Towards a National AI Security Framework for Financial Infrastructure Protection

Rahul Mehta[1], Neal Patwar[2], Xiangang Wei[3], Emily Saunders[4], Xu Zhu[5], Jingwei Liu[6]

[1]Engineering Science, University of Oxford, United Kingdom
[2]University of Utah, USA
[3]Management Science and Engineering, Xi'an University of Architecture and Technology, Shaanxi, China
[4]Artificial Intelligence and Machine Learning, University of Cambridge, United Kingdom
[5]Raffles University, Malaysia
[6]New York University, USA
Correspondence Author

Abstract: ***Artificial intelligence (AI) is increasingly embedded in modern financial infrastructure, including real-time payment systems, anti–money laundering (AML) platforms, credit risk assessment engines, and cross-border settlement networks. As financial institutions accelerate digital transformation, AI-driven decision-making has become essential for ensuring operational efficiency, fraud detection, and systemic stability. However, the rapid adoption of AI technologies within critical financial systems has simultaneously introduced a new class of security vulnerabilities. Emerging threats such as adversarial machine learning, model poisoning, data manipulation, automated fraud orchestration, and AI-enabled cyberattacks pose significant risks to the integrity and resilience of financial infrastructure. These risks extend beyond individual institutions, creating the potential for cascading failures across interconnected financial ecosystems and posing systemic challenges to national economic stability [1]. Despite the growing importance of AI security in financial operations, there is currently no unified national-level framework that comprehensively addresses AI-related threats within financial infrastructure. Existing cybersecurity standards and AI governance models often operate in isolation, lacking integrated mechanisms that align technical safeguards, operational monitoring, and regulatory coordination across institutions. This fragmentation limits the ability of financial systems to respond effectively to sophisticated AI-driven threats and undermines the development of a consistent [2], trustworthy AI security posture at scale. The absence of a standardized national architecture further complicates collaboration among banks, regulators, and technology providers, creating gaps in visibility, accountability, and coordinated defense [3]. To address these challenges, this paper proposes a National AI Security Framework for Financial Infrastructure Protection. The framework introduces a multi-layered architecture designed to secure AI-enabled financial systems through integrated model protection, data integrity assurance, adversarial defense mechanisms, real-time monitoring, and cross-institutional coordination [4]. At the technical level, the framework incorporates secure model development practices, privacy-preserving data sharing mechanisms, adversarial attack detection, explainable AI monitoring, and continuous risk assessment. At the governance level, it establishes mechanisms for inter-organizational collaboration, regulatory integration, and shared threat intelligence to support consistent security practices across financial ecosystems. By aligning technical safeguards with broader national risk management strategies, the framework enables financial institutions to proactively identify vulnerabilities, mitigate emerging threats, and maintain operational continuity in increasingly complex AI-driven environments [5]. The proposed framework also emphasizes alignment with national critical infrastructure protection priorities and trustworthy AI initiatives, providing a structured approach for integrating AI security into existing financial risk management and regulatory oversight processes. Through illustrative use cases in real-time payment networks, fraud detection systems, and cross-border financial transactions, this study demonstrates how a coordinated AI security architecture can enhance resilience, transparency, and trust across financial systems [6]. The framework is designed to be scalable and adaptable, supporting both individual institutional deployment and broader national-level coordination [7]. By establishing a unified and technically grounded approach to AI security, this research contributes to the development of a secure and resilient financial ecosystem capable of withstanding evolving technological and adversarial challenges. This study contributes toward building a resilient and trustworthy AI foundation for protecting national financial infrastructure [8].***

Keywords: Artificial Intelligence Security; Financial Infrastructure Protection; Critical Infrastructure Security; Adversarial Machine Learning; Trustworthy AI; Financial Risk Management; Federated Learning Security; Payment System Security; Cyber-Physical Financial Systems; National AI Security Framework

1. INTRODUCTION

1.1 Background

Artificial intelligence (AI) has rapidly become a foundational technology within modern financial systems, transforming the way institutions manage risk, process transactions, and ensure regulatory compliance [9]. AI-driven capabilities are now deeply embedded in real-time payment networks, anti–money laundering (AML)

monitoring platforms, fraud detection engines, credit assessment tools, and clearing and settlement systems. In payment infrastructures, machine learning models enable instant risk scoring and anomaly detection for high-volume transaction streams [10]. Within AML and compliance environments, AI is used to identify suspicious behavior, detect hidden transaction patterns, and automate regulatory reporting processes. Clearing and settlement systems increasingly rely on predictive analytics and intelligent automation to maintain operational continuity, manage liquidity risk, and support cross-border financial transactions. As a result, AI is no longer an auxiliary tool but a core operational component that underpins the reliability and efficiency of financial services [11].

With the growing dependence on digital technologies and interconnected platforms, financial systems are now widely recognized as critical infrastructure essential to national stability and economic security [12]. Payment systems, interbank networks, and financial data exchanges form the backbone of modern economies, enabling commerce, investment, and public-sector operations [13]. Disruptions to these systems—whether caused by cyberattacks, operational failures, or emerging AI-related threats—can produce cascading effects across markets and institutions. Consequently, the protection of financial infrastructure has become a strategic priority for governments and regulatory bodies worldwide. In this context, the integration of AI into core financial operations introduces both transformative opportunities and unprecedented security challenges [14]. Vulnerabilities in AI models, training data, and decision-making processes can potentially compromise transaction integrity, expose sensitive financial information, and undermine trust in financial institutions [15].

