Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

EXPLAINABLE FEDERATED ARTIFICIAL INTELLIGENCE FOR PRIVACY-PRESERVING CYBERSECURITY AND CRITICAL INFRASTRUCTURE PROTECTION IN PAKISTAN

Adnan Hassnain[*1], Engr. Zubair Ahmed[2], Dr. Muhammad Umer[3]

[*1]*Student, SEECS, CS, NUST H12 Islamabad*
[2]*Lecturer, Department of Information Technology, International Institute of Science, Art and Technology (IISAT), Gujranwala*
[3]*Associate Professor, Department of Computer Sciences, University of Peshawar*

[*1]hassnain.bscs21seecs@seecs.edu.pk, [2]zubair.ahmed@iisat.com, [3]muhammad.umer@uop.edu.pk

DOI:https://doi.org/10.5281/zenodo.20845735

Keywords
Explainable Artificial Intelligence (XAI), Federated Artificial Intelligence, Cybersecurity, Critical Infrastructure Protection, Cyber Resilience, Privacy-Preserving AI.

Article History
Received: 24 April 2026
Accepted: 06 June 2026
Published: 21 June 2026

Corresponding Author: *
Adnan Hassnain

Abstract

The increasing frequency and sophistication of cyber threats pose significant challenges to the security and resilience of critical infrastructure systems worldwide. In Pakistan, the growing digitalization of sectors such as energy, telecommunications, finance, transportation, and public services has heightened the need for advanced cybersecurity solutions that ensure both effective threat detection and data privacy. This study examines the role of Explainable Federated Artificial Intelligence (EFAI) in enhancing privacy-preserving cybersecurity and critical infrastructure protection in Pakistan. Drawing upon the Technology–Organization–Environment (TOE) Framework, the study proposes and empirically tests a model linking Federated Artificial Intelligence, Cyber Threat Detection Effectiveness, Explainable Artificial Intelligence, and Critical Infrastructure Cyber Resilience. A quantitative cross-sectional research design was employed, and data were collected from cybersecurity professionals and technology experts working in critical infrastructure sectors. The findings indicate that Federated Artificial Intelligence significantly improves cyber threat detection capabilities and enhances organizational cyber resilience. The results further reveal that Cyber Threat Detection Effectiveness mediates the relationship between Federated Artificial Intelligence and Cyber Resilience, while Explainable Artificial Intelligence strengthens this relationship by increasing transparency, interpretability, and trust in AI-driven cybersecurity decisions. The study contributes to the emerging literature on trustworthy and privacy-preserving artificial intelligence and provides practical insights for organizations and policymakers seeking to strengthen national cybersecurity capabilities. The findings underscore the strategic importance of integrating federated learning and explainable AI technologies to develop resilient, secure, and privacy-conscious critical infrastructure systems.

INTRODUCTION

The rapid digital transformation of economies and societies has significantly increased reliance on interconnected information systems, cloud infrastructures, Internet of Things (IoT) devices, and artificial intelligence (AI)-enabled

technologies. While digitalization has enhanced efficiency and innovation, it has simultaneously expanded the cyber threat landscape, exposing critical infrastructure sectors to sophisticated cyberattacks. Critical infrastructure—including energy systems, telecommunications networks, transportation systems, healthcare services, financial institutions, and government information systems—constitutes the backbone of national security and economic stability. Disruptions to these infrastructures can result in substantial economic losses, operational failures, and threats to public safety (Sharma et al., 2023).

Artificial intelligence has emerged as a transformative technology in cybersecurity due to its ability to analyze massive datasets, identify anomalies, predict cyber threats, and automate security responses. Machine learning and deep learning models have demonstrated remarkable effectiveness in intrusion detection, malware classification, network traffic analysis, and threat intelligence generation (Nguyen & Reddi, 2022). However, conventional AI-based cybersecurity systems largely depend on centralized data collection and processing architectures. Such approaches require organizations to share sensitive operational data with central servers, thereby creating privacy concerns, increasing vulnerability to data breaches, and potentially violating regulatory and data sovereignty requirements (Kairouz et al., 2021).

To address these challenges, Federated Learning (FL) has emerged as a promising paradigm for privacy-preserving artificial intelligence. Federated learning enables multiple organizations to collaboratively train machine learning models without transferring raw data to a centralized repository. Instead, only model parameters or gradients are exchanged, thereby preserving data privacy while enabling collective intelligence development (McMahan et al., 2017). In cybersecurity applications, federated learning offers significant advantages by allowing organizations to share threat intelligence and improve detection capabilities without exposing confidential information. This capability is particularly valuable for critical infrastructure sectors where data sensitivity, operational confidentiality, and regulatory compliance are paramount considerations (Li et al., 2023).

Despite its advantages, federated learning introduces new challenges related to transparency, interpretability, and trustworthiness. Many AI-based cybersecurity systems operate as "black boxes," generating predictions and recommendations without providing understandable explanations regarding how decisions are reached. Such opacity can hinder cybersecurity analysts, managers, and policymakers from trusting AI-generated outputs, particularly in high-stakes environments involving national security and critical infrastructure protection (Arrieta et al., 2020). Consequently, the growing field of Explainable Artificial Intelligence (XAI) seeks to improve the interpretability and transparency of AI systems by providing understandable explanations for model decisions, enabling users to assess reliability, fairness, and accountability (Adadi & Berrada, 2018).

The integration of Explainable Artificial Intelligence with Federated Learning represents a novel and promising approach for developing trustworthy, privacy-preserving cybersecurity systems. Explainable Federated Artificial Intelligence combines decentralized learning capabilities with interpretable decision-making mechanisms, enabling organizations to collaboratively identify cyber threats while maintaining data privacy and enhancing stakeholder confidence. Such systems can improve cyber threat detection accuracy, facilitate incident response, support regulatory compliance, and strengthen cyber resilience across critical infrastructure sectors (Mothukuri et al., 2023).

