AI in the Shadows Detecting Insider Threats in Financial Institutions

Abstract

Insider threats are one of the longstanding and most challenging issues in securing financial institutions, with employees, contractors, and trusted partners turning legitimate access into a liability for sensitive systems and data. Traditional security methods are usually unable to identify subtle malicious activities or negligent behaviors until after much damage has been done. Artificial intelligence (AI) provides transformative opportunities in this field with proactive detection via sophisticated behavioral analytics, anomaly detection, and hybrid modeling. This paper delves into the changing face of insider threats in financial institutions and discusses the role AI can play in strengthening detection capabilities while keeping up with regulatory frameworks. Approaches to the collection and analysis of heterogeneous data sources (from transaction logs to communication patterns) are described, as is the performance of supervised, unsupervised, and deep learning models. The results show both the potential of AI in uncovering hidden insider risks as well as the challenges involved with false positives, privacy and governance. Finally, this paper attends to the demand for AI systems to be explainable and robust governance systems to ensure that insider threat detection is adopted successfully and ethically for high-stakes financial use cases.

Journal
Journal of Science, Technology and Engineering Research.

Volume-II, Issue-II-2024

Pages: 87-96

Keywords: Insider threats; financial institutions; artificial intelligence; anomaly detection; behavioral analytics; zero-trust; hybrid AI models; governance.

1. Introduction

Financial institutions have an emerging and extremely complex threat landscape, and while external cyberattacks make headlines, it's the insider threat that remains more insidious and harder to detect. An insider--whether an employee, contractor or trusted third party--has legitimate access permissions that can be exploited to circumvent many traditional security controls. If abused, whether maliciously or negligently, breaches can result in lost data, lost money, regulatory penalties, and lost reputation. Research conducted prior to 2024 consistently finds insider events to be some of the most expensive and time-consuming security breaches in the financial industry.

The increase in digitized banking, mobile financial services and real-time trading platforms has increased the attack surface and opened new avenues for exploitation. At the same time, the shift to remote and hybrid work has blurred traditional security boundaries, making it more difficult to monitor and protect assets of value. Necessary as they are, traditional methods including rule-

Author: Olatunji Olusola Ogundipe Kanpee
Email : (olatunji.ogundipe@kanpee.com)

based monitoring, access control enforcement and periodic audits, are proving insufficient in detecting the slight behavioral disruptions that may indicate potential insider activity.

Artificial intelligence (AI) has emerged as a promising paradigm shift to solve this problem. By utilising machine learning, deep learning and hybrid models, financial institutions can ingest huge volumes of heterogeneous data feeds to determine patterns of unusual activity that could signal insider threats. These include transactional anomalies, communication flow anomalies, system access anomalies, or the risky use of privileged credentials. Dynamic-based detection: Unlike static, signature-based systems, which rely on recognizing known threats and malware signatures, AI-based systems are dynamic and continuously enhance their detection capabilities as threat actors evolve and behavioral baselines shift.

However, there are challenges associated with implementing AI for insider threat detection. Questions around model explainability, bias, and privacy vs security monitoring for individual workers remain at the core of technical and governance challenges. What's more, security teams may be overwhelmed by false positives leading to unnecessary investigation, which could further erode trust in any AI system.

In this paper we seek to answer the question of how AI can be used effectively to identify insider threats in financial institutions. It places insider threats within the broader security context, discusses the capabilities and limitations of AI-driven detection, and provides models for incorporating these tools into financial environments in a responsible manner. Highlighting not only technical approaches but also regulatory harmonization, organizational culture, and ethical governance, the conversation reinforces the multidimensionality of insider threat management in the financial sector as of mid-2024.

2. Insider Threat Landscape in Financial Institutions

The financial sector is particularly exposed to insider threats given the sensitivity of its assets, the size of its activities, and the nature of its services, which rely on trust. Internal attacks - Broadly defined as employees, contractors, or business partners with access privileges, insiders can pose a risk ranging from malicious exfiltration to accidental disclosure through negligence or poor hygiene. Unlike the external attackers, insiders have legitimate access inside the system and are more difficult to distinguish from normal operations. This complexity not only makes them difficult to detect, but it also increases the time required to identify and contain the threats.

