World Journal of Information Systems 1(3) 17-24
DOI: https://doi.org/10.17013/wjis.v1i3.16

Review

Security Risks of Generative AI in Financial Systems: A comprehensive review

Nikhil Gupta[1]
[1]Alumnus, IIT Mumbai, Mumbai (Maharashtra), India. Email: guptanikhil28@gmail.com

* Correspondence: guptanikhil28@gmail.com

(Article history: Received June 05, 2024; Received in revised form July 15, 2024; Accepted August 28, 2024; Available online September 15, 2024)

Abstract: The rapid evolution of generative artificial intelligence (AI) presents transformative opportunities and unprecedented security challenges for financial systems. This paper conducts a comprehensive review of the security risks associated with generative AI in finance, focusing on three prominent threats: deepfakes, synthetic identity fraud, and AI-generated phishing attacks. Deepfakes, synthetic media that convincingly replace individuals' likenesses or voices, can undermine authentication and verification processes, potentially enabling unauthorized access to accounts and fraudulent transactions. Synthetic identity fraud involves the creation of fictitious identities using a combination of real and fabricated information, posing significant challenges to financial verification and leading to substantial losses. AI-generated phishing attacks leverage natural language processing (NLP) and machine learning to create highly convincing and personalized phishing emails, text messages, or voice calls, increasing the success rate of social engineering attacks and potentially leading to data breaches and financial losses. This review analyzes the potential impact of these threats on financial authentication, verification, and social engineering, emphasizing the urgent need for enhanced security measures. These measures include robust authentication methods like multi-factor authentication (MFA) and liveness detection, adversarial training of AI models to improve resilience against attacks, and continuous monitoring systems to detect and respond to threats in real-time. Additionally, the paper explores the broader implications of generative AI for financial markets, including the potential for market manipulation, herding behavior, and the challenges of AI explainability and regulatory gaps. By addressing these challenges and opportunities, financial institutions can harness the benefits of generative AI while mitigating its potential risks to ensure the integrity and stability of financial systems in the age of AI.

Keywords: Generative AI, financial systems, security risks, deepfakes, synthetic identity fraud, AI-generated phishing attacks, authentication, verification, social engineering, cybersecurity.

1. Introduction

The financial services industry is rapidly integrating artificial intelligence (AI) to enhance efficiency, automate tasks, and improve customer experience. Generative AI, a subset of AI that can create new content, such as images, text, and audio, is gaining traction for its potential to revolutionize financial services. However, this rapid adoption introduces novel security risks that demand careful consideration. The increasing accessibility of these powerful AI tools is

© 2025 ITMA. All rights reserved.

empowering less sophisticated threat actors to carry out more advanced attacks, making it crucial for financial institutions to stay ahead of the curve. This paper provides a comprehensive review of the security risks associated with generative AI in financial systems. It focuses on three key threats: deepfakes, synthetic identity fraud, and AI-generated phishing attacks. The study analyzes their potential impact on critical aspects of financial operations, including authentication, verification, and social engineering.

2. Background & Research Methodology

This review paper is based on a systematic analysis of research papers, industry reports, and real-world case studies. The research process involved the following steps:

1. **Identification of relevant literature:** A comprehensive search was conducted across academic databases, such as Scopus, IEEE Xplore, and ACM Digital Library, as well as industry publications like the Financial Stability Board (FSB) and the Financial Services Information Sharing and Analysis Center (FS-ISAC) publications, and reputable online sources like Google Scholar to identify research papers, reports, and articles discussing the security risks of generative AI in financial systems.
2. **Selection of key sources:** The identified sources were screened for relevance and quality, with a focus on peer-reviewed research papers, reports from reputable organizations (e.g., FS-ISAC, the Federal Reserve), and articles providing real-world examples of AI-related security incidents in the financial sector. The selection criteria included the recency of the publication, the credibility of the source, and the relevance of the findings to the focus of this review paper.
3. **Data extraction and analysis:** Key findings, data points, and insights were extracted from the selected sources. This information was then analyzed to identify common themes, trends, and challenges related to the security risks of generative AI in financial systems.
4. **Synthesis and interpretation:** The extracted information was synthesized and interpreted to provide a comprehensive overview of the topic, highlighting key security risks, their potential impact, and mitigation strategies.

