Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 3, 2026

AI-DRIVEN CYBER SOVEREIGNTY AND NATIONAL SECURITY IN PAKISTAN: EMERGING LEGAL AND INTERNATIONAL GOVERNANCE CHALLENGES

Qazi Hussain Asghar[1], Dr. Mahboob Ullah[*2], Syeda Duaa Zehra Naqvi[3]

[1]Ex Civil Servant/ PhD Scholar, Quaid-i-Azam University Islamabad
*[*2]Associate Professor, Department of Management Sciences, Abasyn University*
[3]PhD scholar/ Research Analyst, Department of Defence & Strategic Studies, Quaid-i-Azam University Islamabad

[1]qazihussainasgharkhan@gmail.com, [*2]mahboob.ullah@abasyn.edu.pk, [3]duaazehra89@gmail.com

DOI: http://doi.org/10.5281/zenodo.22842546

Keywords
Artificial Intelligence; AI-Driven Cyber Threats; Cyber Sovereignty; Cyber Governance; National Security Resilience; AI Regulation

Article History
Received: 06 January 2026
Accepted: 19 February 2026
Published: 05 March 2026

Copyright @Author
Corresponding Author: *
Dr. Mahboob Ullah

Abstract

The rapid integration of artificial intelligence (AI) into cyberspace has transformed both defensive cybersecurity capabilities and the nature of emerging threats to national security. This study examined the relationship between AI-driven cyber threats, legal and regulatory gaps, cyber governance capacity, and national security resilience in Pakistan. Guided by Socio-Technical Systems Theory, the study proposed a quantitative framework linking technological threats with institutional and regulatory capacity. An illustrative cross-sectional design was employed using a simulated dataset of 300 cybersecurity professionals, IT specialists, legal practitioners, academics, and public-sector officials. Descriptive statistics, reliability analysis, Pearson correlation, multiple regression, and mediation analysis were applied. The illustrative findings indicated that AI-driven cyber threats were negatively associated with national security resilience, while cyber governance capacity demonstrated a significant positive relationship with resilience. Legal and regulatory gaps were negatively associated with cyber governance capacity, and governance capacity partially mediated the relationship between AI-driven cyber threats and national security resilience. The findings underscore the importance of integrating AI-specific cybersecurity capabilities with legal accountability, institutional coordination, regulatory capacity, and international cooperation. The study provides a conceptual and methodological foundation for future empirical research on AI-driven cyber sovereignty and national security in Pakistan.

1. Introduction

Artificial intelligence (AI) is fundamentally transforming the cybersecurity environment by increasing the speed, scale, adaptability, and sophistication of both defensive and offensive cyber operations. Generative AI and large language models can support vulnerability identification, threat detection, security testing, and automated incident analysis; however, the same technologies can facilitate sophisticated phishing, social engineering, malware development, data poisoning, identity manipulation, and other malicious activities (Hilario et al., 2024; Malatji & Tolah, 2025; Paracha et al., 2024). Recent cybersecurity assessments indicate that threat actors are increasingly experimenting with AI-enabled capabilities, including the use of generative AI for reconnaissance, social engineering, malware-related activities, and evasion techniques (European Union Agency for Cybersecurity [ENISA], 2025).

The growing convergence between AI and cyberspace has consequently expanded the

meaning of national security. Contemporary national security increasingly depends not only on territorial and military capabilities but also on the protection of digital infrastructure, government information systems, financial networks, telecommunications, health systems, energy infrastructure, and strategic data. AI-enabled attacks against such systems may produce consequences extending beyond conventional cybercrime because they can affect institutional continuity, economic stability, public trust, and the functioning of critical state services (OECD, 2024).

This transformation also raises questions concerning cyber sovereignty, broadly understood in this study as the state's capacity to protect, regulate, govern, and exercise legitimate authority over its digital infrastructure, information assets, and cyberspace while participating in international digital cooperation. AI complicates cyber sovereignty because digital attacks frequently cross national borders, AI systems may be developed and hosted outside the affected jurisdiction, data may be processed through transnational infrastructure, and responsibility for AI-generated or AI-assisted actions can be difficult to establish. Consequently, national regulatory authority increasingly intersects with international standards, cross-border investigations, private technology providers, and global AI governance mechanisms.

The governance problem is intensified by the rapid development of AI technologies relative to institutional and legal adaptation. The United Nations High-Level Advisory Body on AI identified a significant global governance deficit and emphasized the need for international cooperation and mechanisms capable of adapting to rapidly evolving AI capabilities (United Nations High-Level Advisory Body on Artificial Intelligence, 2024). Similarly, the OECD has identified increasingly sophisticated cyberattacks, manipulation, fraud, incidents affecting critical systems, and unclear liability as important AI-related policy risks (OECD, 2024). Recent scholarship likewise emphasizes the need to align AI control, accountability, transparency, and risk management across developers, organizations, and users (Grote et al., 2024; Novelli et al., 2024).