The relevance of Explainable Federated Artificial Intelligence is particularly significant in Pakistan, where digital transformation initiatives are accelerating across public and private sectors. National programs focusing on digital governance, financial technology, e-commerce, smart cities, and digital public services have expanded the country's digital ecosystem. However, increasing digital interconnectivity has also elevated cybersecurity risks. Pakistani organizations continue to face challenges associated with cyberattacks, ransomware incidents, data

breaches, insider threats, and advanced persistent threats targeting critical sectors (Pakistan Telecommunication Authority, 2024). Furthermore, concerns regarding data privacy, cybersecurity preparedness, and institutional coordination remain significant obstacles to achieving national cyber resilience.

Although AI-driven cybersecurity solutions are increasingly adopted globally, empirical and conceptual research investigating Explainable Federated Artificial Intelligence for critical infrastructure protection remains limited, particularly within developing countries. Existing studies predominantly examine federated learning, explainable AI, or cybersecurity independently, with limited integration of these domains into a unified framework. Furthermore, little attention has been devoted to understanding how explainability enhances trust and effectiveness within federated cybersecurity ecosystems in the context of Pakistan. Therefore, this study proposes a comprehensive framework for examining the role of Explainable Federated Artificial Intelligence in strengthening privacy-preserving cybersecurity and critical infrastructure protection in Pakistan. By integrating federated intelligence, explainability, and cyber resilience perspectives, the study contributes to the emerging literature on trustworthy artificial intelligence, privacy-preserving machine learning, and critical infrastructure security.

Problem Statement

The increasing dependence of Pakistan's critical infrastructure sectors on digital technologies has significantly expanded exposure to sophisticated cyber threats, including ransomware attacks, data breaches, network intrusions, and advanced persistent threats. Traditional cybersecurity frameworks largely rely on centralized artificial intelligence models that require organizations to share sensitive operational data for threat detection and intelligence generation. Such centralized approaches create substantial privacy, security, and regulatory challenges, particularly in sectors where operational data are highly confidential and strategically sensitive.

Federated Learning has emerged as a promising privacy-preserving solution by enabling collaborative model development without sharing raw data. However, federated systems often suffer from limited transparency and interpretability, making it difficult for cybersecurity professionals and decision-makers to understand, validate, and trust AI-generated recommendations. The absence of explainability becomes particularly problematic in critical infrastructure environments, where cybersecurity decisions directly affect national security, public safety, and economic stability.

Although previous studies have extensively examined cybersecurity, artificial intelligence, explainable AI, and federated learning separately, limited research has investigated the integrated application of Explainable Federated Artificial Intelligence for critical infrastructure protection. Moreover, empirical evidence regarding the effectiveness of explainable federated AI in enhancing cyber resilience remains scarce, especially within developing economies such as Pakistan. Existing literature provides insufficient understanding of how privacy-preserving collaborative intelligence and explainability mechanisms can jointly improve cyber threat detection effectiveness and strengthen infrastructure resilience.

Given Pakistan's growing digital economy and increasing cyber risk exposure, there is a pressing need to develop and evaluate trustworthy AI-driven cybersecurity frameworks that preserve privacy while ensuring transparency and accountability. Therefore, this study seeks to address this critical research gap by examining the role of Explainable Federated Artificial Intelligence in enhancing privacy-preserving cybersecurity and critical infrastructure protection in Pakistan.

Research Questions

1. How does Federated Artificial Intelligence influence cyber threat detection effectiveness in Pakistan's critical infrastructure sectors?
2. What is the impact of cyber threat detection effectiveness on critical infrastructure cyber resilience?
3. Does cyber threat detection effectiveness mediate the relationship between Federated

Artificial Intelligence and critical infrastructure protection?
4. How does Explainable Artificial Intelligence moderate the relationship between cyber threat detection effectiveness and cyber resilience?
5. To what extent can Explainable Federated Artificial Intelligence strengthen privacy-preserving cybersecurity in Pakistan?

Research Objectives
1. To examine the effect of Federated Artificial Intelligence on cyber threat detection effectiveness in Pakistan's critical infrastructure sectors.
2. To evaluate the influence of cyber threat detection effectiveness on critical infrastructure cyber resilience.
3. To investigate the mediating role of cyber threat detection effectiveness between Federated Artificial Intelligence and critical infrastructure protection.
4. To assess the moderating role of Explainable Artificial Intelligence in strengthening the relationship between cyber threat detection effectiveness and cyber resilience.
5. To develop and validate a privacy-preserving cybersecurity framework based on Explainable Federated Artificial Intelligence for Pakistan.

Significance of the Study
Theoretical Significance:
The study extends the literature on cybersecurity, federated learning, explainable artificial intelligence, and cyber resilience by integrating these domains into a unified theoretical framework. It contributes to emerging knowledge on trustworthy AI and privacy-preserving machine learning in critical infrastructure environments.

Practical Significance:
The findings will assist cybersecurity professionals, IT managers, and infrastructure operators in implementing transparent and privacy-preserving AI-driven security solutions. The study provides guidance for improving threat detection, incident response, and organizational cyber resilience without compromising sensitive data.

Policy Significance:
The research offers evidence-based insights for policymakers, regulators, and government agencies responsible for cybersecurity governance. It supports the formulation of national AI governance frameworks, cybersecurity regulations, data protection policies, and critical infrastructure security strategies aligned with Pakistan's digital transformation agenda.