As AI assumes a central role in critical financial operations, ensuring the security, robustness, and trustworthiness of AI systems has become inseparable from national security considerations [16]. AI security is no longer solely a technical concern confined to individual organizations; it is increasingly viewed as a matter of national resilience and systemic risk management. Malicious manipulation of AI-driven financial systems, large-scale automated fraud campaigns, or coordinated attacks on intelligent payment infrastructures could have significant economic and societal consequences. Therefore, safeguarding AI-enabled financial infrastructure requires coordinated strategies that extend beyond institutional boundaries and incorporate national-level risk management frameworks [17].

1.2 The rapid integration of artificial intelligence (AI) into financial systems has introduced significant security, operational, and systemic challenges that extend beyond individual institutions. While AI technologies enable real-time decision-making, fraud detection, and automated compliance, their increasing adoption also expands the attack surface and introduces new vulnerabilities that can be exploited at scale. Despite growing awareness of these risks, the financial sector lacks a unified and comprehensive approach to securing AI-driven infrastructure. This section outlines three critical problem domains that motivate the need for a national-level AI security framework for financial infrastructure protection [18].

(1) AI-Driven Threats to Financial Systems

The deployment of AI across payment processing, fraud detection, and compliance systems has created new avenues for sophisticated cyber and operational attacks. Adversaries can exploit vulnerabilities in machine learning models, training data, and inference processes to manipulate financial decision-making and disrupt system integrity. One major category of threats involves adversarial attacks, in which carefully crafted inputs are designed to mislead AI models and cause incorrect classifications or risk assessments. In financial contexts, such attacks can enable fraudulent transactions to bypass detection systems or trigger false positives that disrupt legitimate activities [19].

Model poisoning represents another critical threat vector. In collaborative or continuously trained systems, malicious actors may introduce manipulated or contaminated data into training pipelines, thereby degrading model performance or embedding hidden backdoors. Poisoned models can produce biased or incorrect outputs that remain undetected until significant financial damage has occurred. Additionally, the rise of AI-driven fraud has significantly increased the scale and automation of financial crimes. Attackers increasingly leverage AI tools to generate synthetic identities, automate phishing campaigns, and conduct coordinated transaction fraud at speeds and volumes that exceed traditional detection capabilities. These emerging threats highlight the urgent need for robust AI security mechanisms specifically tailored to financial environments [20].

(2) Lack of a Unified AI Security Architecture

Despite the shared risks posed by AI-enabled threats, financial institutions currently rely on fragmented and institution-specific approaches to AI risk management and cybersecurity. Banks and financial service providers typically develop their own internal risk control mechanisms, resulting in inconsistent security practices and limited interoperability. Existing cybersecurity frameworks often focus on network and data protection but do not adequately address the unique risks associated with AI models, data pipelines, and automated decision systems.

Moreover, there is no standardized national framework that defines best practices, technical safeguards, and governance mechanisms for securing AI within financial infrastructure. The absence of a unified architecture complicates cross-institutional collaboration, limits information sharing, and hinders the development of coordinated defense strategies [21]. Without common standards and shared security models, financial systems remain vulnerable to systemic exploitation and uneven protection across institutions.

(3) Systemic Risks in AI-Enabled Financial Infrastructure

The integration of AI into critical financial operations introduces systemic risks that can affect entire financial ecosystems. Real-time payment systems, which rely on instant processing and automated risk evaluation, are particularly sensitive to AI-related failures or manipulations. Errors or attacks affecting AI models in these systems could disrupt transaction processing, delay settlements, or allow fraudulent transfers to propagate rapidly across networks. Cross-border clearing and settlement systems also depend increasingly on AI-driven analytics and automation to manage liquidity, detect anomalies, and ensure compliance with international regulations. Vulnerabilities in these systems could lead to cross-jurisdictional disruptions and financial instability [22].

Furthermore, anti–money laundering (AML) operations now heavily depend on AI-based monitoring and pattern recognition to identify suspicious activities and comply with regulatory requirements. While AI enhances detection capabilities, overreliance on automated systems without robust security and governance measures introduces the risk of manipulation, evasion, or large-scale operational failures. Given the interconnected nature of modern financial systems, disruptions in one AI-enabled component can rapidly propagate across institutions and markets, amplifying systemic risk [23].

Collectively, these challenges demonstrate the urgent need for a comprehensive and coordinated approach to securing AI within financial infrastructure [24]. Addressing AI-driven threats, architectural fragmentation, and systemic vulnerabilities requires the development of a unified framework capable of protecting financial systems at both institutional and national levels.

1.3 Research Objective

The objective of this research is to develop a comprehensive and structured approach to securing artificial intelligence within modern financial infrastructure. As AI technologies become deeply embedded in payment systems, anti–money laundering platforms, fraud detection engines, and clearing and settlement networks, the security and reliability of these intelligent systems have become essential to maintaining operational continuity and financial stability. However, the rapid adoption of AI has outpaced the development of unified security standards and coordinated defense mechanisms capable of addressing emerging AI-driven threats across financial ecosystems. This gap highlights the need for a systematic and scalable framework that can guide financial institutions and regulatory bodies in implementing secure, trustworthy, and resilient AI systems [25].