3. Research Findings

3.1 Deepfakes

Deepfakes are synthetic media, often videos or audio recordings, created using AI to convincingly replace a person's likeness or voice with someone else's. They can be used to create targeted phishing attacks, manipulate markets through disinformation campaigns, and distort identity verification processes. In the financial sector, deepfakes pose a significant threat to authentication and verification processes. Attackers can use deepfakes to impersonate customers, employees, or executives to gain unauthorized access to accounts, initiate fraudulent transactions, or manipulate markets. On average, businesses across industries have lost nearly $450,000 to deepfakes.

Impact on Financial Authentication: Deepfakes can bypass traditional security measures, such as facial recognition or voice authentication, by creating realistic impersonations. This undermines the integrity of Know Your Customer (KYC) and Anti-Money Laundering (AML) protocols, potentially leading to financial losses and regulatory penalties.

© 2025 ITMA. All rights reserved.

Challenges in Deepfake Detection:

- **Data limitations:** Training effective deepfake detection models requires a vast amount of real-world data, both genuine and manipulated. Financial institutions may struggle to access or share such sensitive data due to privacy concerns.
- **Evolving technology:** Deepfake creators constantly improve their techniques, making detection a moving target. Criminals can adapt their methods to evade current detection methods, necessitating continuous updates to security systems.

Real-world examples:

- In February 2024, a finance worker in Hong Kong was tricked into transferring $25 million after a video call with deepfakes impersonating the company's CFO and other colleagues.
- A Wall Street Journal reporter successfully cloned her voice using AI and used it to fool her bank's voice authentication system.
- Fraudsters have used AI technology to forge real-life video meetings, leading to significant financial losses.
- Deepfakes of children's voices have been used for scam kidnappings.
- Deepfake images of non-existent people have been used to create fake accounts for information operations.

3.2 Synthetic Identity Fraud

Synthetic identity fraud involves creating a fictitious identity by combining real and fabricated information. There are a few different categories of this type of fraud: manipulated synthetics (where a real person's data is modified), Frankenstein synthetics (which combine data from multiple real people), and manufactured synthetics (where the identity is completely fabricated). Fraudsters often use stolen Social Security numbers with fake names and addresses to open accounts, build credit, and eventually "bust out" by maxing out credit lines and disappearing. They may also exploit a technique known as "piggybacking," where they link synthetic identities to legitimate customers' credit accounts to build credibility before launching their attack.

Impact on Financial Verification: Synthetic identities can bypass traditional verification checks as they appear legitimate and may even have seemingly valid credit histories. This poses a significant challenge for financial institutions, leading to substantial financial losses and difficulties in identifying and mitigating fraud.

Real-world examples:

- Losses from synthetic identity fraud in the US were estimated to be $20 billion in 2020.
- A 2022 study by Security Magazine found that 46% of organizations faced synthetic identity fraud.
- Synthetic identities are also used for national security purposes, such as allowing criminals and terrorists to enter countries undetected.
- AI can be used to generate fake driver's licenses and other documents to set up bank accounts for fraudulent purposes.

3.3 AI-Generated Phishing Attacks

© 2025 ITMA. All rights reserved.

AI-powered phishing attacks leverage natural language processing (NLP) and machine learning to generate highly convincing and personalized phishing emails, text messages, or even voice calls. These attacks can bypass traditional security filters and manipulate individuals into divulging sensitive information or making fraudulent payments. Predictions indicate that AI-driven phishing attacks will see a significant increase, leveraging advanced language models to craft more believable and targeted messages.

Impact on Social Engineering: AI-generated phishing attacks exploit human psychology by creating a sense of urgency, mimicking trusted sources, and tailoring messages to individual victims. This increases the success rate of social engineering attacks, leading to data breaches, financial losses, and reputational damage. AI can also be used to analyze company email conversations and impersonate executives for malicious purposes.

Specific Techniques Used in AI-Powered Phishing Attacks:

- **NLP:** AI models use NLP to analyze and understand human language, enabling them to generate highly realistic and contextually appropriate phishing messages.
- **Behavioral Analysis:** AI systems can analyze user behavior patterns on social media and other online platforms to predict their responses to phishing attempts and adjust the timing and content of messages accordingly.

