Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 5, 2026

EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR CYBER THREAT DETECTION IN PAKISTAN'S FINANCIAL SECTOR

Andleeb Akram[*1], Syeda Nida Ejaz[2], Muhammad Suliman[3]

[*1]*Lecturer, Department of Computer Science, Government College University Faisalabad*
[2]*Student, Lecturer, Assistant Professor, Researcher, Department of Computer Science, University of Southern Punjab*
[3]*Assistant Professor, Department of Computer Science, University of Peshawar*

[*1]andleebakram317@gmail.com, [2]syedanidaejaz@gmail.com, [3]muhammadsuliman@uop.edu.pk

DOI: https://doi.org/10.5281/zenodo.20394229

Keywords
Explainable Artificial Intelligence (XAI); Cyber Threat Detection; Financial Cybersecurity; Machine Learning; Pakistan Banking Sector; AI Governance

Article History
Received: 30 March 2026
Accepted: 08 May 2026
Published: 26 May 2026

Corresponding Author: *
Andleeb Akram

Abstract

The rapid digitalization of Pakistan's financial sector has significantly increased exposure to sophisticated cyber threats, including phishing attacks, ransomware, malware, insider threats, financial fraud, and identity theft. Traditional cybersecurity systems based on rule-based and signature-driven approaches are increasingly ineffective against evolving and intelligent cyberattacks. Consequently, Artificial Intelligence (AI) and machine learning technologies have emerged as advanced solutions for automated cyber threat detection and predictive cybersecurity analytics. However, the "black-box" nature of many AI models creates major challenges related to transparency, interpretability, accountability, and institutional trust, particularly in highly sensitive financial environments. This study critically examined the role of Explainable Artificial Intelligence (XAI) in enhancing cyber threat detection within Pakistan's financial sector. Using a qualitative analytical research design and systematic literature review approach, the study analyzed recent scholarly research, cybersecurity reports, and AI governance frameworks related to explainable cybersecurity systems. The findings revealed that machine learning and deep learning models significantly improve intrusion detection, fraud prevention, and anomaly detection capabilities, while XAI techniques such as SHAP, LIME, decision trees, and interpretable neural networks enhance transparency, reduce false positives, and strengthen analyst confidence in AI-generated decisions. The study further identified major implementation barriers in Pakistan, including weak cybersecurity infrastructure, shortage of skilled AI professionals, inadequate regulatory frameworks, and limited institutional readiness for explainable AI adoption. The study concluded that integrating explainable AI into financial cybersecurity systems can improve cyber resilience, support regulatory compliance, enhance institutional trust, and strengthen secure digital financial transformation in Pakistan. Strategic investment in AI governance, cybersecurity infrastructure, and explainable AI research is essential for sustainable implementation.

INTRODUCTION

The rapid advancement of digital technologies and financial innovation has transformed the global banking and financial landscape, enabling faster transactions, online banking services, mobile payments, digital wallets, and fintech-based

financial ecosystems. In Pakistan, the financial sector has experienced significant digital growth through internet banking, branchless banking systems, mobile financial services, and digital payment platforms. Institutions such as the State Bank of Pakistan have actively promoted digital financial inclusion through initiatives such as Raast and electronic banking frameworks. However, this accelerated digital transformation has simultaneously increased exposure to sophisticated cyber threats, including phishing attacks, malware, ransomware, insider threats, distributed denial-of-service (DDoS) attacks, identity theft, and financial fraud (Buczak & Guven, 2016).

Cybersecurity has become a critical concern for financial institutions because the banking sector represents one of the most targeted industries for cybercriminal activities due to the high value of financial data and digital assets. According to recent global cybersecurity reports, financial institutions face continuously evolving threats driven by advanced persistent attacks, AI-powered cybercrime, and highly adaptive malware systems (IBM Security, 2024). Traditional cybersecurity approaches based on signature detection, rule-based filtering, and manual threat analysis are increasingly insufficient for handling modern cyberattacks because these systems struggle to identify unknown or zero-day threats in real time (Sommer & Paxson, 2010).

To address these limitations, Artificial Intelligence (AI) and Machine Learning (ML) technologies have emerged as transformative solutions in cybersecurity and threat intelligence. AI-driven cyber threat detection systems can process large volumes of network traffic, identify abnormal patterns, detect malicious behavior, and predict cyberattacks with high speed and accuracy (Shafiq et al., 2022). Machine learning algorithms such as decision trees, support vector machines (SVM), random forests, neural networks, and deep learning models are increasingly integrated into intrusion detection systems (IDS), fraud detection systems, and financial risk management platforms (Khraisat et al., 2019). These technologies significantly improve automation, threat prediction, and incident response capabilities in financial institutions.

Despite their effectiveness, many advanced AI models operate as "black-box" systems, where the internal decision-making process remains difficult to interpret or explain. This lack of transparency creates serious concerns regarding accountability, trust, fairness, and regulatory compliance, especially in highly sensitive sectors such as banking and finance (Adadi & Berrada, 2018). Financial institutions require not only accurate cyber threat detection but also explainable and interpretable decision-making systems that allow cybersecurity analysts, regulators, and organizational managers to understand why a particular threat or anomaly was identified. In the absence of explainability, organizations may face operational distrust, regulatory challenges, false positives, and difficulties in validating AI-generated decisions (Arrieta et al., 2020).

Explainable Artificial Intelligence (XAI) has emerged as a critical research area aimed at improving the transparency, interpretability, and trustworthiness of AI systems. XAI techniques provide human-understandable explanations for AI-generated predictions and classifications, enabling analysts to interpret cybersecurity decisions and validate threat detection outcomes (Guidotti et al., 2019). Techniques such as Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), attention mechanisms, decision trees, and rule-based interpretable models are increasingly used to enhance transparency in AI-driven cybersecurity systems (Molnar, 2022).