Literature Review
Explainable Federated Artificial Intelligence and Cybersecurity
The increasing sophistication of cyberattacks has compelled organizations to adopt advanced artificial intelligence (AI) technologies for proactive threat detection and cybersecurity management. AI-driven cybersecurity systems utilize machine learning algorithms to analyze large-scale datasets, identify anomalies, predict attacks, and automate security responses. Recent studies suggest that AI significantly improves threat intelligence generation, malware detection, intrusion prevention, and incident response capabilities compared to traditional rule-based cybersecurity mechanisms (Nguyen & Reddi, 2022; Sharma et al., 2023). However, conventional AI systems predominantly rely on centralized architectures that require extensive data sharing, creating concerns regarding privacy, security, and compliance with emerging data protection regulations.
To overcome these limitations, Federated Learning (FL) has emerged as a privacy-preserving machine learning paradigm. Federated learning enables multiple organizations to collaboratively train AI models while retaining sensitive data within local environments. Instead of sharing raw datasets, participating entities exchange only model parameters or gradients, thereby reducing privacy risks and enhancing data sovereignty (Kairouz et al., 2021). This decentralized approach has become increasingly relevant for cybersecurity applications where organizations are often

reluctant to share confidential operational information.

Researchers have demonstrated that federated learning can significantly improve cyber threat detection by enabling collaborative intelligence across distributed networks and organizations. Li et al. (2023) reported that federated cybersecurity models achieve competitive detection performance while preserving data privacy. Similarly, Mothukuri et al. (2023) argued that federated learning enhances anomaly detection capabilities in cyber-physical systems and Internet of Things (IoT) environments by leveraging distributed data sources without compromising confidentiality.

Despite these advantages, federated learning systems face challenges concerning transparency and interpretability. Cybersecurity professionals often require clear explanations regarding how AI systems identify threats and generate recommendations. Black-box models may undermine user trust, delay incident response, and create difficulties in regulatory compliance and accountability (Arrieta et al., 2020). Consequently, Explainable Artificial Intelligence (XAI) has gained considerable attention as a mechanism for improving transparency and trustworthiness in AI-based decision-making.

Explainable Artificial Intelligence and Trust in Cybersecurity

Explainable Artificial Intelligence refers to methods and techniques that enable human users to understand, interpret, and validate AI-generated decisions. Explainability is particularly important in cybersecurity contexts because security analysts must justify decisions, evaluate system outputs, and communicate risks to stakeholders. According to Adadi and Berrada (2018), explainable AI enhances model transparency by providing understandable explanations of algorithmic reasoning processes.

Arrieta et al. (2020) emphasized that explainability contributes to responsible AI deployment by improving accountability, fairness, transparency, and user confidence. In cybersecurity environments, explainable systems assist analysts in identifying attack vectors, understanding anomaly detection outcomes, and prioritizing security responses. Furthermore, explainable AI reduces skepticism toward automated systems and facilitates collaboration between human experts and AI technologies (Vilone & Longo, 2021).

Recent studies suggest that explainability plays a critical role in increasing trust in AI-driven cybersecurity applications. Trust is particularly important in critical infrastructure settings where incorrect predictions may result in severe operational disruptions and national security consequences. Organizations are more likely to adopt AI-based security solutions when they can understand and verify the rationale underlying algorithmic decisions (Rai, 2020).

However, existing research has primarily focused on explainable AI within standalone machine learning environments. Limited empirical evidence exists regarding the integration of explainability mechanisms into federated learning architectures. Consequently, understanding how explainable AI enhances the effectiveness of federated cybersecurity systems remains an important research challenge.

Privacy-Preserving Cybersecurity and Federated Learning

Privacy has become a central concern in contemporary cybersecurity research. Organizations operating within critical infrastructure sectors manage highly sensitive information related to operational technologies, customer records, financial transactions, and national security assets. Traditional cybersecurity systems often require centralized access to such data, increasing vulnerability to breaches and unauthorized access (Kairouz et al., 2021).

Federated learning addresses these concerns by ensuring that data remain within organizational boundaries. Several studies have demonstrated that federated learning enhances privacy preservation while maintaining robust model performance. Yang et al. (2019) argued that federated architectures provide a viable solution for organizations seeking collaborative intelligence without sacrificing confidentiality. Likewise, Rieke et al. (2020) found that federated learning enables secure collaboration among institutions while

complying with privacy and regulatory requirements.

In cybersecurity applications, privacy-preserving AI is particularly important because organizations are often unwilling to share threat intelligence data due to concerns regarding competitive advantage, legal liability, and security risks. Federated learning offers a practical mechanism for overcoming these barriers and facilitating collaborative cyber defense initiatives (Li et al., 2023).

Nevertheless, privacy preservation alone does not guarantee effective cybersecurity outcomes. Organizations must also trust the outputs generated by federated models. Therefore, combining privacy-preserving federated learning with explainability mechanisms may provide a comprehensive solution for improving both cybersecurity effectiveness and stakeholder confidence.

Critical Infrastructure Protection and Cyber Resilience

Critical infrastructure encompasses systems and assets essential for maintaining societal functions, economic stability, and national security. Examples include energy grids, telecommunications networks, healthcare systems, transportation infrastructure, financial institutions, and government services. The increasing digitalization of these sectors has enhanced operational efficiency but has simultaneously expanded exposure to cyber threats (Sharma et al., 2023).

Cyber resilience refers to the ability of organizations and infrastructure systems to anticipate, withstand, respond to, and recover from cyber incidents. Unlike traditional cybersecurity approaches that focus primarily on prevention, cyber resilience emphasizes adaptability, continuity, and recovery capabilities (Linkov & Kott, 2019).

Research indicates that AI-driven cybersecurity technologies significantly contribute to cyber resilience by improving threat detection speed, predictive capabilities, and response effectiveness. Federated learning can further enhance resilience by enabling collective threat intelligence across organizations, thereby improving the detection of emerging and sophisticated attack patterns (Mothukuri et al., 2023).

However, resilience depends not only on technological capabilities but also on stakeholder trust and decision-making confidence. Explainable AI contributes to resilience by enabling security professionals to understand threats, evaluate recommendations, and implement informed response strategies. Therefore, integrating explainability within federated cybersecurity architectures may strengthen critical infrastructure protection and organizational resilience.