This paper proposes a national AI security framework designed to protect financial infrastructure from emerging AI-driven threats. The framework aims to provide a unified architecture that integrates technical safeguards, operational monitoring mechanisms, and governance-level coordination to address vulnerabilities associated with AI deployment in critical financial environments [26]. By establishing a multi-layered security model, the proposed framework seeks to enhance the protection of AI models, data pipelines, and decision-making processes against adversarial manipulation, model poisoning, automated fraud, and other sophisticated attack vectors.

By establishing a unified and technically grounded approach to AI security in financial infrastructure, this study aims to support the development of consistent protection strategies and foster greater collaboration among stakeholders responsible for safeguarding financial systems. The proposed framework is intended to serve as a foundational reference for future research, industry implementation, and policy development related to secure and trustworthy AI deployment in critical financial environments [26].

1.4 Contributions

This paper makes the following key contributions to the study of artificial intelligence security in financial infrastructure:

Proposal of a national-level AI security framework for financial infrastructure.

This study introduces a comprehensive national AI security framework specifically designed to protect AI-enabled financial systems from emerging threats. The proposed framework addresses security challenges across payment networks, anti–money laundering systems, and clearing and settlement infrastructures, offering a unified approach to safeguarding critical financial operations in the era of intelligent automation [27].

Design of a multi-layered security architecture for AI-driven financial systems.

The paper develops a structured, layered security architecture that integrates AI model protection, data integrity assurance, adversarial defense mechanisms, real-time monitoring, and governance-level coordination. This architecture provides a systematic approach for mitigating vulnerabilities across the entire AI lifecycle, from data ingestion and model training to deployment and operational monitoring.

Alignment with national AI and financial security strategies.

The proposed framework is designed to align with major U.S. initiatives related to trustworthy AI, critical infrastructure protection, and financial system resilience. By incorporating principles consistent with national AI risk management and cybersecurity strategies, the framework supports coordinated efforts to enhance the security, reliability, and accountability of AI systems deployed within financial infrastructure [28].

Demonstration through practical financial use case scenarios.

The study presents representative application scenarios, including real-time payment risk monitoring, AI-driven fraud detection and anti–money laundering operations, and cross-border transaction security. These use cases illustrate how the proposed framework can be implemented in real-world financial environments to improve resilience, strengthen risk management capabilities, and support secure collaboration across financial institutions.

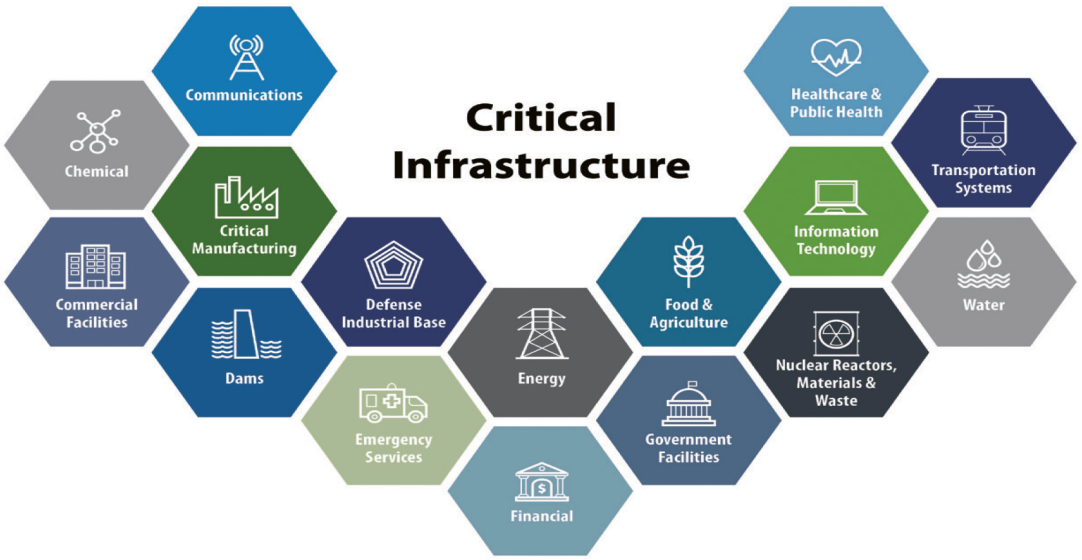
2. RELATED WORK

2.1 AI in Financial Infrastructure

Artificial intelligence has become a core enabling technology in modern financial infrastructure, supporting a wide range of operational and risk management functions. Financial institutions increasingly rely on machine learning and advanced analytics to enhance payment processing efficiency, strengthen fraud detection capabilities, and improve regulatory compliance. In payment systems, AI-driven models are widely deployed for real-time transaction monitoring, anomaly detection, and dynamic risk scoring. These systems enable financial institutions to process large transaction volumes while identifying suspicious behavior with high accuracy and speed [29].

In the domain of risk management and credit assessment, AI techniques are used to evaluate borrower profiles, forecast default risks, and optimize portfolio performance. Anti–money laundering (AML) operations have also undergone significant transformation through the adoption of AI-based monitoring systems capable of detecting complex transaction patterns and hidden relationships across accounts and institutions. By leveraging machine learning, financial organizations can improve detection rates and reduce false positives compared to traditional rule-based approaches [30].