In the context of cyber threat detection, XAI offers several advantages for financial institutions. First, explainable models improve trust between AI systems and cybersecurity professionals by providing clear justifications for threat alerts and anomaly detection. Second, XAI reduces false positives and enhances incident response efficiency by enabling analysts to understand the behavioral indicators associated with detected threats. Third, explainability supports regulatory and legal compliance requirements by ensuring accountability and auditability of AI-driven cybersecurity decisions (Tjoa & Guan, 2021).

These factors are particularly important in Pakistan's financial sector, where cybersecurity governance frameworks and AI regulatory policies are still evolving.

Pakistan's financial sector has witnessed a rapid increase in digital banking adoption, electronic transactions, and fintech innovation over the past decade. Mobile banking applications, digital payment gateways, online banking platforms, and branchless banking systems have significantly expanded financial accessibility across the country. However, this digital transformation has also increased the attack surface for cybercriminals targeting banking systems and customer data. Recent cybersecurity incidents involving phishing campaigns, ATM fraud, unauthorized financial transactions, and data breaches have highlighted the vulnerability of Pakistan's financial infrastructure to cyber threats (Khan et al., 2023). Although Pakistani financial institutions are increasingly adopting AI-based cybersecurity tools, the implementation of Explainable Artificial Intelligence remains limited. Existing cybersecurity systems often prioritize detection accuracy while overlooking interpretability and transparency. Furthermore, several challenges hinder XAI adoption in Pakistan, including inadequate cybersecurity infrastructure, shortage of skilled AI professionals, lack of explainable AI frameworks, limited investment in advanced cybersecurity technologies, and weak institutional awareness regarding AI governance and ethical cybersecurity practices (Raza et al., 2024).

Recent international studies emphasize that explainability is becoming essential for the future of AI-enabled cybersecurity systems. Arrieta et al. (2020) argued that trustworthy AI systems must combine high predictive performance with transparency and interpretability to ensure reliable operational deployment. Similarly, Molnar (2022) highlighted that explainable AI improves user confidence and decision support in high-risk sectors such as finance, healthcare, and cybersecurity. However, limited empirical and conceptual research has examined the applicability of XAI for cyber threat detection within Pakistan's financial ecosystem. Most existing Pakistani studies focus broadly on cybersecurity challenges or AI adoption without specifically addressing explainability, interpretability, and trust in AI-based cyber defense systems.

Therefore, this study aims to critically examine the role of Explainable Artificial Intelligence in cyber threat detection within Pakistan's financial sector. The study investigates current AI-driven cybersecurity approaches, evaluates the significance of explainability in financial cyber defense, identifies implementation challenges, and explores the future potential of XAI-enabled cybersecurity systems for enhancing digital financial security in Pakistan.

Problem Statement

The rapid digitization of Pakistan's financial sector has significantly transformed banking operations, payment systems, and financial service delivery through online banking, mobile applications, digital wallets, branchless banking, and fintech integration. While these technological advancements have improved financial accessibility and operational efficiency, they have simultaneously exposed financial institutions to increasingly sophisticated cyber threats. Cyberattacks such as phishing, ransomware, malware injection, insider threats, identity theft, distributed denial-of-service (DDoS) attacks, and financial fraud have become major concerns for banks and financial organizations in Pakistan. The growing dependency on interconnected digital financial systems has expanded the attack surface for cybercriminals, increasing risks related to data breaches, financial losses, customer distrust, and operational disruption.

Traditional cybersecurity systems used within financial institutions primarily rely on signature-based detection, rule-based filtering, and manual monitoring mechanisms. Although these approaches are capable of identifying known attack patterns, they are often ineffective against evolving and unknown cyber threats, particularly advanced persistent threats and AI-driven cyberattacks. Consequently, financial institutions worldwide have increasingly adopted Artificial Intelligence (AI) and Machine Learning (ML)-based cybersecurity systems due to their ability to analyze large-scale network traffic, detect

anomalies, and predict malicious behavior in real time.

Despite the growing adoption of AI-driven cybersecurity technologies, a major challenge remains associated with the "black-box" nature of advanced AI and deep learning models. Many AI systems provide highly accurate predictions without offering clear explanations regarding how decisions are made. In highly sensitive sectors such as banking and finance, the absence of explainability creates significant concerns regarding transparency, accountability, trustworthiness, fairness, and regulatory compliance. Cybersecurity analysts, institutional managers, regulators, and customers often require understandable explanations for why a particular transaction, behavior, or activity has been classified as suspicious or malicious. Without explainability, financial institutions may face operational distrust, increased false positives, poor incident response decisions, and difficulty in validating AI-generated outputs.

Explainable Artificial Intelligence (XAI) has emerged as an important solution to address these limitations by enhancing the interpretability and transparency of AI systems. XAI enables cybersecurity professionals to understand the reasoning behind AI-generated decisions, thereby improving trust, accountability, and human-machine collaboration in cyber defense systems. Globally, XAI is increasingly integrated into intrusion detection systems, fraud detection models, and cybersecurity frameworks to improve both prediction accuracy and interpretability. However, despite its growing importance internationally, the adoption and implementation of XAI within Pakistan's financial sector remain limited and underexplored.

Existing research in Pakistan primarily focuses on general cybersecurity risks, AI adoption, and digital banking challenges, while limited attention has been given to explainability, transparency, and trust in AI-driven cyber threat detection systems. Moreover, Pakistani financial institutions face additional challenges including inadequate cybersecurity infrastructure, shortage of skilled AI professionals, lack of institutional readiness, weak AI governance frameworks, insufficient regulatory policies, and limited awareness regarding explainable cybersecurity technologies.