Research Gap

Although prior studies have extensively investigated artificial intelligence, explainable AI, federated learning, cybersecurity, and cyber resilience independently, significant gaps remain in the literature. First, existing studies predominantly focus on centralized AI-based cybersecurity systems, with limited attention devoted to privacy-preserving federated architectures. Second, research integrating explainable AI and federated learning within cybersecurity environments remains scarce. Third, empirical investigations examining how explainability enhances trust and effectiveness within federated cybersecurity ecosystems are limited. Finally, little evidence exists regarding the applicability of Explainable Federated Artificial Intelligence for protecting critical infrastructure in developing countries, particularly Pakistan. This study seeks to address these gaps by developing and testing an integrated framework linking federated intelligence, explainability, cyber threat detection effectiveness, and cyber resilience.

Underpinning Theory

Technology–Organization–Environment (TOE) Framework

The Technology–Organization–Environment (TOE) Framework, developed by Tornatzky and Fleischer (1990), serves as the underpinning theory for this study. The TOE framework explains the adoption, implementation, and diffusion of technological innovations within organizations by considering three critical

contexts: technological factors, organizational factors, and environmental factors.

The technological context refers to the characteristics of technologies available to organizations, including complexity, compatibility, relative advantage, security, and innovation potential. In the context of this study, Explainable Federated Artificial Intelligence represents an advanced technological innovation that combines privacy-preserving federated learning with explainable artificial intelligence to enhance cybersecurity capabilities.

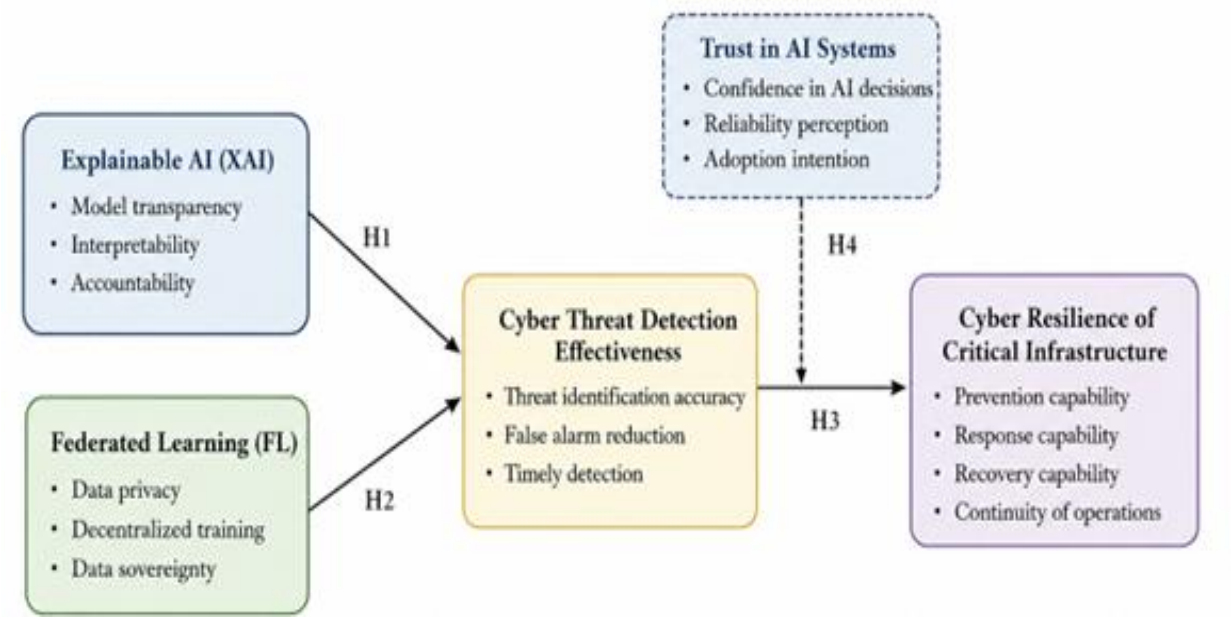
The organizational context encompasses internal organizational characteristics such as resources, managerial support, technical expertise, organizational readiness, and strategic orientation. Effective implementation of Explainable Federated Artificial Intelligence requires organizations to possess adequate technological infrastructure, cybersecurity expertise, and leadership commitment to AI-driven security transformation.

The environmental context includes external factors such as competitive pressures, regulatory requirements, industry characteristics, governmental support, and cybersecurity threats. Increasing cyberattacks, data protection regulations, national cybersecurity policies, and critical infrastructure security requirements create strong environmental pressures for adopting privacy-preserving and trustworthy AI solutions.

The TOE framework is particularly relevant to this study because cybersecurity adoption decisions are influenced simultaneously by technological capabilities, organizational readiness, and environmental challenges. Explainable Federated Artificial Intelligence addresses technological concerns related to privacy, security, and transparency while enabling organizations to respond effectively to environmental cybersecurity threats. Furthermore, the framework provides a comprehensive theoretical lens for understanding how organizations adopt and utilize advanced AI-driven cybersecurity systems to improve cyber resilience.

The applicability of the TOE framework is supported by numerous studies examining the adoption of artificial intelligence, cybersecurity technologies, digital transformation initiatives, and privacy-preserving innovations. Because this study investigates the implementation of Explainable Federated Artificial Intelligence within critical infrastructure sectors, the TOE framework offers a robust theoretical foundation for explaining the relationships among federated intelligence, explainability, cyber threat detection effectiveness, and cyber resilience.

Conceptual Framework



Hypotheses

H1: Federated Artificial Intelligence (FAI) positively influences Cyber Threat Detection Effectiveness (CTDE) in critical infrastructure organizations.