Clearing and settlement systems increasingly incorporate predictive analytics and intelligent automation to manage liquidity, detect anomalies, and support cross-border transactions. As financial ecosystems continue to evolve toward real-time processing and digital platforms, AI has become indispensable in maintaining operational efficiency and ensuring compliance with regulatory requirements [31]. However, the growing reliance on AI also introduces new risks related to model integrity, data quality, and automated decision-making, highlighting the need for robust security mechanisms tailored to AI-driven financial environments.

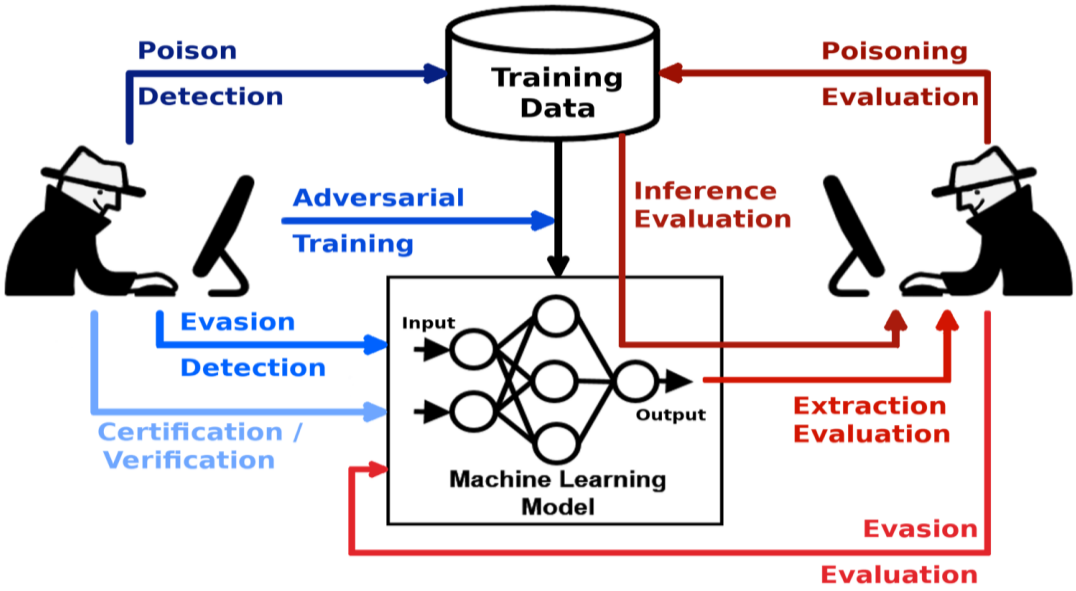


© US Cybersecurity & Infrastructure Security Agency (CISA)

2.2 AI Security and Adversarial Machine Learning

The rapid expansion of AI applications has been accompanied by increasing concerns regarding the security and robustness of machine learning systems. Research in adversarial machine learning has demonstrated that AI models are vulnerable to a variety of attacks that can manipulate their behavior and degrade their reliability. Adversarial attacks involve crafting carefully designed inputs that cause models to produce incorrect outputs while appearing normal to human observers. In financial systems, such attacks may enable fraudulent transactions to evade detection or trigger incorrect risk classifications.

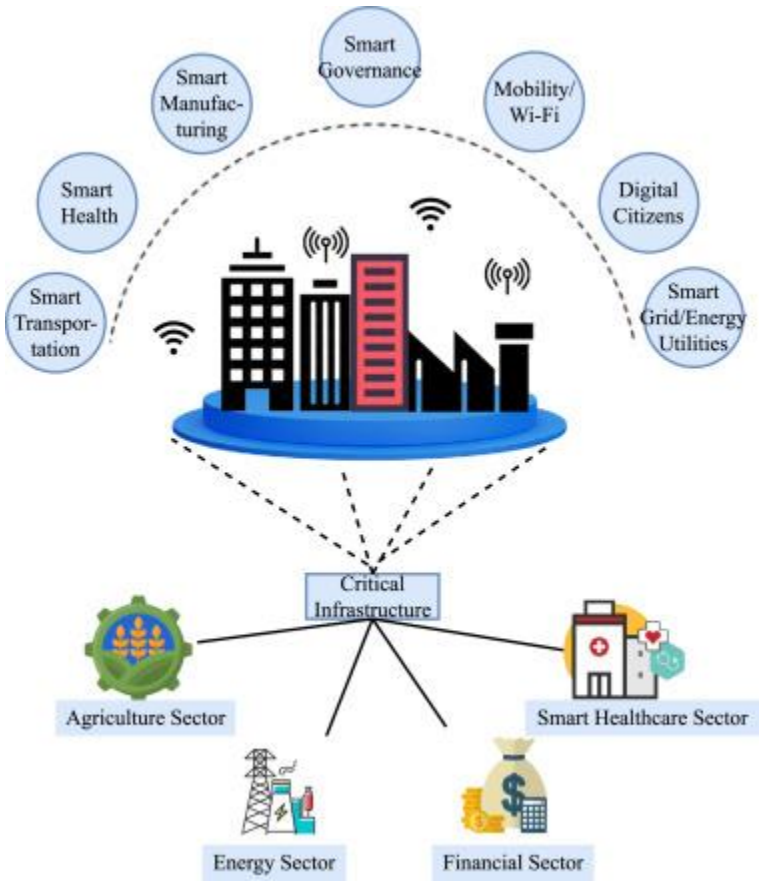
Recent research has also examined the use of AI by malicious actors to automate fraud, generate synthetic identities, and conduct coordinated cyberattacks. These developments have intensified the need for robust, trustworthy, and resilient AI systems capable of withstanding adversarial manipulation. While significant progress has been made in developing defensive techniques such as adversarial training, anomaly detection, and explainable AI, the application of these methods within large-scale financial infrastructure remains limited and fragmented [32].