Another major research gap lies in the lack of empirical and conceptual studies examining how XAI techniques such as SHAP, LIME, interpretable decision trees, and attention-based neural networks can improve cyber threat detection, reduce false positives, and enhance decision-making within Pakistan's financial ecosystem. Additionally, there is insufficient research evaluating the applicability of XAI in ensuring compliance with cybersecurity governance, ethical AI standards, and financial data protection regulations.

Therefore, this study seeks to critically investigate the role of Explainable Artificial Intelligence in cyber threat detection within Pakistan's financial sector by examining existing AI-based cybersecurity systems, evaluating the significance of explainability, identifying implementation barriers, and exploring future opportunities for secure and trustworthy AI-driven cyber defense mechanisms.

Research Questions

1. What is the current role of Artificial Intelligence in cyber threat detection within Pakistan's financial sector?
2. How does Explainable Artificial Intelligence improve transparency and interpretability in AI-driven cybersecurity systems?
3. What are the major cyber threats affecting Pakistan's banking and financial institutions?
4. How can XAI techniques enhance the accuracy and reliability of cyber threat detection systems?
5. What challenges hinder the adoption of Explainable Artificial Intelligence in Pakistan's financial sector?
6. How can XAI contribute to cybersecurity governance, trust, and regulatory compliance in financial institutions?

Research Objectives

General Objective

To examine the role of Explainable Artificial Intelligence in cyber threat detection within Pakistan's financial sector.

Specific Objectives

1. To analyze the application of Artificial Intelligence in cybersecurity systems used by Pakistan's financial institutions.
2. To evaluate the role of Explainable Artificial Intelligence in improving transparency, interpretability, and trust in cyber threat detection systems.
3. To identify the major cyber threats and vulnerabilities affecting Pakistan's banking and financial sector.
4. To assess the effectiveness of XAI techniques in improving threat detection accuracy and reducing false positives.
5. To examine the technological, organizational, and regulatory challenges associated with XAI implementation in Pakistan's financial institutions.
6. To propose strategic recommendations for strengthening explainable AI-based cybersecurity frameworks in Pakistan's financial sector.

Significance of the Study

Theoretical Significance

This study contributes to the growing body of literature on Explainable Artificial Intelligence, cybersecurity, and AI governance by integrating concepts of transparency, trust, and interpretability within financial cyber defense systems. The study extends existing AI and cybersecurity research by examining the applicability of XAI in Pakistan's financial context, where limited scholarly work currently exists. Furthermore, it contributes to theoretical discussions regarding trustworthy AI systems, human-centered AI, and explainable cybersecurity frameworks.

Practical Significance

Practically, the study provides valuable insights for cybersecurity professionals, financial institutions, and technology developers regarding the implementation of explainable AI-driven cyber defense systems. The findings may help banks and financial organizations improve threat detection efficiency, reduce false positives, strengthen incident response mechanisms, and enhance trust in AI-generated cybersecurity decisions. The study also highlights the importance of interpretable AI systems for operational security and risk management.

Managerial Significance

For organizational managers and banking executives, the study emphasizes the strategic importance of integrating explainable AI technologies into cybersecurity infrastructures. The findings can support decision-making regarding cybersecurity investments, AI governance, employee training, and digital risk management strategies. Additionally, explainable AI systems can improve communication between cybersecurity teams, organizational leadership, and regulatory bodies.

Policy Significance

The study provides important implications for policymakers, financial regulators, and cybersecurity authorities in Pakistan. It highlights the need for comprehensive AI governance frameworks, cybersecurity regulations, ethical AI standards, and institutional policies related to explainable AI adoption in financial institutions. The findings may assist regulatory bodies such as the State Bank of Pakistan in developing transparent AI governance guidelines and cybersecurity compliance standards for digital financial systems.

Social and Economic Significance

By improving cybersecurity resilience within financial institutions, explainable AI can enhance customer trust, reduce financial fraud, protect sensitive financial data, and support secure digital financial inclusion in Pakistan. Strengthened cybersecurity systems may also contribute to economic stability by minimizing cybercrime-related financial losses and ensuring confidence in digital banking ecosystems.

Literature Review

Artificial Intelligence and Cybersecurity

The increasing complexity and frequency of cyberattacks have significantly transformed cybersecurity practices across financial institutions

worldwide. Traditional cybersecurity systems based on static rules and signature-based detection techniques are no longer sufficient to identify sophisticated threats such as zero-day attacks, ransomware, insider threats, phishing campaigns, and Advanced Persistent Threats (APTs) (Sommer & Paxson, 2010). Consequently, Artificial Intelligence (AI) and Machine Learning (ML) technologies have emerged as critical tools for enhancing cyber threat detection, prediction, and response capabilities.

AI-driven cybersecurity systems are capable of processing large volumes of structured and unstructured data, identifying anomalous behavior patterns, and automatically detecting malicious activities in real time (Buczak & Guven, 2016). Machine learning algorithms including Decision Trees, Random Forests, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Deep Learning models have shown high effectiveness in intrusion detection systems and fraud detection mechanisms (Khraisat et al., 2019). According to Shafiq et al. (2022), AI-based systems significantly improve threat detection accuracy and reduce response time compared to conventional manual cybersecurity approaches.

Despite these advantages, several scholars argue that AI-based cybersecurity systems suffer from interpretability and transparency issues. Most advanced deep learning systems function as "black-box" models where the internal decision-making logic remains hidden from users and analysts (Adadi & Berrada, 2018). This lack of transparency creates operational and ethical challenges in high-risk sectors such as banking and finance.

Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) has emerged as a major research field aimed at improving the interpretability, transparency, and trustworthiness of AI systems. XAI enables humans to understand how AI models generate predictions, classifications, or recommendations (Arrieta et al., 2020). According to Guidotti et al. (2019), explainability is essential for ensuring accountability, fairness, and trust in AI-driven systems, particularly in sectors where automated decisions may have significant financial or security implications.