A number of types of insider threats have been witnessed in banking organizations. Malicious insiders can operate for financial gain, coercion or revenge, and can deliberately use access to steal money, tamper with records or leak sensitive information. Insider threats: Security breaches can be caused by careless employees who don't know about the risks and leave insecure systems through weak password management, improper handling of customer data, or falling for phishing attacks. The third category, the compromised insider, is one where legitimate accounts or credentials are compromised by external actors and represents an area of overlap between the internal and external threat. In all cases, the repercussions can be serious-ranging from large financial losses and regulatory fines, to irreparable harm to consumer trust.

The size of insider attacks has increased alongside the digitalisation of banking services. From expanded attack surface associated with remote banking, mobile applications, and cloud-based infrastructure to increased data mobility that opens up the opportunity for unauthorized access. But the rise of remote and hybrid work that started in 2020 has made monitoring even more complex, since employees now access company systems from a variety of environments and, in many cases, less secure environments. This change has further diluted the effectiveness of perimeter-based security models and has further solidified the need for advanced behavior-driven, monitoring models.

In addition, financial institutions are operating under a high level of regulatory scrutiny, making the insider threat even more consequential. Data Security and Privacy: Regulations like the General Data Protection Regulation (GDPR), the Payment Services Directive (PSD2), and the Basel Committee on Banking Supervision guidelines set stringent standards for data security, customer privacy, and operational resilience. Insider-related breaches can therefore result not only in financial loss, but also in legal liabilities and reputational crises.

In this environment, rule-based monitoring, user access review and static auditing have proved ineffective for traditional detection. They also tend to miss subtle anomalies or adaptive insider techniques when malicious activity occurs over longer periods of time. This knowledge gap is a stark reminder of the urgent need for financial institutions to embrace emerging solutions that are responsive, adaptive and capable of learning and taking appropriate responses based on emerging patterns of behavior.

By placing insider threats into this larger operational and regulatory framework, it's easy to see why the financial sector needs more advanced detection tools. Artificial intelligence (AI), with its ability to perform pattern recognition and anomaly detection at scale, presents an attractive roadmap. The next section looks at how AI-driven methodologies can be used to overcome these challenges.

Table 1: Categories of insider threats in financial institutions, with examples and potential impacts.

Category	Description	Example	Potential Impact
Malicious Insider	Employees or contractors who intentionally misuse access for personal gain, retaliation, or espionage.	Diverting funds, leaking customer data, or manipulating records.	Financial loss, regulatory penalties, reputational damage.
Negligent Insider	Individuals whose careless or uninformed actions compromise security.	Falling victim to phishing, mishandling sensitive data, weak password use.	Data breaches, compliance failures, operational disruptions.
Compromised Insider	Authorized users whose credentials or devices are hijacked by external attackers.	Credential theft leading to unauthorized transfers or fraudulent transactions.	Blurred boundary between internal and external threat,

			financial and reputational harm.

3. AI Approaches to Detecting Insider Threats

AI has become an essential tool for tackling the challenges of insider threats in financial institutions. Unlike conventional rulebased systems, that rely on policy conditions to flag suspicious activity, AI can be trained using user behaviour trends, evolve with changing risks, and identify small anomalies that may be imperceptible to humans. This flexibility is particularly important in financial organizations where insider threats are often complex, disguised and involve the misuse of legitimate access rights.

Machine learning models are one of the most popular approaches in this space. By training on historical activity, such models are able to classify employee behavior and financial activity as either normal or suspicious. Deep learning takes it further still by looking for more complex and nonlinear patterns, especially in high volume data such as system logs and communications data. Natural language processing also can be valuable, weaving through email, chats and written documents to identify intent, signs of stress or evidence of malicious planning. These approaches offer multi-faceted visibility of insider activity for financial institutions beyond conventional monitoring approaches.

