-Artificial-Intelligence-Specific-Cybersecurity-Risks-In-The-Financial-Services-Sector.pdf.

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*.

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/

Xu, H., Niu, K., Lu, T., & Li, S. (2024). Leveraging artificial intelligence for enhanced risk management in financial services: Current applications and future prospects. *Engineering Science & Technology Journal*. https://doi.org/10.5281/zenodo.13765819

Zhang, X., Chan, F.T., Yan, C., & Bose, I. (2022). Towards risk-aware artificial intelligence and machine learning systems: an overview. *Decis. Support Syst., 159*, 113800. https://doi.org/10.1016/j.dss.2022.113800