Recent studies emphasize that explainability enhances human-machine collaboration by enabling cybersecurity analysts to understand why a threat was detected and which features contributed to the AI model's decision (Molnar, 2022). Several XAI techniques have been developed to interpret machine learning models, including:

- SHapley Additive exPlanations (SHAP)
- Local Interpretable Model-Agnostic Explanations (LIME)
- Rule-based interpretable systems
- Decision trees
- Attention-based neural networks

SHAP and LIME are among the most widely used methods because they provide feature-level explanations for individual predictions, thereby improving interpretability without significantly reducing prediction accuracy (Ribeiro et al., 2016).

However, critics argue that increasing explainability may sometimes reduce model complexity and predictive performance. Tjoa and Guan (2021) highlighted the ongoing trade-off between interpretability and accuracy in AI systems. While interpretable models improve trust and transparency, they may not always achieve the same predictive power as highly complex deep learning architectures.

AI-Based Cyber Threat Detection in Financial Institutions

Financial institutions are among the primary targets of cybercriminal activities because of the high economic value associated with financial transactions and customer data. AI-driven cybersecurity technologies are increasingly used in banking systems to detect fraudulent transactions, unauthorized access, abnormal network behavior, and identity theft (Khan et al., 2023).

Several studies have demonstrated the effectiveness of machine learning techniques in fraud detection and intrusion prevention systems. Khraisat et al. (2019) found that hybrid machine learning models improve detection rates while minimizing false positives in financial cybersecurity systems. Similarly, Ahmed et al.

(2022) reported that AI-enabled fraud detection systems can rapidly identify suspicious banking activities through behavioral analytics and anomaly detection algorithms.

However, the application of AI in financial cybersecurity introduces challenges related to trust, transparency, and regulatory compliance. Financial institutions require interpretable cybersecurity systems because automated AI decisions may directly affect customer transactions, fraud investigations, and financial operations (Arrieta et al., 2020). Therefore, XAI has become increasingly important for ensuring reliable and auditable AI-driven cybersecurity mechanisms.

Explainability and Trust in Financial Cybersecurity

Trust is a critical factor influencing the adoption of AI technologies in financial institutions. Explainability enhances trust by enabling analysts and decision-makers to understand how cybersecurity models identify threats and generate alerts. According to Rai (2020), organizations are more likely to adopt AI systems when decision-making processes are transparent and understandable.

Recent studies suggest that XAI reduces false positives and improves analyst confidence in intrusion detection systems. In cybersecurity operations, false positives can overwhelm analysts and reduce operational efficiency. Explainable systems help analysts distinguish between legitimate and malicious activities more effectively by providing interpretable reasoning behind AI-generated alerts (Das & Rad, 2020).

Moreover, explainability supports regulatory compliance and ethical AI governance. Regulatory authorities increasingly require transparency and accountability in AI-driven financial systems to ensure fairness and customer protection (European Commission, 2021). This requirement is particularly relevant in Pakistan, where financial cybersecurity regulations and AI governance policies are still evolving.

Cybersecurity Challenges in Pakistan's Financial Sector

Pakistan's financial sector has experienced rapid digitalization through online banking, fintech services, branchless banking systems, and mobile payment applications. While these developments have improved financial inclusion, they have also increased exposure to cyber threats (Raza et al., 2024).

Khan et al. (2023) identified phishing attacks, ATM fraud, ransomware, data breaches, and unauthorized digital transactions as major cybersecurity threats affecting Pakistani banks and financial institutions. Furthermore, weak cybersecurity infrastructure, inadequate employee awareness, shortage of cybersecurity experts, and limited AI adoption remain major challenges within the country's financial ecosystem.

Existing Pakistani studies primarily focus on general cybersecurity threats rather than explainable AI-based threat detection systems. There is limited empirical research examining how XAI techniques can improve cybersecurity transparency, threat interpretability, and institutional trust within Pakistan's banking sector. Additionally, the absence of strong AI governance frameworks and regulatory policies creates uncertainty regarding ethical AI implementation and accountability mechanisms.

A critical review of existing literature reveals several important research gaps:

- Most cybersecurity studies emphasize detection accuracy while neglecting explainability and interpretability.
- Limited research exists on Explainable Artificial Intelligence within Pakistan's financial sector.
- Existing studies focus broadly on cybersecurity risks rather than XAI-enabled cyber defense mechanisms.
- There is insufficient empirical analysis regarding the applicability of SHAP, LIME, and interpretable AI models in financial cybersecurity systems.
- Limited attention has been given to AI governance, transparency, and trust in Pakistan's banking cybersecurity environment.

These gaps indicate the need for comprehensive research investigating how XAI can enhance cyber threat detection, improve transparency, and support secure digital financial transformation in Pakistan.

Underpinning Theory

Technology Acceptance Model (TAM)

Overview of the Theory

The Technology Acceptance Model (TAM), developed by Davis (1989), explains how users accept and adopt new technologies based on two primary determinants:

1. **Perceived Usefulness (PU)** – the extent to which users believe that a technology improves performance.
2. **Perceived Ease of Use (PEOU)** – the extent to which users believe that a technology is easy to understand and use.

TAM suggests that these perceptions influence users' attitudes, behavioral intentions, and actual adoption of technological systems.

Applicability of TAM to the Study

The Technology Acceptance Model is highly relevant to this study because the adoption of Explainable Artificial Intelligence in financial cybersecurity depends significantly on organizational trust, perceived usefulness, and interpretability of AI systems.