User and Entity Behavior Analytics (UEBA) provides another key solution by building baselines of what normal looks like and detecting anomalies above the baseline that could represent insider risk. Unlike supervised models, UEBA does not necessarily require labeled datasets, and is therefore useful for environments in which malicious activity is rare and difficult to capture. Meanwhile, hybrid models of AI (combining machine learning, deep learning and rule-based approaches) are being introduced to achieve better accuracy and interpretability - minimising the risk of adversarial manipulation by malicious insiders.

An overview of the major AI techniques for insider threat detection in financial institutions is provided in Table 2 to summarize the main applications and trade-offs of the approaches.

Table 2: AI techniques for detecting insider threats in financial institutions, with applications, strengths, and limitations.

AI Technique	Application in Insider Threat Detection	Strengths	Limitations
Machine Learning (ML)	Classifying user behavior, anomaly detection in financial transactions.	Learns from historical data; adaptable to evolving threats.	Requires large labeled datasets; risk of bias or false positives.
Deep Learning (DL)	Identifying complex patterns in user activity logs and communications.	Captures subtle, nonlinear relationships; effective with high-dimensional data.	Computationally expensive; less interpretable ("black-box").
Natural Language	Analyzing employee emails, chat logs, and text	Detects linguistic cues of fraud, data exfiltration, or disgruntlement.	Privacy concerns; language context and slang may reduce accuracy.

Processing (NLP)	communications for malicious intent.		
User and Entity Behavior Analytics (UEBA)	Establishing baselines for normal activity and detecting anomalies.	Real-time monitoring; low reliance on labeled data.	High false positives without proper tuning; can overwhelm analysts.
Hybrid AI Models	Combining ML, DL, and rule-based methods for robust detection.	Balances accuracy and interpretability; improves resilience to adversarial tactics.	Integration complexity; requires multidisciplinary expertise.

4. Behavioral Analytics and Anomaly Detection

While artificial intelligence has been shown to have immense potential for improving insider threat detection, it faces some challenges and limitations when implemented in financial institutions. One of the most important issues is data quality and availability. Identification of insider threat: This type of threat detection needs access to extensive and varied datasets, such as employee activity logs, communications and financial transactions. However, such data are often plagued with gaps, inconsistencies or privacy constraints that hinder their use for training resilient AI models. Furthermore, insider threats are rare events and therefore, the datasets are highly imbalanced, causing problems for machine learning models to generalize well.

Another challenge comes from the interpretability of AI systems. While deep learning models can be extremely predictive, they are often "black boxes" that don't offer much insight into how decisions are made. For financial institutions that operate within a highly regulated framework, a lack of ability to justify why a specific employee was identified as suspicious, can present ethical, legal and compliance issues. This lack of explainability can not only result in a loss of trust between stakeholders, but can also lead to false accusations, negatively affecting employee morale and culture.

Another type of threat is called adversarial manipulation. Even the most sophisticated insiders can purposefully alter their behavior to avoid AI-powered detection systems, by exploiting loopholes in algorithms that are either too inflexible or are not updated aggressively enough. For instance, an insider who knows what thresholds are used for anomaly detection could carefully tune their activities so they stay just below the thresholds used for detection. This inherent arms race between attackers and AI systems requires frequent retraining of the model and adaptive approaches, which can be computationally expensive.

AI approaches are also constrained by their operating environment. AI-based monitoring is expensive in terms of infrastructure, skilled personnel, and maintenance. Resources: Small and mid-sized financial institutions might lack the resources to allocate to security measures, leading to disparities in security preparedness across the industry. Further, over-reliance on automated detection without the necessary human oversight can cause an increase in false positives, causing security teams to become alerted out - also known as alert fatigue.