AI's Role in Expanding Cyberattacks:

- **Enhancing Existing Attacks:** AI is being used to improve the sophistication and efficiency of existing cyberattacks, such as phishing and business email compromise (BEC).
- **Investigation and Reconnaissance Support:** AI can automate and improve the efficiency of human decision-making in pre-attack investigation and reconnaissance.
- **Vulnerability Investigation:** AI can streamline the investigation of potential attack surfaces
- **Incident Response Assistance:** AI can assist in incident investigation and response operations.
- **Creating New Cybercrime Strategies:** Cybercriminals are using AI to create new ideas for cybercrimes.
- **Bulk Invoice Creation and Swapping:** AI tools can scrutinize compromised email data, find invoices that are to be paid, and alter payee details.
- **Analyzing Data Sets and Spotting Opportunities:** AI tools can analyze large datasets to quickly find vulnerabilities and high-value targets.
- **Refining Password Cracking:** AI is being used to refine password cracking techniques.
- **Intellectual Property Theft:** AI algorithms can analyze high volumes of data to identify valuable trade secrets or sensitive information.
- **Creating Fake News and Misinformation:** AI can generate fake news, manipulated articles, and false reports to influence public opinion and spread misinformation.
- **Social Media Impersonation and Fake Accounts:** AI-powered tools create thousands of fake social media accounts to spread scams, manipulate discussions, and gain trust.
- **Automating Fraud Schemes:** AI technologies can be used to create codes and scripts to automate complex fraud schemes like credential stuffing.
- **Document Forgery:** AI can be used to create or alter documents, such as bank statements, audit reports, and financial records.
- **Writing Exploit Code:** AI has the potential to write exploit code based on Common

© 2025 ITMA. All rights reserved.

Vulnerabilities and Exposures (CVE) descriptions.

Real-world examples:

- A 2024 report by CyberPeace Foundation analyzed an AI-driven phishing attack on a leading Indian bank where attackers used AI to generate highly realistic emails that mimicked the bank's internal communications.
- A 2023 survey found that 75% of security professionals reported an uptick in attacks over the past year, with most attributing the increase to generative AI.
- Cybercriminals are using AI to improve the tailoring of their phishing attacks through enhanced social media and other publicly available information searches.
- AI is making it easy to mimic a real person's voice or create a scam website that looks exactly like a real one.
- AI is being used to develop bots to attack financial institutions.

4. Impact on Financial Markets

Generative AI agents have the potential to destabilize the financial system by enabling malicious actors to defraud consumers, execute cyberattacks against financial institutions, and engage in market manipulation. The reliance on a small number of AI providers for financial decisions can lead to herding behavior that results in bank runs or flash crashes. Additionally, AI agents deployed by financial institutions can "hallucinate" false or misleading information, potentially causing harm to customers and the institutions themselves.

5. Challenges and Opportunities

Challenges:

- **Evolving Threat Landscape:** The rapid advancement of generative AI technology means that attackers are constantly developing new and more sophisticated techniques.
- **Limited Explainability:** The "black box" nature of some AI models makes it difficult to understand their decision-making processes and identify vulnerabilities.
- **Data Dependency:** AI models rely heavily on data, making them susceptible to data poisoning and bias, which can perpetuate or exacerbate existing inequalities in financial systems.
- **Regulatory Gaps:** Existing regulations may not adequately address the unique challenges posed by generative AI in financial services.

Opportunities:

- **Enhanced Security Measures:** Generative AI can also be used to develop more robust security solutions, such as advanced fraud detection systems and AI-powered authentication tools.
- **Proactive Risk Management:** AI can help financial institutions proactively identify and mitigate potential threats by analyzing vast amounts of data and detecting anomalies.
- **Improved Customer Experience:** AI-powered chatbots and virtual assistants can enhance customer service and provide personalized financial advice.

© 2025 ITMA. All rights reserved.