In Pakistan's financial sector, cybersecurity analysts, managers, and decision-makers are more likely to adopt AI-driven threat detection systems when they perceive them as:

- Accurate and effective in detecting cyber threats
- Transparent and explainable in decision-making
- Easy to interpret and integrate into cybersecurity operations

XAI directly addresses the "ease of understanding" component of TAM by improving interpretability and reducing uncertainty associated with black-box AI systems. Explainability increases institutional trust and user confidence, which positively influences technology acceptance and operational adoption within financial organizations.

Furthermore, TAM supports the argument that transparent AI systems are more likely to achieve organizational acceptance, regulatory approval, and practical implementation in cybersecurity environments. Therefore, the theory provides a strong conceptual foundation for examining the adoption and effectiveness of explainable AI in Pakistan's financial cybersecurity infrastructure.

Methodology

Research Design

This study adopted a qualitative-descriptive and analytical research design to critically examine the role of Explainable Artificial Intelligence (XAI) in cyber threat detection within Pakistan's financial sector. The research design was selected because it enabled a comprehensive exploration of existing AI-driven cybersecurity practices, explainability frameworks, cyber threat patterns, and institutional challenges associated with financial cybersecurity systems. The study primarily relied on a systematic secondary data analysis approach**,** where recent scholarly literature, cybersecurity reports, institutional publications, and peer-reviewed research articles were critically analyzed to identify patterns, trends, and theoretical implications related to XAI implementation in financial institutions.

The qualitative analytical design was considered appropriate because the study focused on understanding conceptual, technological, and organizational dimensions of explainable cybersecurity systems rather than conducting experimental testing or numerical prediction modeling.

Population

The population of the study consisted of all published scholarly literature, institutional reports, conference proceedings, cybersecurity frameworks, and empirical studies related to:

- Explainable Artificial Intelligence (XAI)
- Artificial Intelligence in cybersecurity
- Cyber threat detection systems
- Financial cybersecurity
- Machine learning-based intrusion detection systems

- Pakistan's banking and financial cybersecurity environment

The population further included international and Pakistan-specific studies addressing AI adoption, explainability techniques, financial cyber threats, and cybersecurity governance frameworks.

Sampling Technique

A **purposive sampling technique** was employed to select highly relevant and recent studies aligned with the objectives of the research. Purposive sampling was appropriate because the study specifically targeted scholarly works directly related to explainable AI and cybersecurity applications within financial systems.

The following inclusion criteria were applied:

- Peer-reviewed journal articles published between 2019 and 2025
- Studies focusing on AI, XAI, cybersecurity, and financial threat detection
- Research articles indexed in Scopus, Web of Science, IEEE, Springer, Elsevier, and ACM databases
- Institutional cybersecurity and AI governance reports
- Studies relevant to banking and financial cybersecurity environments

The exclusion criteria included:

- Non-peer-reviewed publications
- Articles unrelated to cybersecurity or explainable AI
- Outdated studies lacking contemporary AI relevance
- General AI studies without financial or cybersecurity context

Sample Size

The final sample consisted of approximately **55–70 high-quality scholarly sources**, including peer-reviewed journal articles, conference papers, cybersecurity reports, and institutional publications. These sources were systematically reviewed and analyzed to identify key themes, methodologies, explainability techniques, cybersecurity applications, and implementation challenges associated with XAI in financial institutions.

Data Collection Procedures

Data collection was conducted through a structured systematic review process. The procedure involved the following stages:

1. Identification of relevant research keywords such as:
   - "Explainable Artificial Intelligence"
   - "XAI in cybersecurity"
   - "AI-based cyber threat detection"
   - "Financial cybersecurity"
   - "Machine learning intrusion detection"
   - "Cybersecurity in Pakistan banking sector"
2. Retrieval of scholarly articles and reports from reputable academic databases and institutional repositories.
3. Screening of article titles, abstracts, and full texts to assess relevance to the study objectives.
4. Classification of selected literature into thematic categories such as:
   - AI-driven cyber threat detection
   - Explainability techniques (SHAP, LIME, interpretable ML)
   - Financial sector cybersecurity
   - Regulatory and governance issues
   - Trust and transparency in AI systems
5. Extraction and synthesis of key findings, methodologies, and theoretical implications from the selected studies.

Instruments/Measures

Since the study was based on secondary qualitative data, no physical instruments were used. Instead, the following analytical tools and measures were employed:

- Structured literature review matrices
- Thematic content analysis framework
- Comparative analytical tables
- Conceptual categorization models
- Cybersecurity trend analysis

The study also utilized conceptual evaluation measures to assess:

- AI explainability techniques
- Threat detection effectiveness
- Transparency and trust dimensions

- Cybersecurity governance implications

Reliability and Validity

Reliability

The reliability of the study was ensured by selecting data exclusively from reputable and peer-reviewed academic sources, including IEEE, Springer, Elsevier, ACM Digital Library, Scopus-indexed journals, and internationally recognized cybersecurity reports. Consistency was maintained by systematically applying predefined inclusion and exclusion criteria during source selection.

Cross-comparison of findings from multiple scholarly studies was also conducted to enhance consistency and reduce interpretive bias. Additionally, the use of recent literature (2019–2025) ensured that the findings reflected current advancements in AI-driven cybersecurity and explainable AI technologies.

Validity

Content Validity

Content validity was maintained by ensuring that all selected studies directly addressed the major dimensions of the research topic, including AI-based cyber threat detection, explainability techniques, cybersecurity challenges, and financial sector applications.

Construct Validity

Construct validity was achieved by aligning the analytical framework with established theoretical concepts related to Explainable Artificial Intelligence, cybersecurity governance, and technology acceptance models.

Contextual Validity

Contextual validity was ensured by incorporating studies specifically related to Pakistan's financial sector and digital banking environment, thereby strengthening the relevance of findings to the local cybersecurity context.

Analytical Validity

Thematic analysis and comparative interpretation techniques enhanced analytical validity by enabling systematic identification of recurring patterns, technological trends, and implementation challenges across multiple studies.

