International Journal of Scientific Research and Modern Technology (IJSRMT) *ijsrmt.com*
Volume 4, Issue 8, 2025
DOI: https://doi.org/10.38124/ijsrmt.v4i8.716

Challenges in Implementing Explainable AI (XAI) for Threat Detection in Banking Networks

Giwa Olajumoke Sherifat[1]

Publication Date 2025/08/27

Abstract

This research work looks into the challenges of implementing Explainable Artificial Intelligence (XAI) for threat detection in the banking network. The introduction of XAI is geared towards providing more transparency and accountability in AI-driven decision-making process, which will also address the concerns of reliability and trust in its process. The research also highlights the benefits and prospects of XAI in the banking network and its role in improving threat detection and reducing cases of false alarm in financial transactions. It also further discussed key challenges that impedes the successes of XAI in banking network such as complexity, data quality and compliance among others. The Author went further to provide recommendations and way further in addressing these challenges in implementing XAI in banking networks. The research therefore concluded that the emergence of XAI has the potential to enhance threat detection in the banking networks, but its implementation calls for more careful consideration as XAI will be more significant in banking networks going forward.

I. INTRODUCTION

The banking sector has undergone a significant transformation in recent years, driven by advances in technology and the increasing demand for digital services. The integration of artificial intelligence (AI) in threat detection systems has been a key development in this space, enabling banks to detect and respond to cyber threats more effectively. However, people often question the lack of transparency in AI decision-making processes which has raised concerns about the reliability and trustworthiness of these systems. Explainable AI (XAI) has therefore emerged as a reliable solution to address these concerns by providing adequate insights into AI-driven decision-making process. This research work discusses the challenges in implementing XAI for threat detection in banking networks, highlighting the benefits and limitations of this technology.

II. EVOLUTION OF THE BANKING NETWORK

The banking sector as seen massive advancement in technology over the years, from it days of traditional banking approaches. These days, technology has made financial activities seamless with the use of internet and mobile banking. The increased use of these technologies have also led to massive increase in cyber threats, making it essential for banks to implement robust security measures (Neelam, 2021).

As a result of this shift towards internet banking and the use of mobile Apps, there has been massive rise in cyber-attacks by hackers and criminal elements. It has become imperative for the financial institutions to find a way to safeguard funds and sensitive financial information of their customers, hence the need for a more advanced and sophisticated solutions by the financial institutions. To achieve a more secure security, many banks have introduced the use of AI-driven threat detection systems to help mitigate the incessant cyber-attacks. However, the operation of AI-driven threat detection systems have been clouded with lack of transparency in their decision-making processes. The emergence of Explainable AI (XAI) is important to address these challenges. XAI enables banking personnel to have a better understanding of its decision-making approach and thereby created more trust in its outcome and cyber defense tactics.

III. CYBER THREATS IN BANKING NETWORK

In contemporary banking era, there are more sophistication in technological techniques used by hackers to bypass banking network security. These threats can result in significant financial losses and damage to the reputation of banks (Muhammad, 2022). Some of the common cyber threats faced by banks include phishing, ransomware, and social engineering attacks (Ishika, 2023). Cyber threat is a significant challenge faced by financial institutions these days, and quite often it results in the loss of funds and important financial information of the

Sherifat, G. O. (2025). Challenges in Implementing Explainable AI (XAI) for Threat Detection in Banking Networks. *International Journal of Scientific Research and Modern Technology*, *4*(8), 32–35. https://doi.org/10.38124/ijsrmt.v4i8.716

customer. Due to the advanced techniques of these criminal hackers, the banks have been forced to look for increased sophistication in their security solutions. This is where AI-driven threat detection systems comes in, as an effort to limit and prevent these frequent cyber-attacks. Nevertheless, despite the benefits associated with AI-driven threat detection systems, there is a limitation in its transparency and accountability. This limitation makes it difficult to fully understand the systems. XAI has the potential to address these concerns by providing insights into AI-driven decisions, enabling banks to improve the security and reliability of their threat detection systems.