6. Discussion

The findings of this review highlight the urgent need for financial institutions to adopt a proactive and multi-layered approach to mitigate the security risks associated with generative AI. This includes:

- **Strengthening Authentication and Verification:** Implementing robust authentication methods, such as multi-factor authentication (MFA) and liveness detection, can help prevent deepfake attacks and synthetic identity fraud. MFA requires multiple forms of verification, such as something the user knows (e.g., a password), something the user has (e.g., a smartphone), and something the user is (e.g., biometric data).
- **Adversarial Training:** Training AI models with adversarial examples can improve their resilience against attacks and enhance their ability to detect anomalies.
- **Continuous Monitoring:** Implementing continuous monitoring systems that leverage AI can help detect suspicious activities and respond to threats in real-time.
- **Data Security and Privacy:** Protecting data used for training and operating AI models is crucial to prevent data poisoning and privacy breaches. Data poisoning involves attackers introducing additional, mislabelled instances into the training data to train the AI model incorrectly.
- **Collaboration and Information Sharing:** Sharing information about AI-related threats and best practices across the financial sector can help institutions stay ahead of evolving risks. * **AI Security Posture Management (AI-SPM):** AI security can't be an afterthought. To secure themselves from adversarial AI attacks, businesses must choose a unified cloud security tool in which AI-SPM is a central capability.
- **GenAI Tenant Architectures:** Enterprises should be very deliberate with what tenant architecture model they use for GenAI-incorporating services. The three fundamental types of GenAI tenant architectures are multi-tenant, single-tenant, and hybrid architectures. Businesses should spread some components of their GenAI services across multiple tenants and provide others with dedicated tenants. * **Input Preprocessing:** Applying preprocessing techniques to inputs before feeding them into the model can help detect and mitigate adversarial perturbations.
- **Network Distillation:** During the model training phase, combine outputs from multiple deep neural networks (DNNs). * **Responsible AI Hygiene Measures:** AI model security is also intricately related to other responsible AI tenets, including fairness, transparency, and explainability.
- **Red Teaming:** Red teaming simulates real-world attacks on an AI system from the point of view of an adversary, testing its inherent defences and vulnerabilities.
- **Defense Platforms:** Defense platforms detect adversarial attacks, with targeted responses for each threat vector.
- **Integrating AI Model Security with Enterprise Security:** Firms must reassess their existing cybersecurity initiatives, keeping in mind the vulnerabilities of enterprise AI models in production.
- **Rotating and Hardening Encryption Keys:** Frequently rotate encryption keys and apply hardware security modules (HSMs) for storing model artifacts and datasets.
- **Implementing Differential Privacy for Data Protection:** Apply differential privacy techniques when training AI models to ensure sensitive data cannot be reverse-engineered from the model.
- **Using Adversarial Example Detection Tools:** Implement tools that actively monitor and

© 2025 ITMA. All rights reserved.

identify adversarial inputs designed to mislead AI systems.

- **Leveraging Model Explainability to Detect Anomalies:** Use explainability techniques to continuously validate AI decision-making processes.
- **Employing Federated Learning for Sensitive Environments:** Employ federated learning for sensitive environments. * **Gradient Masking:** Gradient masking involves obscuring gradient information exploited by attackers to craft adversarial examples. * **Defensive Distillation:** Defensive distillation aims to reduce model sensitivity to minor input perturbations by generating smoother, less precise predictions.
- **Ensuring AI Regulatory Compliance:** Maintain strict adherence to legal frameworks governing AI usage.
- **Crafting an Internal Governance Framework for AI:** An ethics committee should oversee AI practices, ensuring they align with regulatory requirements and institutional values.
- **Continuously Upgrading Fraud Prevention Infrastructure:** Regularly enhance fraud prevention infrastructure by integrating the latest AI and machine learning technologies.
- **Empowering Teams with Continuous AI Learning:** Empower teams with continuous AI learning to stay ahead of evolving threats and best practices.
- **Strengthening Multi-Factor Identity Verification Processes:** Strengthen multi-factor identity verification processes to add layers of security and prevent unauthorized access.
- **Data Sanitization:** Implement proper data sanitization processes to ensure that sensitive or personally identifiable information is protected.
- **Secure Model Development and Deployment:** Secure coding practices should be followed throughout the development lifecycle of generative AI models to minimize vulnerabilities.
- **Continuous Monitoring and Vulnerability Management:** Continuous monitoring and vulnerability management are essential to identify and address potential security weaknesses
- **Adversarial Testing and Defense:** Adversarial testing and defense strategies should be implemented to proactively identify and mitigate potential attack vectors.
- **Leveraging Explainable AI:** Leverage explainable AI to understand the decision-making processes of AI models and identify potential biases or vulnerabilities.
- **Strict Access Controls and Authentication:** Implement strict access controls and authentication measures to prevent unauthorized access to AI systems and data.
- **Improving Privacy and Data Protection Systems:** Improve privacy and data protection systems to safeguard sensitive information and comply with relevant regulations.
- **Establishing Proper Model Governance and Tracking:** Establish proper model governance and tracking mechanisms to ensure the responsible development and deployment of AI models.
- **Investing in AI Ethical and Security Training:** Invest in AI ethical and security training to educate employees about the potential risks and best practices associated with AI.
- **Working with Cybersecurity Professionals and AI Researchers:** Work with cybersecurity professionals and AI researchers to stay informed about the latest threats and develop effective mitigation strategies.
- **Developing an AI Governance Framework:** Develop a comprehensive AI governance framework that outlines ethical guidelines, security protocols, and risk management strategies for AI development and deployment.
- **Training Employees on Generative AI Security Risks:** Train employees on generative AI security risks to raise awareness and promote a security-conscious culture.
- **Controlling the Use of Sensitive Work Data in Generative AI Systems:** Control the use of sensitive work data in generative AI systems to prevent data leakage and ensure