2.3Critical Infrastructure Protection

Financial services are widely recognized as a critical infrastructure sector essential to national economic stability and public trust. Governments and regulatory agencies have introduced frameworks and guidelines to strengthen cybersecurity and risk management across critical infrastructure domains. The Cybersecurity and Infrastructure Security Agency (CISA) has identified the financial services sector as a priority area for advanced cybersecurity protections and coordinated threat response. CISA's guidance emphasizes resilience, risk management, and cross-sector collaboration to mitigate evolving cyber threats.

Despite these efforts, existing frameworks primarily address general cybersecurity or AI governance challenges and do not provide detailed technical architectures specifically tailored to AI-driven financial systems. Financial institutions often adapt these guidelines independently, leading to variations in implementation and limited coordination across the sector [33].



2.4 Research Gap

Although substantial research has been conducted on AI applications in finance, adversarial machine learning, and critical infrastructure protection, there remains a significant gap in the development of a unified and comprehensive AI security framework tailored to financial infrastructure. Current approaches are typically fragmented, focusing either on institutional cybersecurity practices or isolated AI risk management techniques without providing an integrated architecture capable of addressing systemic and cross-institutional risks.

In particular, there is no unified national AI security framework for finance that systematically integrates technical safeguards, operational monitoring, and governance-level coordination across financial institutions and regulatory bodies. The absence of such a framework limits the ability of financial systems to respond effectively to emerging AI-driven threats and undermines efforts to establish consistent and trustworthy AI security practices at scale [34].

This gap underscores the necessity of developing a structured and nationally coordinated AI security framework designed specifically for financial infrastructure. The framework proposed in this study seeks to address this need by providing a comprehensive architecture that aligns technical protection mechanisms with broader risk management and critical infrastructure protection objectives.

3. THREAT LANDSCAPE FOR AI-DRIVEN FINANCIAL SYSTEMS

3.1 AI Attack Surface

The adoption of artificial intelligence in financial systems introduces new security risks unique to AI-driven operations. These threats target models, data pipelines, and automated decision-making, requiring specialized mitigation strategies.

Model Poisoning: Malicious actors can inject corrupted data into training pipelines, causing AI models to make biased or incorrect predictions. In finance, this can result in fraudulent transactions being misclassified or credit assessments being unreliable.

Prompt Injection: Adversaries can manipulate AI inputs to produce unintended outputs, potentially bypassing compliance checks or generating unsafe financial instructions.

Data Manipulation: Altering transaction records or training data can mislead AI models, affecting risk assessment, fraud detection, and AML reporting.

Synthetic Identity Fraud: AI can be misused to generate synthetic identities that evade verification systems, allowing fraudulent accounts or transactions at scale.

These vectors highlight the expanded attack surface in AI-driven financial infrastructure, emphasizing the need for robust model security, data integrity, and monitoring mechanisms.

3.2 Systemic Risk to Payment Networks

AI integration in financial infrastructure not only introduces direct threats but also amplifies systemic risks that can affect entire payment networks.

Real-Time Payment Failure: Failures in AI-driven risk assessment or anomaly detection can disrupt real-time payment processing, causing delays, transaction errors, or service outages.

Fraud Automation: AI tools used maliciously can automate fraudulent transactions at high volume and speed, overwhelming detection systems and increasing financial losses.

Cross-Border Risk: AI vulnerabilities in international payment and settlement systems can propagate across jurisdictions, creating cascading effects that compromise global transaction integrity and financial stability.

These risks demonstrate that AI threats extend beyond individual institutions, potentially impacting the resilience and reliability of national and international financial systems.

3.3 National-Level Implications

The disruption of AI-enabled financial infrastructure poses significant national security concerns. Critical financial systems—including payment networks, clearinghouses, and anti–money laundering platforms—are deeply interconnected and form the backbone of the national economy. Failures or attacks targeting these systems can lead to widespread financial instability, loss of public trust, and cascading operational disruptions across multiple sectors.

Given the strategic importance of financial services, vulnerabilities in AI-driven processes not only threaten individual institutions but also create systemic risks that can impact national economic security. Ensuring the resilience and reliability of these systems is therefore essential for safeguarding the nation’s financial stability and supporting broader national security objectives.