IV. EXPLAINABLE AI (XAI) EXPLAINED

XAI is a type of AI that provides insights into its decision-making processes, making it possible for humans to understand and interpret the results. XAI has the potential to improve the transparency and accountability of AI systems, which is essential for high-stakes applications such as threat detection in banking networks (Zhibo, 2024). Explainable AI (XAI) is increasingly critical in the banking sector's efforts to strengthen security protocols and give banking personnel the ability to understand its decision-making process which will help for a better threat detection approach. This will significantly reduce the occurrence of undetected cyber-attacks and increase trust in the banking sector.

The clarity been afforded by XAI will also enable the banking personnel to easily distinguish between genuine threats and false alarm as they can interpret the reason why an AI will flag a transaction or an account. This will overtime empower the financial institutions to enhance both their accuracy and effectiveness and gain trust from customers.

V. CATEGORIZATION OF XAI

According to Barredo 2020, XAI can be categorized into several types. However, given the complexity of XAI techniques, categorization methods may not be mutually exclusive, and a single technique can fit into multiple categories, to provide a nuanced understanding, it would be beneficial to categorize XAI techniques from multiple perspectives (Carvalho, 2019). This multi-faceted approach can reveal a richer set of characteristics and information about each technique, enabling a more comprehensive evaluation and application (Zhibo, 2024).

- *Intrinsic:*

This type of XAI is built into the AI model itself, providing insights into the decision-making process.

- *Model-Specific:*

This type of XAI is designed for specific AI models, providing insights into the decision-making process of those models.

- *Local or Global:*

XAI can provide insights into local or global decision-making processes, depending on the specific application.

- *Explanation Output:*

XAI can provide various types of explanation outputs, including text, images, and graphs (Muhammad 2022).

Properly categorizing XAI isn't just theoretical, it's a crucial step for banks aiming to understand which methods actually suit their needs. By choosing the right XAI categorization approach, banks and other financial institutions can finally bring the some much needed transparency to their AI-driven threat detection systems. This will help not only with accountability but also reduces the risk of frequent and costly errors and reputational fallout. In short, it's not something banks can afford to overlook.

VI. EXISTING CHALLENGES OF XAI

In-spite of the benefits of Explainable AI (XAI), there are several challenges associated with its implementation, these includes:

- *XAI Security:*

XAI systems can also be vulnerable to cyber-attacks, which can adversely affect their integrity and reliability. Researchers like Guo 2020 emphasize the need for defense mechanisms to detect targeted attacks on Explainable AI (XAI) systems, crucial for building trust in 6G-enabled autonomous industries. Similarly, Fatima et al (Hussain, 2021). suggest exploring adversarial machine learning in XAI, identifying key security factors for AI models as input data changes, bias, and fairness concerns.

- *Performance Evaluation:*

Evaluating the performance of XAI systems can be challenging, particularly in complex and dynamic environments.

- *Legal and Privacy Issues:*

XAI systems may raise legal and privacy concerns, particularly in applications where sensitive data is involved.

- *Disparity Between Interpretation and Accuracy:*

There may be a trade-off between the interpretability and accuracy of XAI systems, which can make it challenging to achieve both simultaneously. There's a growing need for high-performance models that are also explainable, as top-performing algorithms like deep learning often lack transparency (Barredo, 2020). While simple models are typically preferred for their interpretability, their explainability can be limited when complex or high-dimensional features are involved (Lipton, 2018).

VII. CHALLENGES IN IMPLEMENTING EXPLAINABLE AI (XAI) FOR THREAT DETECTION IN BANKING NETWORKS

Implementing XAI for threat detection in banking networks comes with several challenges that need to be looked into. These challenges include:

- *Complexity:*

Banking networks are complex and involve multiple applications, systems and stakeholders. This complexity can make it challenging to implement XAI solutions that can effectively detect and respond to threats (Neelam, 2021).

- *Lack of Transparency:*

Traditional AI models used in threat detection are often black boxes, making it difficult to understand the reasoning behind their decisions. XAI can help address this challenge, but it requires significant investment in data collection, model training, and testing (Muhammad 2022).

- *Data Quality and Availability:*

XAI models require high-quality and relevant data to learn and make accurate predictions. However, banking networks often generate vast amounts of data, making it challenging to collect, process, and analyze the data effectively (Ishika, 2023).