© 2025 ITMA. All rights reserved.

compliance with privacy regulations.

- **Classifying, Anonymizing, and Encrypting Data Used with Generative AI:** Classify, anonymize, and encrypt data used with generative AI to protect sensitive information and minimize the risk of data breaches.
- **Limiting Behavior with Additional Prompts in the Service:** Limit the behavior of AI models by using additional prompts to restrict their responses and prevent unintended actions.
- **Formal Checks of I/O:** Implement formal checks of input and output data to ensure the integrity and accuracy of AI-generated content.
- **Checking of I/O with an Independent Generative AI:** Check the input and output data with an independent generative AI model to identify potential biases or inconsistencies.
- **Vulnerability Testing Including Implementation of Adversarial Prompts:** Conduct vulnerability testing, including the implementation of adversarial prompts, to identify and address potential security weaknesses.

7. Conclusion

Generative AI presents both significant opportunities and challenges for the financial services industry. While it has the potential to drive innovation and improve efficiency, it also introduces novel security risks that demand careful consideration. By adopting a proactive and multi-layered approach to security, financial institutions can harness the benefits of generative AI while mitigating its potential risks. This includes implementing robust authentication, adversarial training, continuous monitoring, and data protection measures. Collaboration and information sharing across the industry are also essential to counter the evolving threat landscape and ensure the long-term integrity and stability of financial systems in the age of AI.

References

[1] Joshi, R., Pandey, K. and Kumari, S. (2025). Generative AI. In Generative Artificial Intelligence in Finance (eds P.R. Chelliah, P.K. Dutta, A. Kumar, E.D.R.S. Gonzalez, M. Mittal and S. Gupta). https://doi.org/10.1002/9781394271078.ch7

[2] Lee, D. K. C., Guan, C., Yu, Y., & Ding, Q. (2024). A Comprehensive Review of Generative AI in Finance. FinTech, 3(3), 460-478. https://doi.org/10.3390/fintech3030025

[3] Haidar, A. and Abbass, A. (2025). Navigating the Frontier of Finance. In Generative Artificial Intelligence in Finance (eds P.R. Chelliah, P.K. Dutta, A. Kumar, E.D.R.S. Gonzalez, M. Mittal and S. Gupta). https://doi.org/10.1002/9781394271078.ch12

[4] Moharrak, M. and Mogaji, E. (2024), "Generative AI in banking: empirical insights on integration, challenges and opportunities in a regulated industry", International Journal of Bank Marketing, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/IJBM-08-2024-0490

[5] Mallikarjuna Paramesha, Nitin Liladhar Rane, & Jayesh Rane. (2024). Artificial Intelligence, Machine Learning, Deep Learning, and Blockchain in Financial and Banking Services: A Comprehensive Review. Partners Universal Multidisciplinary Research Journal, 1(2), 51–67. https://doi.org/10.5281/zenodo.12826933

[6] Sai, S., Arunakar, K., Chamola, V., Hussain, A., Bisht, P. and Kumar, S. (2025), Generative AI for Finance: Applications, Case Studies and Challenges. Expert Systems, 42: e70018. https://doi.org/10.1111/exsy.70018

© 2025 ITMA. All rights reserved.