4. PROPOSED NATIONAL AI SECURITY FRAMEWORK

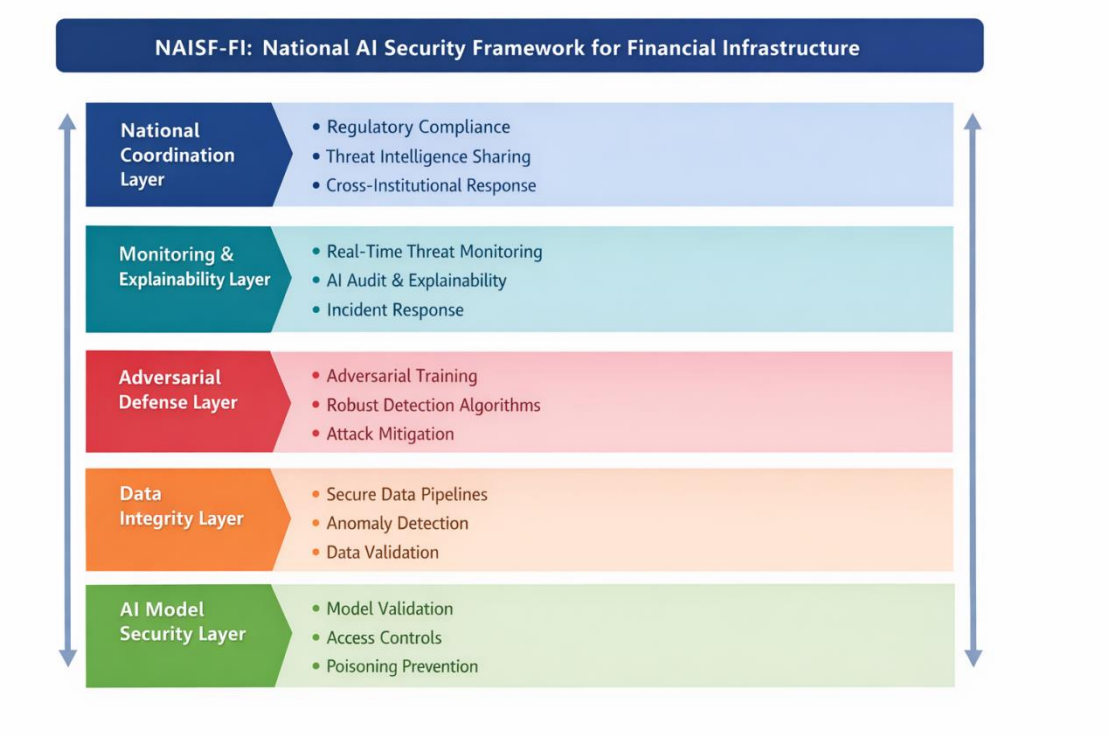
The core contribution of this paper is the design of a comprehensive security framework for AI-enabled financial

infrastructure, referred to as NAISF-FI (National AI Security Framework for Financial Infrastructure). This framework provides a structured, multi-layered approach to protecting AI-driven financial systems against emerging threats while ensuring alignment with national security and regulatory priorities.

NAISF-FI integrates technical safeguards, operational monitoring, and governance mechanisms across institutions to address vulnerabilities at every stage of the AI lifecycle. Its design emphasizes scalability, interoperability, and resilience, enabling coordinated defense across banks, payment networks, and regulatory agencies. By combining model-level protection, data integrity, adversarial defense, real-time monitoring, and national-level coordination, NAISF-FI establishes a unified architecture that can guide financial institutions and policymakers in securing AI-enabled infrastructure effectively.

The framework is structured into distinct layers, each addressing specific dimensions of AI security while collectively forming a comprehensive national-level defense system. The layered architecture ensures that threats are mitigated systematically—from model training and deployment to cross-institutional operations—providing both operational robustness and regulatory compliance.

4.1 Framework Overview



4.2 AI Model Security Layer

The AI Model Security Layer focuses on protecting AI models throughout their lifecycle, from training to deployment, ensuring their reliability and resistance to malicious manipulation. Key components include:

Secure Training: Implementing controlled and monitored model training processes to prevent injection of malicious or biased data.

Model Validation: Systematic evaluation of model performance and behavior to ensure accuracy, fairness, and compliance with security standards.

Adversarial Robustness: Incorporating defensive techniques such as adversarial training, robust optimization, and input sanitization to protect models from adversarial attacks.

Model Audit: Periodic auditing of model decisions and training data lineage to detect anomalies, backdoors, or unauthorized modifications.

These measures collectively safeguard AI models from manipulation, maintain operational integrity, and ensure trustworthiness within critical financial systems.

4.3 Data & Privacy Protection

The Data & Privacy Protection Layer ensures that AI-driven financial systems maintain data confidentiality, integrity, and compliance with privacy regulations. Core mechanisms include:

Federated Learning: Enabling multiple institutions to collaboratively train AI models without sharing raw data, reducing exposure of sensitive information.

Encrypted Data Collaboration: Applying encryption and secure computation techniques to allow joint analytics while protecting underlying datasets.

Privacy-Preserving Analytics: Leveraging differential privacy, homomorphic encryption, and secure multiparty computation to analyze sensitive financial data without compromising individual privacy.

Together, these strategies provide a strong foundation for secure data handling, regulatory compliance, and privacy-aware AI operations across financial institutions.

4.4 Real-Time Risk Monitoring

The Monitoring & Explainability Layer provides continuous oversight of AI-driven financial systems, ensuring operational transparency, early threat detection, and actionable insights for regulators and institutions. Key functions include:

Anomaly Detection: Real-time monitoring of transactions, model outputs, and system behavior to identify unusual patterns or potential attacks.