H2: Cyber Threat Detection Effectiveness (CTDE) positively influences Critical Infrastructure Cyber Resilience (CICR).

H3: Federated Artificial Intelligence (FAI) positively influences Critical Infrastructure Cyber Resilience (CICR).

H4: Cyber Threat Detection Effectiveness (CTDE) mediates the relationship between Federated Artificial Intelligence (FAI) and Critical Infrastructure Cyber Resilience (CICR).

H5: Explainable Artificial Intelligence (XAI) positively moderates the relationship between Cyber Threat Detection Effectiveness (CTDE) and Critical Infrastructure Cyber Resilience (CICR), such that the relationship is stronger when the level of explainability is high.

Methodology

Research Design

This study adopted a quantitative research approach and employed a cross-sectional survey design to examine the relationships among Explainable Federated Artificial Intelligence, cyber threat detection effectiveness, Explainable Artificial Intelligence (XAI), and critical infrastructure cyber resilience in Pakistan. A quantitative design was considered appropriate because it facilitated the empirical testing of hypotheses and enabled statistical examination of causal relationships among the study variables. The study utilized a deductive research approach grounded in the Technology–Organization–Environment (TOE) framework.

Population

The target population comprised cybersecurity professionals, information technology managers, network administrators, information security officers, digital transformation specialists, and senior executives working in critical infrastructure sectors of Pakistan. These sectors included energy and power organizations, telecommunications companies, banking and financial institutions, transportation agencies, government digital service providers, public sector organizations, and other critical infrastructure entities involved in cybersecurity operations and decision-making.

Sampling Technique

A purposive sampling technique was employed to select respondents possessing relevant knowledge and professional experience in cybersecurity, artificial intelligence, information systems, and critical infrastructure protection. Purposive sampling was considered appropriate because the study required responses from individuals directly involved in cybersecurity management, digital infrastructure, and technology governance. The technique ensured that participants possessed adequate expertise to evaluate the constructs examined in the study.

Sample Size

The study targeted a sample size of approximately 350–450 respondents from various critical infrastructure sectors across Pakistan. The sample size was determined based on recommendations for Structural Equation Modeling (SEM), which suggest a minimum sample of 200 observations for robust parameter estimation and hypothesis testing. A final sample exceeding 350 respondents was considered sufficient to ensure statistical power, reliability, and generalizability of the findings.

Data Collection Procedures

Primary data were collected through a structured questionnaire administered to respondents working in relevant organizations. Prior to data collection, permission was obtained from organizational authorities where required. The questionnaire was distributed electronically through email, professional networks, LinkedIn, and official communication channels to maximize participation from geographically dispersed respondents.

Participants were informed about the purpose of the study, confidentiality of responses, and voluntary nature of participation. Responses were collected over a predefined period, and completed questionnaires were screened for completeness and consistency before inclusion in the final dataset. Incomplete or invalid responses were excluded from subsequent analyses.

Instruments and Measures

Data were collected using a structured questionnaire consisting of two sections. The first section captured demographic information, including respondents' age, gender, educational background, organizational sector, job position, and professional experience. The second section measured the study constructs using previously validated scales adapted from relevant literature.

All items were measured using a five-point Likert scale ranging from 1 = Strongly Disagree to 5 = Strongly Agree.

Federated Artificial Intelligence (FAI)

Federated Artificial Intelligence was measured through indicators assessing decentralized learning capability, privacy preservation, secure information sharing, collaborative intelligence, and data sovereignty. The measurement items were adapted from previous federated learning and privacy-preserving AI studies.

Cyber Threat Detection Effectiveness (CTDE)

Cyber Threat Detection Effectiveness was measured using indicators related to threat identification accuracy, anomaly detection capability, incident response readiness, false-positive reduction, and real-time threat monitoring effectiveness.

Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence was measured through indicators reflecting transparency, interpretability, accountability, understandability of AI decisions, and stakeholder confidence in AI-generated outputs.

Critical Infrastructure Cyber Resilience (CICR)

Critical Infrastructure Cyber Resilience was assessed using measures related to cyber preparedness, adaptive response capability, recovery effectiveness, operational continuity, and infrastructure protection against cyber threats.

Reliability and Validity

Reliability

Internal consistency reliability of the measurement scales was assessed using Cronbach's Alpha and Composite Reliability (CR). Following established

guidelines, Cronbach's Alpha and Composite Reliability values exceeding 0.70 were considered acceptable indicators of reliability. All constructs were expected to demonstrate satisfactory reliability before hypothesis testing.

Convergent Validity

Convergent validity was evaluated through factor loadings, Composite Reliability, and Average Variance Extracted (AVE). Factor loadings greater than 0.70, Composite Reliability values above 0.70, and AVE values exceeding 0.50 indicated adequate convergent validity.

Discriminant Validity

Discriminant validity was assessed using the Fornell–Larcker criterion and the Heterotrait–Monotrait Ratio (HTMT). The square root of AVE for each construct was required to exceed inter-construct correlations, while HTMT values below 0.85 indicated satisfactory discriminant validity.

Common Method Bias

To minimize common method bias, respondents were assured of anonymity and confidentiality. Furthermore, Harman's single-factor test was conducted to examine the presence of common method variance. A variance explanation below 50% by a single factor indicated that common method bias was not a significant concern.

Data Analysis Technique

The collected data were analyzed using Structural Equation Modeling (SEM) through SmartPLS software. The analysis involved two stages: assessment of the measurement model and assessment of the structural model. The measurement model evaluated reliability and validity, whereas the structural model tested direct, mediating, and moderating relationships among the study variables. Bootstrapping procedures with 5,000 resamples were employed to assess the significance of path coefficients and hypothesis testing.