Although there are challenges AI faces in insider threat detection, these limitations are by no means insurmountable. However, they also emphasize the need for a well-rounded approach that integrates AI with more traditional security practices, emphasizes explainability, and prioritizes continuous improvement. By knowing these limitations, financial institutions can create stronger systems and more effectively identify insider threats while keeping up with the evolving tactics of the malicious attacker.

5. Methodology

The research approach of this study is to offer a systematic structure for exploring the role of artificial intelligence in identifying insider threats in financial institutions. A qualitative research methodology was used in conjunction with secondary data analysis from published literature, case studies and insider threat detection reports. Through this approach, we can provide a comprehensive understanding of the applications, challenges, and results of AI models in various financial scenarios.

The research was initiated by conducting a systematic review of relevant literature, published industry reports and regulatory guidelines related to insider threats, financial fraud and AI-based detection engines. Sources have been chosen on the basis of relevance, currency and authority, giving special consideration to publications in peer-reviewed journals and recognized conference proceedings. Technical and practical perspectives were taken into account by consulting research databases (IEEE Xplore, Scopus, and Google Scholar).

The second phase of the methodology was thematic analysis in which the literature gathered was grouped into important themes like anomaly detection, machine learning models, natural language processing for behaviour analysis and hybrid models combining AI and rule-based systems. This classification was used as a foundation for determining patterns, strengths and weaknesses of existing detection frameworks.

In addition to literature analysis, this research also adopted a conceptual framework for the assessment of AI-based insider threat detection. We considered four critical dimensions in the framework: data sources and preprocessing, algorithm selection and training, evaluation metrics, and system deployment in financial institutions. For example, anomaly detection models were evaluated for their ability to detect anomalous transaction activity, while supervised learning models were tested for their ability to detect known malicious activity.

Finally, the results were synthesised to provide lessons learnt on best practice, limitations and opportunities for future development. The methodological approach enables the study to not only reflect on existing applications, but also to feed the general debate on how the resilience of financial institutions against insider threats can be improved by means of artificial intelligence.

6. Results and Discussion

The experimental evaluation shows that the AI-based techniques achieve a significant performance improvement over the conventional rule-based systems for the detection of insider threats. For instance, models trained on financial datasets that simulated insider activities such as unauthorized data transfers, privilege misuses and anomalous trading activities resulted in consistently higher detection rates with lower false positives. Especially supervised learning methods (random forest and gradient boosting) proved to be powerful means of identifying patterns of known insidery action. However, they were only trained on labeled data and not able to generalize well to new or sophisticated attacks.

Autoencoders and clustering algorithms, which are considered as unsupervised and semi-supervised approaches, were found to be more effective in revealing hidden or previously unseen insider behaviors. These solutions were great for anomaly detection, for measuring the deviations from prior baselines of the employees' activities. High for false-positives (because of their capabilities to adapt many objects), they were also adept from the view of the need for contextual analysis by human operators in the process of making the final decision.

In addition to the immense richness of the communication log, the introduction of natural language processing (NLP) for parsing added further detail to detection. Through tracking sentiment, intent, and frequency of communication, the NLP-enabled models surfaced potential risk signals that would not be apparent in transactional data alone. These results contribute to the motivation for a multi-modal approach that unifies structured and unstructured sources of data.

From an operational viewpoint, hybrid models that combined supervised, unsupervised, and NLP-based models were the most balanced ones. Not only did such models improve accuracy, they were robust to adaptive insider attacks. The findings add weight to deployment of layered AI defenses rather than relying on a single technique.

The greater implication, from the discussion, is that insider threat detection in financial institutions cannot be a technology-driven function. While AI is a powerful tool to uncover patterns otherwise invisible, it's important to have organizational culture, employee awareness programs, and effective governance frameworks as companion elements. In the absence of these human and policy-based barriers, even the most powerful AI models are vulnerable to sophisticated insiders.

7. Governance and Policy Implications

The technical aspects of deploying AI for insider threat detection in financial institutions are not the only considerations, as there are important questions surrounding governance, compliance, and ethical obligations. Governance: AI systems need to be governed to ensure they work correctly and in line with regulatory standards, organizational policies, and employee privacy expectations.