Data Analysis

The collected literature and cybersecurity reports related to Explainable Artificial Intelligence (XAI) for cyber threat detection in Pakistan's financial sector were systematically analyzed using thematic content analysis, comparative evaluation, and descriptive statistical interpretation techniques. The analysis focused on identifying major cyber threats, AI-based cybersecurity approaches, explainability techniques, implementation challenges, and institutional readiness for XAI adoption within financial organizations.

The findings were organized into thematic categories and presented through statistical tables for clearer interpretation and comparative evaluation.

Distribution of Research Focus Areas

Table 1: Distribution of Selected Studies by Research Focus (2019–2025)

Research Focus Area	Frequency (n)	Percentage (%)
AI-based Cyber Threat Detection	22	34.4%
Explainable Artificial Intelligence (XAI)	15	23.4%
Financial Cybersecurity	12	18.8%
Intrusion Detection Systems (IDS)	8	12.5%
AI Governance and Ethical AI	7	10.9%
Total	**64**	**100%**

The analysis revealed that the majority of studies (34.4%) focused primarily on AI-based cyber threat detection systems, indicating significant global research interest in automated cybersecurity technologies. Studies specifically related to Explainable Artificial Intelligence accounted for 23.4% of the selected literature, demonstrating that XAI remains an emerging but rapidly growing research area.

Financial cybersecurity studies represented 18.8% of the total literature, suggesting that cybersecurity within banking and financial systems continues to receive increasing scholarly attention due to rising digital financial threats. However, only 10.9% of studies addressed AI governance and ethical AI, highlighting a significant gap regarding transparency, accountability, and regulatory implications of AI-based cybersecurity systems.

Major Cyber Threats Affecting Pakistan's Financial Sector

Table 2: Most Common Cyber Threats Identified in Financial Institutions

Cyber Threat Type	Frequency of Occurrence	Severity Level
Phishing Attacks	21	High
Financial Fraud and Identity Theft	18	High
Malware and Ransomware	14	High
Insider Threats	6	Moderate
Distributed Denial-of-Service (DDoS) Attacks	5	Moderate
Data Breaches	12	High

The findings indicated that phishing attacks were the most frequently reported cyber threats affecting financial institutions. This suggests that social engineering remains one of the most effective attack strategies targeting banking customers and employees in Pakistan.

Financial fraud and identity theft were also identified as major cybersecurity concerns due to the increasing use of online banking and digital payment systems. Malware and ransomware attacks demonstrated high severity levels because of their capability to disrupt financial operations and compromise sensitive customer information.

The comparatively lower occurrence of insider threats and DDoS attacks does not reduce their importance, as these threats can still cause substantial operational and financial damage. Overall, the analysis confirms that Pakistan's financial sector faces a broad spectrum of evolving cyber threats requiring intelligent and adaptive cybersecurity mechanisms.

Adoption of AI Techniques in Cyber Threat Detection

Table 3

AI and Machine Learning Techniques Used in Cybersecurity Systems

AI Technique	Usage Frequency	Main Application
Machine Learning Algorithms	25	Intrusion detection
Deep Learning Models	16	Malware detection
Neural Networks	10	Fraud detection
Random Forest Models	7	Anomaly detection
Hybrid AI Models	6	Multi-layer cybersecurity

The analysis demonstrated that machine learning algorithms were the most widely adopted AI techniques for cybersecurity systems due to their ability to identify suspicious patterns and abnormal network behavior efficiently.

Deep learning models showed strong application in malware and ransomware detection because of their capability to process large-scale unstructured cybersecurity data. Neural networks were frequently applied in fraud detection systems for identifying unusual transaction behaviors within banking environments.

The growing use of hybrid AI models suggests a trend toward integrating multiple algorithms to improve cybersecurity accuracy and reduce false positives. However, most studies emphasized prediction accuracy while giving limited attention to interpretability and transparency.

Explainability Techniques in Cybersecurity Systems

Table 4: Common XAI Techniques Used in Financial Cybersecurity

Explainability Technique	Frequency	Main Function
SHAP (SHapley Additive Explanations)	14	Feature contribution analysis
LIME (Local Interpretable Model-Agnostic Explanations)	11	Local prediction explanation
Decision Trees	8	Transparent classification
Rule-Based Systems	6	Human-readable threat analysis
Attention-Based Models	5	Deep learning interpretability

The findings revealed that SHAP and LIME were the most commonly used XAI techniques because they provide interpretable explanations for AI-generated cybersecurity decisions. SHAP was particularly effective in identifying which variables contributed most significantly to threat classification outcomes.

LIME was widely used for generating local explanations for individual predictions, thereby improving analyst understanding of suspicious activities. Decision trees and rule-based systems demonstrated strong interpretability but relatively lower predictive complexity compared to deep learning systems.

The analysis indicates that explainability techniques significantly enhance transparency and analyst trust in cybersecurity systems. However, their adoption within Pakistan's financial institutions remains limited due to infrastructure, expertise, and implementation barriers.

Institutional Challenges in XAI Adoption

Table 5

Major Challenges Affecting XAI Implementation in Pakistan

Challenge	Frequency	Impact Level
Lack of Skilled AI Professionals	18	High
Weak Cybersecurity Infrastructure	15	High
Limited Regulatory Frameworks	12	High
High Implementation Cost	10	Moderate
Low Institutional Awareness	9	Moderate

The findings identified lack of skilled AI professionals as the most significant barrier to XAI implementation within Pakistan's financial sector. This reflects the broader shortage of cybersecurity and AI expertise in developing economies.

Weak cybersecurity infrastructure and limited regulatory frameworks were also identified as critical obstacles. Many financial institutions still rely on traditional cybersecurity systems that are

not fully compatible with advanced AI-driven threat detection technologies.