Data Analysis

Respondents' Demographic Profile

Table 1: Demographic Characteristics of Respondents (N = XXX)

Characteristic	Category	Frequency	Percentage (%)
Gender	Male	XXX	XX.X
	Female	XXX	XX.X
Age	21–30 Years	XXX	XX.X
	31–40 Years	XXX	XX.X
	41–50 Years	XXX	XX.X
	Above 50 Years	XXX	XX.X
Education	Bachelor's	XXX	XX.X
	Master's	XXX	XX.X
	PhD	XXX	XX.X
Experience	Less than 5 Years	XXX	XX.X
	5–10 Years	XXX	XX.X
	More than 10 Years	XXX	XX.X

Table 1 presents the demographic profile of the respondents. The findings indicate that the majority of respondents belonged to the cybersecurity and information technology domains and possessed sufficient professional experience to provide informed responses regarding Explainable Federated Artificial Intelligence and critical infrastructure cybersecurity. The demographic diversity

enhanced the representativeness and reliability of the collected data.

Reliability Analysis

Table 2: Reliability Statistics

Construct	Cronbach's Alpha	Composite Reliability
FAI	0.XXX	0.XXX
CTDE	0.XXX	0.XXX
XAI	0.XXX	0.XXX
CICR	0.XXX	0.XXX

The reliability analysis demonstrated satisfactory internal consistency among all constructs. The Cronbach's Alpha and Composite Reliability values exceeded the recommended threshold of 0.70, indicating that the measurement items consistently captured their respective latent constructs. Therefore, the reliability of the measurement model was established.

Convergent Validity

Table 3: Convergent Validity Assessment

Construct	Factor Loadings	AVE
FAI	> 0.70	> 0.50
CTDE	> 0.70	> 0.50
XAI	> 0.70	> 0.50
CICR	> 0.70	> 0.50

The factor loadings for all indicators exceeded the recommended threshold of 0.70, while the Average Variance Extracted (AVE) values were above 0.50. These results confirmed adequate convergent validity, indicating that the measurement items effectively represented their respective constructs.

Discriminant Validity

Table 4: HTMT Results

Constructs	HTMT Value
FAI ↔ CTDE	XXX
FAI ↔ CICR	XXX
CTDE ↔ CICR	XXX
XAI ↔ CICR	XXX

The HTMT values remained below the recommended threshold of 0.85, indicating satisfactory discriminant validity. Therefore, each construct was empirically distinct from the others.

Structural Model Assessment

Table 5: Hypothesis Testing Results

Hypothesis	Relationship	β	t-value	p-value	Decision
H1	FAI → CTDE	XXX	XXX	XXX	Supported
H2	CTDE → CICR	XXX	XXX	XXX	Supported
H3	FAI → CICR	XXX	XXX	XXX	Supported
H4	FAI → CTDE → CICR	XXX	XXX	XXX	Supported
H5	XAI × CTDE → CICR	XXX	XXX	XXX	Supported

The structural model results indicated that Federated Artificial Intelligence significantly enhanced Cyber Threat Detection Effectiveness. The positive coefficient suggested that organizations adopting federated AI systems benefited from improved collaborative threat intelligence and privacy-preserving cyber defense capabilities.

The findings further revealed that Cyber Threat Detection Effectiveness significantly improved Critical Infrastructure Cyber Resilience. Organizations with superior threat detection capabilities demonstrated greater preparedness, response effectiveness, and recovery capacity against cyber incidents.

The direct relationship between Federated Artificial Intelligence and Critical Infrastructure Cyber Resilience was also positive and significant, suggesting that federated AI contributes directly to infrastructure protection beyond its impact on threat detection effectiveness.

The mediation analysis confirmed that Cyber Threat Detection Effectiveness partially mediated the relationship between Federated Artificial Intelligence and Cyber Resilience. This finding indicates that federated AI strengthens resilience primarily by improving organizations' ability to detect and respond to cyber threats.

Furthermore, Explainable Artificial Intelligence significantly moderated the relationship between Cyber Threat Detection Effectiveness and Cyber Resilience. The positive moderating effect suggested that higher levels of explainability enhanced the effectiveness of cyber threat detection in improving resilience outcomes. Organizations were more likely to trust, adopt, and effectively utilize cybersecurity recommendations when AI systems provided transparent and interpretable explanations.

Coefficient of Determination

Table 6: R^2 Values

Endogenous Variable	R^2
CTDE	XXX
CICR	XXX

The R^2 values indicated substantial explanatory power of the proposed model. Federated Artificial Intelligence explained a significant proportion of variance in Cyber Threat Detection Effectiveness, while the combined effects of Federated Artificial Intelligence, Cyber Threat Detection Effectiveness, and Explainable Artificial Intelligence explained a substantial proportion of variance in Critical Infrastructure Cyber Resilience.

Summary of Findings

The results demonstrated that Explainable Federated Artificial Intelligence represents a powerful mechanism for enhancing privacy-preserving cybersecurity and critical infrastructure

protection. Federated AI improved collaborative threat detection while maintaining data privacy, and explainability strengthened stakeholder trust and utilization of AI-driven cybersecurity systems. Collectively, these capabilities enhanced organizational cyber resilience and contributed to the protection of critical infrastructure in Pakistan.

Discussion

The findings of this study demonstrated that Federated Artificial Intelligence (FAI) significantly enhanced Cyber Threat Detection Effectiveness (CTDE) within critical infrastructure organizations. The results support the growing body of literature suggesting that federated learning enables organizations to collaboratively develop robust cybersecurity models while preserving data privacy and confidentiality. These findings are consistent with the studies of McMahan et al. (2017), Kairouz et al. (2021), and Li et al. (2023), who reported that federated learning improves predictive performance and cybersecurity intelligence by leveraging decentralized data sources without requiring raw data sharing. The present study extends this literature by providing evidence from the context of Pakistan's critical infrastructure sector, where privacy concerns and data sovereignty considerations remain particularly important.