Explainable AI for Regulators: Generating interpretable explanations of AI decisions to support compliance, auditing, and regulatory review.

Risk Scoring: Quantifying operational, financial, and systemic risks based on model outputs and detected anomalies, enabling proactive mitigation strategies.

This layer ensures that AI systems remain accountable, auditable, and resilient, supporting both institutional oversight and national-level financial security objectives.

5. EVALUATION AND DISCUSSION

5.1 Security Effectiveness

The proposed NAISF-FI framework is designed to enhance the resilience and robustness of AI-enabled financial infrastructure. Key aspects of its security effectiveness include:

Defense Capability Analysis: Each layer of the framework—model security, data integrity, adversarial defense, monitoring, and national coordination—is evaluated for its ability to prevent, detect, and mitigate AI-driven attacks such as model poisoning, prompt injection, and synthetic identity fraud.

Risk Reduction: By integrating multi-layered protections and coordinated oversight, the framework reduces operational, fraud-related, and systemic risks across payment networks, AML systems, and cross-border settlement processes.

5.2 Operational Challenges

Implementing a national AI security framework in financial infrastructure presents several practical challenges:

Deployment Costs: Establishing secure AI pipelines, monitoring systems, and inter-institutional coordination mechanisms requires significant investment in infrastructure, personnel, and training.

Data Sharing Issues: Collaborative mechanisms such as federated learning require standardized protocols and trust frameworks to enable secure data exchange without violating privacy regulations.

Standards Harmonization: Variations in institutional practices and regulatory requirements can complicate the adoption of unified AI security standards across the financial sector.

5.3 Comparison with Existing Approaches

Compared to traditional financial security architectures, NAISF-FI provides several advantages:

Focuses specifically on AI-driven threats, unlike conventional cybersecurity frameworks that primarily address network and data security.

Introduces a multi-layered, nationally coordinated architecture that aligns with regulatory bodies and cross-institutional operations, unlike isolated bank-level solutions.

Integrates explainable AI and continuous risk scoring, enhancing transparency and auditability relative to traditional rule-based risk management systems.

This evaluation demonstrates that NAISF-FI not only strengthens AI model and data security but also provides systemic resilience, operational oversight, and regulatory alignment that current approaches lack.

6. FUTURE RESEARCH DIRECTIONS

The rapidly evolving landscape of AI-driven financial systems presents ongoing challenges and opportunities for enhancing security and resilience. Future research directions for NAISF-FI include:

Enhanced Adversarial AI Defense: Developing more robust techniques to defend against sophisticated adversarial attacks, including model poisoning, prompt injection, and AI-driven fraud, to ensure long-term resilience of financial AI systems.

Digital Twin Test Environments: Leveraging digital twin simulations of financial infrastructure to evaluate system behavior under cyber and operational stress, enabling high-fidelity testing of AI security measures before deployment.

Automated Security Auditing: Implementing AI-powered auditing tools that continuously monitor models, data pipelines, and system outputs to detect anomalies, policy violations, or potential vulnerabilities in real time.

Global Financial AI Security Collaboration: Establishing frameworks for international cooperation, information sharing, and coordinated defense to address cross-border risks and ensure the security of AI-enabled global financial networks.

These directions aim to advance both the technical robustness and operational governance of AI in financial infrastructure, supporting secure, resilient, and trustworthy systems at national and international levels.

7. CONCLUSION

This paper introduces NAISF-FI (National AI Security Framework for Financial Infrastructure), a comprehensive framework designed to protect AI-enabled financial systems from emerging threats. By integrating multi-layered defenses—spanning AI model security, data integrity, adversarial protection, monitoring, and national coordination—the framework provides a unified approach to safeguarding critical financial infrastructure.

NAISF-FI enhances the security and resilience of financial systems, mitigating risks associated with real-time payments, fraud detection, anti–money laundering, and cross-border settlement operations. The framework supports the stable and trustworthy operation of future AI-driven financial systems while aligning with national

security priorities and regulatory requirements.

Furthermore, this work establishes a foundational architecture and methodology that can guide future research, policy development, and practical deployment of AI security measures in the financial sector. By providing both technical and governance-level solutions, NAISF-FI serves as a reference for building secure, resilient, and accountable AI infrastructure capable of addressing evolving financial and systemic risks.

REFERENCES

[1] Chen, H., et al. (2024). *Threat detection driven by artificial intelligence: Enhancing cybersecurity with machine learning algorithms*. Cybersecurity Innovations Conference, Nov. 2024, p. 45. https://doi.org/10.53469/wjimt.2024.07(06).09

[2] Liu, Y., et al. (2021). *Deep reinforcement learning for cybersecurity: A survey*. IEEE Communications Surveys & Tutorials, 23(2), 1022–1048.

[3] Wang, Y., et al. (2025). *AI end-to-end autonomous driving*. International Journal of Operations and Management Science Research, 8(1), Article 8. https://doi.org/10.53469/wjimt.2025.08(01).08

[4] Papernot, N., et al. (2016). *Transferability in machine learning: From phenomena to black-box attacks using adversarial samples*. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1–10).

[5] Kaelbling, L. P., et al. (1996). *Reinforcement learning: A survey*. Journal of Artificial Intelligence Research, 4, 237–285.