- *Regulatory Compliance:*

Banking institutions are subject to strict regulatory requirements, and XAI solutions must comply with these regulations. This can be a challenge, especially when implementing XAI solutions that require access to sensitive customer data (Sina, 2025).

- *Model Interpretability:*

XAI models can be complex and difficult to interpret, making it challenging for security teams to understand the reasoning behind the model's decisions. This can lead to a lack of trust in the XAI solution and make it less effective in detecting and responding to threats (Zhibo, 2024).

VIII. RECOMMENDATIONS

To overcome the challenges of implementing XAI for threat detection in banking networks, the following recommendations can be made:

- *Develop a Clear Strategy:*

Develop a clear strategy for implementing XAI solutions that aligns with the bank's security goals and objectives. This strategy should include a roadmap for implementation, data collection, and model training (Muhammad 2022).

- *Invest in Data Quality:*

Invest in data quality and availability by implementing data governance policies and procedures that ensure data accuracy, completeness, and relevance (Ishika, 2023).

- *Choose the Right XAI Model:*

Choose an XAI model that is suitable for the bank's specific security needs and provides transparent and interpretable results. The model should also be able to handle complex and dynamic banking networks (Zhibo, 2024).

- *Provide Training and Support:*

Provide training and support to security teams to ensure they understand the XAI solution and can effectively use it to detect and respond to threats (Neelam, 2021).

- *Continuously Monitor and Evaluate:*

Continuously monitor and evaluate the XAI solution to ensure it is effective in detecting and responding to threats. This includes monitoring model performance, data quality, and security metrics (Sina, 2025).

IX. CONCLUSION

In conclusion, Explainable AI (XAI) faces daunting challenges when deployed for threat detection in the banking networks. However, the main advantage of XAI is its ability to deliver not only accurate threat detection, but also clear explanations for its decisions. This helps banking personnel to better track and monitor threat with better information.

Nevertheless, it is also important to take cognizance of the fact that XAI solutions still require refinement in some quarters as networks criminals are constantly evolving in the strategies and techniques so must XAI will refined to face the constantly evolving threat of cyber-attacks.

These research work has successfully delved into important areas of the implementation and challenges faced with XAI with the aim of highlighting areas of challenges as well as presenting recommendation to strengthen the implementation of XAI which has come to stay.

REFERENCES

[1]. Barredo A. A. et al (2020). *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*. Information Fusion, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[2]. Carvalho D. V. et al (2019). *Machine Learning Interpretability: A Survey on Methods and Metrics.* Electronics, vol. 8, no. 8, Art. no. 8, Aug. 2019, doi: 10.3390/electronics8080832.

[3]. Guo W. (2020). *Explainable Artificial Intelligence for 6G: Improving Trust between Human and Machine.* IEEE Communications Magazine, vol. 58, no. 6, pp. 39–45

[4]. Hussain F. et al (2021). *Explainable Artificial Intelligence (XAI): An Engineering Perspective.* arXiv, Jan. 10, 2021. doi: 10.48550/arXiv.2101.03613.

[5]. Ishika P. M. (2023). *A Study on Cyber Security Affecting Online Banking and Online Transaction.* Department of Banking and Insurance, faculty of Commerce, University of Mumbai

[6]. Lipton Z. C. (2018). *The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.* Queue, vol. 16, no. 3, pp. 31–57, Jun. 2018, doi: 10.1145/3236386.3241340.

[7]. Muhammad A. F. et al (2022). *The Role of Explainable AI in Cybersecurity: Improving Analyst Trust in Automated Threat Assessment Systems.* IRE Journals, Volume 6 Issue 4, ISSN: 2456-8880

[8]. Neelam S. (2021). *Cyber security analysis in banking sector.* International Journal of Advanced Research in Commerce, Management & Social Science (IJARCMSS) ISSN: 2581-7930, Impact Factor: 5.880, Volume 04, No. 03(I), pp 59-64

[9]. Sina A. (2025). *Advancing Fraud Detection in Banking: Real-Time Applications of Explainable AI (XAI).* HAL Open Science, hal-04881704 https://hal.science/hal-04881704v1

[10]. Zhibo Z. et al (2024). *Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research.* IEEE Access; open access journal