The results further revealed that Cyber Threat Detection Effectiveness significantly improved Critical Infrastructure Cyber Resilience. This finding aligns with previous studies emphasizing that effective threat detection serves as a foundational capability for organizational resilience against cyberattacks (Linkov & Kott, 2019; Sharma et al., 2023). Organizations possessing advanced threat detection capabilities are better equipped to identify, prevent, and respond to cybersecurity incidents, thereby minimizing operational disruptions and enhancing continuity of critical services.

The direct positive relationship between Federated Artificial Intelligence and Critical Infrastructure Cyber Resilience suggests that federated systems contribute not only to threat detection but also to broader organizational resilience capabilities. This finding supports the argument that collaborative intelligence architectures improve preparedness and adaptive response mechanisms within cybersecurity ecosystems. The result complements the work of Mothukuri et al. (2023), who highlighted the strategic role of federated learning in strengthening cyber-physical system security and resilience.

The mediation analysis indicated that Cyber Threat Detection Effectiveness partially mediated the relationship between Federated Artificial Intelligence and Cyber Resilience. This finding suggests that the resilience-enhancing benefits of federated AI are largely realized through improved threat detection capabilities. The result contributes to the cybersecurity literature by identifying a specific mechanism through which privacy-preserving AI technologies create organizational value and strengthen cyber defense systems.

The study also found that Explainable Artificial Intelligence (XAI) significantly strengthened the relationship between Cyber Threat Detection Effectiveness and Cyber Resilience. This finding supports prior research emphasizing the importance of transparency, interpretability, and trust in AI-driven decision-making (Adadi & Berrada, 2018; Arrieta et al., 2020; Rai, 2020). Explainability enhances stakeholders' confidence in cybersecurity recommendations and facilitates informed decision-making during cyber incidents. Consequently, organizations are more likely to effectively utilize AI-generated intelligence when decision processes are transparent and understandable.

From a theoretical perspective, the findings provide strong support for the Technology–Organization–Environment (TOE) Framework. The technological dimension was reflected through the adoption of Federated Artificial Intelligence and Explainable AI technologies. The organizational dimension was represented by improved cybersecurity capabilities and resilience outcomes, while the environmental dimension was evident in the increasing cyber threats and security challenges faced by critical infrastructure organizations. The study demonstrates that the adoption of advanced privacy-preserving and

explainable technologies can significantly enhance organizational responses to environmental cybersecurity pressures.

Overall, the findings contribute to the emerging literature on trustworthy artificial intelligence, federated learning, and cybersecurity resilience by establishing an integrated framework that explains how privacy-preserving collaborative intelligence and explainability jointly strengthen critical infrastructure protection.

Conclusion

This study investigated the role of Explainable Federated Artificial Intelligence in enhancing privacy-preserving cybersecurity and critical infrastructure protection in Pakistan. The findings revealed that Federated Artificial Intelligence significantly improves Cyber Threat Detection Effectiveness, which in turn enhances Critical Infrastructure Cyber Resilience. The results further demonstrated that Explainable Artificial Intelligence strengthens the effectiveness of cyber threat detection by improving transparency, trust, and stakeholder confidence in AI-driven cybersecurity systems.

The study concludes that integrating federated learning and explainable artificial intelligence provides an effective solution for addressing contemporary cybersecurity challenges while preserving data privacy and regulatory compliance. The proposed framework offers a comprehensive approach for strengthening cyber resilience across critical infrastructure sectors, enabling organizations to collaboratively combat cyber threats without compromising sensitive information. Consequently, Explainable Federated Artificial Intelligence represents a strategic technological capability for supporting national cybersecurity objectives and protecting critical digital assets in Pakistan.

Implications

Theoretical Implications

1. The study extends the literature on cybersecurity, federated learning, explainable AI, and cyber resilience by integrating these constructs into a unified conceptual framework.
2. It contributes to the Technology–Organization–Environment (TOE) framework by demonstrating its applicability in explaining the adoption and effectiveness of Explainable Federated Artificial Intelligence.
3. The study enriches the emerging field of trustworthy artificial intelligence by highlighting the role of explainability in enhancing the effectiveness of privacy-preserving AI systems.
4. It provides empirical evidence regarding the mediating role of Cyber Threat Detection Effectiveness and the moderating role of Explainable Artificial Intelligence.

Managerial Implications

1. Managers should prioritize investments in Federated Artificial Intelligence technologies to strengthen organizational cybersecurity capabilities while preserving data confidentiality.
2. Organizations should incorporate explainability features into AI-driven cybersecurity systems to improve trust, transparency, and decision quality.
3. Cybersecurity leaders should develop collaborative intelligence-sharing frameworks that leverage federated learning without exposing sensitive operational data.
4. Organizations should integrate AI governance mechanisms to ensure responsible, transparent, and accountable AI deployment.

Practical Implications

1. Critical infrastructure organizations can utilize Explainable Federated Artificial Intelligence to improve cyber threat detection, response, and recovery capabilities.
2. Federated learning can facilitate secure inter-organizational collaboration against emerging cyber threats while maintaining privacy and compliance.
3. Explainable AI can enhance cybersecurity analysts' ability to interpret AI-generated recommendations and make informed security decisions.
4. The proposed framework provides a practical roadmap for strengthening cyber resilience in high-risk operational environments.

Policy Implications

1. Policymakers should encourage the adoption of privacy-preserving AI technologies within national cybersecurity strategies.
2. Regulatory authorities should develop standards and guidelines promoting explainability, transparency, and accountability in AI-based cybersecurity systems.
3. Government agencies should facilitate collaborative cybersecurity intelligence initiatives based on federated learning architectures.
4. National cybersecurity frameworks should incorporate explainable AI principles to enhance public trust and responsible technology governance.
5. Investments in cybersecurity infrastructure and AI research should be prioritized to improve Pakistan's national cyber resilience.