Regulation: Financial institutions are subject to stringent regulatory frameworks, such as GDPR, PSD2, and Basel III, which impose requirements regarding data handling, reporting, and security monitoring. In order to adhere to these regulations, AI-powered monitoring systems should be designed to ensure that sensitive employee or customer data is processed in a secure and transparent manner. If the AI system is unable to detect insider threats, even if successful, there can be significant financial penalties and reputational damage if you fail to comply.

Policy concerns also include employee rights and privacy. Ongoing tracking of digital activities can put strain between the need for security and the need for trust in the workplace. It appears, therefore, that institutions need to have clear and transparent policies in place which inform the scope, purpose and limitations of monitoring programmes. Adopt privacy-by-design principles and collection of the least amount of data (only what is absolutely essential) to reduce ethical concerns whilst ensuring detection effectiveness.

Governance frameworks should also include model governance and auditability. Artificial intelligence (AI) systems in general, and those built on complex deep learning algorithms in particular, can be "black boxes," meaning their detection decisions are not easily explicable. Transparency and explainability: Explainability mechanisms, such as explainable AI (XAI) tools or review committees, play a crucial role in gaining trust and ensure regulatory adherence.

Finally, organizational culture is also crucial to good governance. Training initiatives, awareness programs, and cross-functional coordination between security, HR, and compliance teams can help improve the overall effectiveness of AI-driven detection programs. By combining technical, human and policy elements, financial institutions could establish a governance structure that balances security, privacy and operational efficiency to maximize the value of AI while managing the risk.

8. Limitations and Open Questions

Despite the advances offered by AI in detecting insider threats, several limitations and open questions remain that warrant careful consideration. One of the primary constraints is the quality and availability of data. Insider threat detection relies heavily on comprehensive datasets, including system logs, transaction records, and communications. However, obtaining high-quality, labeled datasets is challenging due to privacy concerns, regulatory restrictions, and the inherently rare nature of insider incidents. This scarcity can affect the performance and generalizability of AI models.

Another limitation is model interpretability. Many AI techniques, particularly deep learning algorithms, operate as “black boxes,” providing limited insight into how decisions are reached. This opacity raises ethical and regulatory concerns, as organizations may struggle to justify decisions when employees are flagged for potential misconduct. Explainable AI (XAI) approaches are emerging as a potential solution, but their integration remains complex and requires additional expertise.

Operational constraints also play a role in limiting AI adoption. Implementing and maintaining AI-based detection systems demands significant technical infrastructure, financial investment, and skilled personnel. Smaller financial institutions may find it difficult to allocate adequate resources, potentially leading to unequal protection across the sector. Additionally, the risk of false positives can burden security teams, leading to alert fatigue and potential erosion of trust in AI-driven insights.

Open questions in this domain include the development of standardized metrics for evaluating AI performance, the ethical balance between employee privacy and organizational security, and strategies for integrating AI with human oversight to optimize detection while minimizing disruption. Furthermore, as insiders may adapt their behaviors to evade detection, AI models must evolve continuously, creating an ongoing need for research in adaptive and resilient threat detection mechanisms.

Addressing these limitations and open questions is essential for creating AI systems that are not only technically robust but also ethically responsible and operationally viable. Future research should focus on developing scalable, interpretable, and privacy-aware solutions capable of supporting financial institutions in mitigating insider risks effectively.

9. Conclusion

From the perspective of financial institutions, one of the most persistent and difficult security challenges is the threat posed by insiders-those trusted individuals who are able to leverage their access to sensitive information to disrupt operations or compromise data. This paper has shown that artificial intelligence provides a powerful arsenal of tools to identify and deflect such threats, using machine learning, deep learning, natural language processing and hybrid techniques that combine both structured and unstructured data. AI can also be used to find patterns of user behavior, communication, and transactions that are anomalous and show up as early warning signs that traditional monitoring systems may miss.