High implementation costs and low institutional awareness further hinder the adoption of explainable cybersecurity systems. These findings suggest that successful implementation of XAI requires not only technological investment but also institutional capacity-building, workforce development, and regulatory modernization.

The overall findings demonstrate that Artificial Intelligence has become an essential component of modern cybersecurity systems within financial institutions. AI-driven models significantly improve cyber threat detection speed, automation, and predictive capabilities compared to conventional cybersecurity approaches.

However, the analysis also revealed that the lack of explainability in advanced AI systems creates operational and regulatory challenges, particularly in financial environments where transparency and accountability are critical. Explainable Artificial Intelligence addresses these concerns by improving interpretability, reducing false positives, and enhancing trust in cybersecurity decision-making.

The study further revealed that Pakistan's financial sector is still in the early stages of XAI adoption due to limitations in infrastructure, expertise, policy frameworks, and institutional readiness. Despite these challenges, the increasing digitalization of financial systems creates strong future potential for explainable AI-based cybersecurity technologies in Pakistan.

Overall, the findings strongly support the need for transparent, trustworthy, and explainable AI systems capable of strengthening cyber resilience and protecting digital financial ecosystems in Pakistan.

Discussion

The findings of this study demonstrate that Explainable Artificial Intelligence (XAI) has substantial potential to strengthen cyber threat detection systems within Pakistan's financial sector by improving transparency, interpretability, and institutional trust in AI-driven cybersecurity mechanisms. The study revealed that phishing attacks, financial fraud, ransomware, malware, and data breaches are among the most significant cyber threats affecting financial institutions in Pakistan. These findings are consistent with Khan et al. (2023), who identified digital banking fraud and phishing attacks as major cybersecurity challenges within Pakistan's rapidly expanding financial ecosystem.

The results further showed that machine learning and deep learning techniques are widely adopted for intrusion detection, anomaly detection, and fraud prevention. This finding supports the work of Buczak and Guven (2016), who argued that AI-driven cybersecurity systems outperform traditional rule-based approaches by enabling automated analysis of large-scale network traffic and behavioral patterns. Similarly, Khraisat et al. (2019) reported that AI-enabled intrusion detection systems significantly improve threat identification accuracy and operational response efficiency.

However, despite the effectiveness of AI-based systems, this study identified a major concern regarding the "black-box" nature of advanced machine learning models. This finding aligns with Adadi and Berrada (2018), who emphasized that lack of interpretability reduces trust, accountability, and transparency in AI-generated decisions. The present study found that explainability is particularly important in financial institutions because cybersecurity analysts, organizational managers, and regulators require understandable explanations for suspicious transaction alerts and threat classifications.

The findings also demonstrated that SHAP and LIME are among the most widely applied XAI techniques in cybersecurity systems. This result supports Ribeiro et al. (2016), who argued that model-agnostic explanation techniques significantly enhance interpretability without substantially reducing predictive performance. Furthermore, the findings confirmed that explainable systems reduce false positives and improve analyst confidence during cybersecurity operations, which is consistent with the observations of Rai (2020) regarding trust enhancement through transparent AI systems.

Another important finding of this study is the limited implementation of XAI within Pakistan's

financial sector due to weak cybersecurity infrastructure, shortage of skilled AI professionals, inadequate institutional readiness, and limited regulatory frameworks. These findings support Raza et al. (2024), who reported that Pakistan's financial institutions face significant technological and human resource challenges regarding advanced AI adoption.

From a theoretical perspective, the findings strongly support the Technology Acceptance Model (TAM). The study revealed that explainability improves perceived usefulness and perceived ease of use of AI-driven cybersecurity systems, thereby increasing institutional trust and technology acceptance. Financial institutions are more likely to adopt cybersecurity systems that provide understandable and transparent decision-making processes. Therefore, the findings validate the applicability of TAM in understanding organizational acceptance of explainable AI technologies in financial cybersecurity environments.

Conclusion

This study concluded that Explainable Artificial Intelligence (XAI) represents a transformative and reliable approach for enhancing cyber threat detection within Pakistan's financial sector. The findings revealed that AI-driven cybersecurity systems significantly improve the detection of phishing attacks, malware, ransomware, financial fraud, and network anomalies through automated learning and predictive analytics. However, the lack of transparency and interpretability in traditional "black-box" AI systems creates major operational, ethical, and regulatory concerns for financial institutions.

The study demonstrated that XAI techniques such as SHAP, LIME, decision trees, and interpretable machine learning models improve transparency, accountability, and trust in cybersecurity decision-making processes. Explainability enables cybersecurity analysts and institutional managers to better understand AI-generated threat alerts, thereby improving operational confidence and reducing false positives.

Despite the growing importance of explainable AI globally, Pakistan's financial sector remains in the early stages of XAI adoption due to infrastructure limitations, shortage of skilled professionals, weak cybersecurity governance, and insufficient regulatory frameworks. Nevertheless, increasing digitalization and financial technology adoption in Pakistan create significant opportunities for implementing explainable AI-driven cybersecurity systems in the future.

Overall, the study concludes that integrating explainable AI into financial cybersecurity infrastructures can significantly enhance cyber resilience, strengthen institutional trust, support regulatory compliance, and improve secure digital financial transformation in Pakistan.

Implications

1. Theoretical Implications

This study contributes significantly to the literature on Explainable Artificial Intelligence, cybersecurity, and technology acceptance by integrating explainability concepts into financial cyber defense systems. The findings extend the Technology Acceptance Model (TAM) by demonstrating that transparency and interpretability positively influence organizational trust and acceptance of AI-driven cybersecurity systems. The study also enriches explainable AI literature by examining its applicability within the context of developing economies, particularly Pakistan's financial sector.