[6] Sutton, R. S., et al. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

[7] Liu, Y., et al. (2023). *Grasp and inspection of mechanical parts based on visual image recognition technology*. Journal of Theory and Practice of Engineering Science, 3(12), 22–28.

[8] Truong, N. B., et al. (2022). *A comprehensive survey on digital twin for future networks and emerging services*. IEEE Communications Surveys & Tutorials, 24(4), 2253–2289.

[9] Chen, W., et al. (2024). *Applying machine learning algorithm to optimize personalized education recommendation system*. Journal of Theory and Practice of Engineering Science, 4(1), 101–108.

[10] Cheng, S., et al. (2023). *Poster graphic design with your eyes: An approach to automatic textual layout design based on visual perception*. Displays, 79, 102458.

[11] Du, S., et al. (2024). *Improving science question ranking with model and retrieval-augmented generation*. In Proceedings of the 6th International Scientific and Practical Conference "Old and New Technologies of Learning Development in Modern Conditions".

[12] Hashem, I. A. T., et al. (2021). *The role of digital twin in cybersecurity: Opportunities and challenges*. Future Generation Computer Systems, 115, 453–465.

[13] Scarfone, K., et al. (2007). *Guide to intrusion detection and prevention systems (IDPS)* (NIST Special Publication 800-94). National Institute of Standards and Technology.

[14] Ferrag, M. A., et al. (2020). *Privacy-preserving schemes for adversarial machine learning in cybersecurity: A survey*. IEEE Communications Surveys & Tutorials, 22(3), 1869–1895.

[15] Lin, S., et al. (2024). *Artificial intelligence and electroencephalogram analysis: Innovative methods for optimizing anesthesia depth*. Journal of Theory and Practice in Engineering and Technology, 1(4), 1–10. https://doi.org/10.5281/zenodo.14457933

[16] Huang, L., et al. (2017). *Adversarial machine learning in cybersecurity: A tutorial*. In Proceedings of the ACM Workshop on Artificial Intelligence and Security (pp. 1–10).

[17] Wang, Z., et al. (2025). *Intelligent construction of a supply chain finance decision support system and financial benefit analysis based on deep reinforcement learning and particle swarm optimization*. International Journal of Management Science Research, 8(3), 28–41.

[18] Al-Garadi, M. A., et al. (2020). *A survey of machine and deep learning methods for cybersecurity*. IEEE Access, 8, 122512–122531.

[19] Schulman, J., et al. (2017). *Proximal policy optimization algorithms*. In Proceedings of the 34th International Conference on Machine Learning (pp. 1–12).

[20] Sangaiah, A. K., et al. (2022). *Digital twin-driven cybersecurity for critical infrastructure: A systematic review*. IEEE Transactions on Industrial Informatics, 18(5), 3512–3524.

[21] Bhuyan, M. H., et al. (2014). *Network anomaly detection: Methods, systems and tools*. IEEE Communications Surveys & Tutorials, 16(1), 303–336.

[22] Cheng, S., et al. (2024). 3D Pop-Ups: *Omnidirectional image visual saliency prediction based on crowdsourced eye-tracking data in VR*. Displays, 83, 102746. https://doi.org/10.1016/j.displa.2024.102746

[23] Tian, M., et al. (2023). *The application of artificial intelligence in medical diagnostics: A new frontier*. Academic Journal of Science and Technology, 8(2), 57–61. https://doi.org/10.54097/ajst.v8i2.14945
[24] Jordan, M. I., et al. (2015). *Machine learning: Trends, perspectives, and prospects*. Science, 349(6245), 255–260.
[25] Chu, D., et al. (2024). *Research progress and challenges in end-to-end autonomous driving*. Journal of Highway and Transportation Research and Development, 1–29.
[26] Wei, K., et al. (2024). *Strategic application of AI in network threat detection using enhanced K-means clustering*. Journal of Theory and Practice of Engineering Science, 4(2), 26–35. https://doi.org/10.53469/jtpes.2024.04(01).07
[27] Lee, R. M., et al. (2016). *Analysis of the cyber attack on the Ukrainian power grid (Report)*. SANS Industrial Control Systems.
[28] Chew, J., et al. (2025). *Artificial intelligence optimizes the accounting data integration and financial risk assessment model of the e-commerce platform*. International Journal of Management Science Research, 8(2), 7–17.
[29] Khan, M. M. R., et al. (2022). *Digital twin-enabled cyber-physical systems: A review*. IEEE Internet of Things Journal, 9(1), 45–65.
[30] Wang, Y., et al. (2025). *Research on the cross-industry application of autonomous driving technology in the field of FinTech*. International Journal of Management Science Research, 8(3), 13–27.
[31] Gardner, M. T., et al. (2014). *Using GENI for experimental evaluation of software-defined networking (SDN) resilience*. In Proceedings of the IEEE Conference on Computer Communications Workshops (pp. 391–396).
[32] Hasan, S. R., et al. (2021). *A survey on digital twin: Definitions, characteristics, applications, and design implications*. IEEE Access, 9, 32091–32112.
[33] Xu, J., et al. (2025). *Adversarial machine learning in cybersecurity: Attacks and defenses*. International Journal of Management Science Research, 8(2), 26–33.
[34] Liu, Y., et al. (2024). *Grasp and inspection of mechanical parts based on visual image recognition technology*. Journal of Theory and Practice of Engineering Science, 3(12), 22–28.