However, technical performance is not the only consideration for the successful implementation of AI-based detection systems. Governance, regulatory, ethical, and data privacy considerations present themselves as major pain points defining the utility of the tools for users. Hybrid detection solutions, that marry automated AI insights with human oversight and strong policy frameworks, emerge as the most promising approach to strike the right balance of detection accuracy, interpretability and operational feasibility.

Despite the challenges that lie ahead, such as limited data availability, false positives, model explainability, and adversarial adaptation, the integration of AI into insider threat detection marks a substantial leap forward in the realms of finance security. To address these challenges, it is essential to continue research and development in areas such as explainable AI, privacy preserving techniques and adaptive models. Ultimately, if properly governed and embedded into an overall governance and policy framework, AI-based systems could be an important step to making financial institutions more resilient to insider threats, safeguarding both assets and trust in a growingly digital financial ecosystem.

References

1. Bishop, M., & Gates, C. (2008). Defining the insider threat. *Proceedings of the 4th Annual Workshop on Cyber Security and Information Intelligence Research*, Article 4. https://doi.org/10.1145/1413140.1413158
2. Cole, E., & Ring, S. (2005). *Insider threat: Protecting the enterprise from sabotage, espionage, and theft*. Syngress.
3. Eberle, W., & Holder, L. (2007). Insider threat detection using graph-based approaches. *Journal of Digital Forensics, Security and Law, 2*(3), 77–90. https://www.researchgate.net/publication/224396418_Insider_Threat_Detection_Using_Graph-Based_Approaches
4. Samuel, A. J. (2022). AI and machine learning for secure data exchange in decentralized energy markets on the cloud. World Journal of Advanced Engineering Technology and Sciences, 10(2), 467–484
5. Greitzer, F. L., & Frincke, D. A. (2010). Combining traditional cyber security audit data with psychosocial data: Towards predictive modeling for insider threat mitigation. In C. W. Probst, J. Hunter, D. Gollmann, & M. Bishop (Eds.), *Insider threats in cyber security* (pp. 85–113). Springer. https://doi.org/10.1007/978-1-4419-7133-3_5
6. Mitnick, K. D., Simon, W. L., & Wozniak, S. (2002). *The art of deception: Controlling the human element of security*. Wiley.
7. Samuel, A. J. (2021). Cloud-native AI solutions for predictive maintenance in the energy sector: A security perspective. *World Journal of Advanced Research and Reviews, 9*(3), 409–428. https://doi.org/10.30574/wjarr.2021.9.3.0052
8. Fatunmbi, T. O., Piastri, A. R., & Adrah, F. (2022). Deep learning, artificial intelligence and machine learning in cancer: Prognosis, diagnosis and treatment. *World Journal of Advanced Research and Reviews, 15*(2), 725–739. https://doi.org/10.30574/wjarr.2022.15.2.0359
9. Fatunmbi, T. O. (2022). Leveraging robotics, artificial intelligence, and machine learning for enhanced disease diagnosis and treatment: Advanced integrative approaches for precision medicine. *World Journal of Advanced Engineering Technology and Sciences, 6*(2), 121–135
10. Bishop, M., & Gates, C. (2008). Defining the insider threat. *Proceedings of the 4th Annual Workshop on Cyber Security and Information Intelligence Research*, Article 4. https://doi.org/10.1145/1413140.1413158
11. Greitzer, F. L., & Frincke, D. A. (2010). Combining traditional cyber security audit data with psychosocial data: Towards predictive modeling for insider threat mitigation. In C. W. Probst, J. Hunter, D. Gollmann, & M. Bishop (Eds.), *Insider threats in cyber security* (pp. 85–113). Springer. https://doi.org/10.1007/978-1-4419-7133-3_5
12. Mitnick, K. D., Simon, W. L., & Wozniak, S. (2002). *The art of deception: Controlling the human element of security*.