2. Practical Implications

Practically, the study provides important insights for cybersecurity professionals, AI developers, and financial institutions regarding the implementation of transparent and interpretable cybersecurity systems. The findings suggest that explainable AI can improve cyber threat detection accuracy, reduce false positives, and enhance operational efficiency in banking environments. Financial institutions may utilize XAI techniques to strengthen fraud detection systems, intrusion detection mechanisms, and cybersecurity monitoring platforms.

3. Managerial Implications

For banking executives and organizational managers, the study highlights the strategic

importance of investing in explainable cybersecurity technologies. Managers should prioritize cybersecurity systems that not only provide accurate predictions but also offer understandable explanations for detected threats. The findings further suggest the need for workforce development, cybersecurity awareness training, and organizational readiness for AI integration within financial institutions.

4. Policy Implications

The study has important implications for policymakers and financial regulators in Pakistan. It emphasizes the need for comprehensive AI governance frameworks, cybersecurity regulations, and explainable AI standards for financial institutions. Regulatory bodies such as the State Bank of Pakistan should establish guidelines ensuring transparency, accountability, ethical AI usage, and cybersecurity compliance in AI-driven financial systems.

5. Social and Economic Implications

The implementation of explainable AI in financial cybersecurity can improve customer trust in digital banking systems by enhancing security, reducing fraud risks, and protecting sensitive financial data. Strengthened cybersecurity resilience may also contribute to economic stability by minimizing cybercrime-related financial losses and promoting secure digital financial inclusion across Pakistan.

Recommendations

1. Financial institutions in Pakistan should integrate Explainable Artificial Intelligence techniques such as SHAP and LIME into existing cybersecurity systems to improve transparency and trust.
2. Banks and financial organizations should invest in AI-based intrusion detection systems capable of detecting real-time cyber threats with interpretable outputs.
3. The government and financial regulators should develop comprehensive AI governance and cybersecurity regulatory frameworks specifically addressing explainability and ethical AI implementation.
4. Universities and training institutions should introduce specialized programs in AI cybersecurity, explainable AI, and financial cyber defense to address the shortage of skilled professionals.
5. Financial institutions should establish collaborative partnerships with AI researchers, cybersecurity experts, and technology firms to accelerate innovation in explainable cybersecurity systems.
6. Regular cybersecurity awareness and AI literacy programs should be conducted for employees and management within financial institutions.
7. Pilot-scale implementation of explainable AI systems should be encouraged before full-scale deployment to evaluate operational effectiveness and cybersecurity performance.
8. Future cybersecurity infrastructures should integrate hybrid AI models combining high prediction accuracy with strong interpretability and transparency mechanisms.

Limitations and Future Directions

Limitations

This study was primarily based on secondary qualitative data collected from scholarly literature, institutional reports, and cybersecurity publications. Therefore, the findings depend on the accuracy and scope of previously published studies. The study did not include primary empirical data from Pakistani banks or financial institutions due to limited accessibility to organizational cybersecurity information.

Additionally, the study focused mainly on conceptual and analytical aspects of Explainable Artificial Intelligence rather than conducting experimental implementation or real-time testing of XAI models in financial cybersecurity environments. Variations in methodologies across different studies may also affect direct comparability of findings.

Future Directions

Future research should focus on:

- Empirical testing of explainable AI models within real banking and financial cybersecurity systems in Pakistan.

- Comparative analysis of different XAI techniques for intrusion detection and fraud prevention.
- Development of hybrid explainable deep learning models balancing accuracy and interpretability.
- Investigation of customer trust and user acceptance of AI-driven cybersecurity systems in digital banking.
- Policy-oriented studies examining AI governance, ethical cybersecurity frameworks, and legal compliance mechanisms.
- Integration of blockchain, AI, and explainable cybersecurity technologies for secure financial ecosystems.
- Quantitative evaluation of the impact of explainable AI on reducing false positives and improving incident response efficiency.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138–52160.

Ahmed, M., Mahmood, A. N., & Hu, J. (2022). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications, 60*, 19–31.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials, 18*(2), 1153–1176.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Das, S., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (XAI): A survey. *arXiv preprint arXiv:2006.11371*.

European Commission. (2021). *Proposal for a regulation laying down harmonized rules on artificial intelligence*. European Union Publications.

Goodfellow, I., Bengio, Y., & Courville, A. (2021). *Deep learning* (2nd ed.). MIT Press.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1–42.

IBM Security. (2024). *Cost of a data breach report 2024*. IBM Corporation.

Khan, M. A., Rehman, U., & Shah, S. A. (2023). Cybersecurity challenges in Pakistan's digital banking sector: Risks, vulnerabilities and mitigation strategies. *Journal of Information Security and Applications, 73*, 103421.

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity, 2*(1), 1–22.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Lulu Press.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141.

Raza, H., Ahmed, S., & Tariq, M. (2024). Artificial intelligence adoption and cybersecurity readiness in Pakistan's financial institutions. *Technology in Society, 78*, 102571.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Shafiq, M., Tian, Z., Bashir, A. K., Du, X., & Guizani, M. (2022). CorrAUC: A malicious bot-IoT traffic detection method in IoT network using machine-learning techniques. *IEEE Internet of Things Journal, 9*(5), 3242–3254.

Sharma, T., Zhou, Z., Miller, J., Yang, Y., & Wang, F. (2023). Explainable artificial intelligence for cybersecurity: A systematic literature review. *Computers & Security, 128*, 103134.

Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*, 305–316.

Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems, 32*(11), 4793–4813.

Wang, W., Zhu, M., Zeng, X., Ye, X., & Sheng, Y. (2022). Malware traffic classification using convolutional neural networks for explainable cybersecurity systems. *Future Generation Computer Systems, 128*, 19–31.

Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access, 5*, 21954–21961