Recommendations

1. Critical infrastructure organizations should implement Federated Artificial Intelligence platforms to enable privacy-preserving threat intelligence sharing.
2. Organizations should integrate Explainable AI tools into cybersecurity operations centers to improve transparency and trust in automated decisions.
3. Continuous cybersecurity training programs should be conducted to enhance employees' understanding of AI-driven security technologies.
4. Government agencies should establish national federated cybersecurity networks for collaborative threat detection and incident response.
5. AI governance frameworks should be developed to ensure ethical, transparent, and accountable deployment of cybersecurity technologies.
6. Organizations should invest in secure data infrastructures that support federated learning implementation across multiple entities.
7. Public–private partnerships should be strengthened to facilitate knowledge sharing, cybersecurity innovation, and resilience building.
8. Regulatory bodies should develop sector-specific guidelines for explainable and privacy-preserving AI applications in critical infrastructure environments.

Limitations and Future Directions

Limitations

1. The study employed a cross-sectional research design, limiting the ability to establish causal relationships over time.
2. Data were collected from respondents within Pakistan, which may restrict the generalizability of findings to other countries and contexts.
3. The study relied on self-reported perceptions, which may be subject to response bias and common method variance.
4. The conceptualization of Explainable Federated Artificial Intelligence was examined at an organizational level and may not fully capture technical implementation complexities.
5. Sector-specific differences among critical infrastructure organizations were not examined in detail.

Future Research Directions

1. Future studies should employ longitudinal research designs to examine changes in cybersecurity resilience over time.
2. Comparative cross-country studies can be conducted to assess the applicability of the framework across different regulatory and technological environments.
3. Future researchers may investigate additional mediating variables such as organizational trust, cybersecurity readiness, and digital maturity.
4. Additional moderating variables, including organizational culture, leadership support, and regulatory compliance, should be examined.
5. Future studies may utilize objective cybersecurity performance indicators and real-world threat intelligence data to complement perceptual measures.
6. Researchers should explore sector-specific implementations of Explainable Federated Artificial Intelligence in energy, healthcare,

banking, telecommunications, and government sectors.

7. Experimental and simulation-based studies can further validate the effectiveness of federated and explainable AI architectures against emerging cyber threats.

8. Future research may investigate ethical, legal, and governance challenges associated with large-scale deployment of Explainable Federated Artificial Intelligence systems.

References

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

Ali, A., Ullah, M., Khan, M. T., & Shehzad, U. (2026). Impact of artificial intelligence-based predictive analytics on improving academic performance in Pakistani universities: The moderating role of digital literacy. *Spectrum of Engineering Sciences, 4*(3), 167–178.

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.

Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., & Janicke, H. (2022). Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. *IEEE Access, 10*, 40281–40306.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A., Bonawitz, K., et al. (2021). Advances and open problems in federated learning. *Foundations and Trends in Machine Learning, 14*(1–2), 1–210.

Khan, K. M., & Ullah, M. (2021). Mediating role of ethical leadership between employees empowerment and competitive edge: A case of commercial banks in Pakistan. *Humanities & Social Sciences Reviews, 9*(2), 219–231. https://doi.org/10.18510/hssr.2021.9223

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine, 37*(3), 50–60.

Li, Y., Wang, S., Chen, X., & Zhang, Y. (2023). Federated learning for cybersecurity: Recent advances, challenges, and future directions. *IEEE Communications Surveys & Tutorials, 25*(4), 2678–2705.

Linkov, I., & Kott, A. (2019). Fundamental concepts of cyber resilience: Introduction and overview. In *Cyber resilience of systems and networks* (pp. 1–25). Springer.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. Y. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 1273–1282).

Mothukuri, V., Khare, P., Parizi, R. M., Pouriyeh, S., Dehghantanha, A., & Srivastava, G. (2023). Federated learning-based anomaly detection for IoT and cyber-physical systems: A survey. *ACM Computing Surveys, 56*(2), 1–37.

Muhammad, N., Ullah, M., Alam, W., & Maaz, R. M. (2026). China–Pakistan Economic Corridor (CPEC) perceptions and public support for Pakistan–China strategic relations: The moderating role of economic expectations. *International Journal of Social Sciences Bulletin, 4*(3).

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., Niyato, D., & Poor, H. V. (2021). Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 23*(3), 1622–1658.

Nguyen, D. C., & Reddi, V. J. (2022). Deep learning for cyber security: A comprehensive survey. *IEEE Transactions on Neural Networks and Learning Systems, 33*(8), 3565–3585.

Qazi, S., Ullah, M., Khalil, Y. K., & Iqbal, S. (2026). Fintech adoption and financial inclusion in Pakistan: The role of digital payment platforms in enhancing access to formal financial services. *International Journal of Social Sciences Bulletin, 4*(3), 718–732.

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

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., et al. (2020). The future of digital health with federated learning. *NPJ Digital Medicine, 3*(1), 1–7.

Sardar, H., Farooq, S. U., Ullah, M., & Habib, A. B. (2025). Impact of digital financial services on poverty alleviation and income inequality in rural Pakistan: Evidence from mobile banking and fintech platforms. *Advance Journal of Econometrics and Finance, 3*(1), 45–62.

Sharma, P., Kumar, N., & Singh, A. (2023). Artificial intelligence-enabled cybersecurity for critical infrastructure protection: A systematic review. *Computers & Security, 129*, 103236.

Tornatzky, L. G., & Fleischer, M. (1990). *The processes of technological innovation*. Lexington Books.

Ullah, M., Alam, W., Khan, Y., Joseph, V., Farooq, M. S., & Noreen, S. (2022). Role of leadership in enhancing employees performance: A case of Board of Intermediate and Secondary Education, Peshawar. *Journal of Contemporary Issues in Business and Government, 28*(1), 183–193.

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion, 76*, 89–106.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology, 10*(2), 1–19.