The Pakistani context is particularly important because the country has developed formal policy instruments addressing both cybersecurity and AI. Pakistan's National Cyber Security Policy 2021 identifies critical infrastructure protection, cybersecurity assurance, vulnerability management, incident response, capacity building, and international cooperation as important national priorities. It also explicitly recognizes the need for next-generation cybersecurity forensic and screening capabilities in an AI-driven environment (Ministry of Information Technology and Telecommunication [MoITT], 2021). More recently, Pakistan's National Artificial Intelligence Policy, approved in 2025, established a national framework for responsible AI adoption and includes attention to secure AI, cybersecurity, data protection, risk management, human oversight, and responsible deployment (MoITT, 2025).

Nevertheless, the emergence of AI-specific cyber risks creates challenges that cannot be addressed exclusively through conventional cybercrime legislation or general cybersecurity policies. Recent research on Pakistan identifies continuing legal and regulatory questions concerning privacy, accountability, ethical AI, data protection, and the adaptation of existing digital laws to emerging AI applications (Ahmad et al., 2024). The evolving threat environment also raises questions about attribution, liability, AI-assisted cybercrime, deepfakes, autonomous systems, cross-border attacks, and the responsibilities of private technology providers.

Accordingly, this study develops an integrated framework linking AI-driven cyber threats, legal and regulatory gaps, cyber governance capacity, and national security resilience in Pakistan. The study uses an illustrative quantitative model to examine whether AI-driven cyber threats and legal-regulatory weaknesses are associated with reduced national security resilience and whether stronger cyber governance capacity can provide an institutional mechanism for strengthening resilience. The study thereby connects technological risk with legal, institutional, and international governance dimensions of cyber sovereignty.

2. Problem Statement

The increasing integration of AI into cyberspace has created a complex national security environment in which the same technologies can strengthen cyber defence while simultaneously increasing offensive cyber capabilities. AI can automate reconnaissance, improve social engineering, generate convincing synthetic content, identify vulnerabilities, and accelerate malicious operations. Such capabilities can reduce the technical and financial barriers to cyberattacks while increasing their speed and scalability (Hilario et al., 2024; Malatji & Tolah, 2025). The resulting environment creates challenges for conventional approaches to cybersecurity, particularly where attacks are automated, adaptive, distributed, and difficult to attribute.

For Pakistan, the challenge is amplified by the growing dependence of public administration, financial services, telecommunications, digital commerce, and critical infrastructure on interconnected information systems. A successful AI-enabled attack against strategically important digital infrastructure could potentially generate consequences extending beyond financial or technical losses to institutional disruption and national security concerns. Pakistan's National Cyber Security Policy 2021 recognizes the importance of protecting critical infrastructure and developing advanced cybersecurity capabilities, including capabilities designed for an AI-driven environment (MoITT, 2021). However, the rapid evolution of generative AI and autonomous or semi-autonomous cyber capabilities creates a continuing need to evaluate whether existing institutional and legal arrangements remain adequate.

A major problem concerns the legal and regulatory gap between technological development and regulatory adaptation. Existing cybercrime and information-security laws were largely developed before the widespread availability of generative AI. Consequently, questions may arise concerning responsibility when an AI system contributes to a cyberattack, when malicious content is generated automatically, when deepfakes are used for strategic manipulation, or when an AI-enabled attack originates from infrastructure located across several jurisdictions. Research specifically examining Pakistan's AI regulatory environment has identified challenges related to privacy, data protection, ethical governance, accountability, and the development of an appropriate legal framework (Ahmad et al., 2024).

A second problem concerns institutional cyber governance capacity. Effective cyber sovereignty requires more than legislation. It depends on technical expertise, institutional coordination, threat intelligence, incident-response capability, critical-infrastructure protection, digital forensics, regulatory enforcement, and cooperation between government and private-sector organizations. The existence of a policy framework does not automatically ensure implementation capacity. The increasing complexity of AI-enabled threats therefore requires empirical examination of the relationship between cyber governance capacity and national security resilience.

A third problem concerns the international nature of AI-driven cyber threats. Cyber operations can cross territorial boundaries almost instantaneously, while AI systems, cloud infrastructures, data repositories, software platforms, and threat actors may be distributed across multiple jurisdictions. This creates difficulties involving jurisdiction, evidence collection, attribution, extradition, liability, data access, and international cooperation. The United Nations has consequently emphasized the need for more inclusive and internationally coordinated AI governance arrangements (United Nations High-Level Advisory Body on Artificial Intelligence, 2024).

Although the international literature increasingly examines AI governance, AI cybersecurity, and responsible AI, there remains a contextual gap concerning the interaction among AI-driven cyber threats, cyber sovereignty, legal-regulatory capacity, and national security resilience in Pakistan. Existing discussions often examine AI regulation, cybersecurity, or national security separately rather than integrating these dimensions into a single analytical framework. This study addresses that gap by conceptualizing cyber governance capacity as an institutional mechanism through which AI-related threats and regulatory

weaknesses may influence national security resilience.

The study is therefore designed to provide a structured framework for understanding Pakistan's emerging AI-related cybersecurity challenges and to identify the legal, institutional, and international governance dimensions that require further empirical investigation.

3. Research Questions

RQ1: What is the relationship between AI-driven cyber threats and national security resilience in Pakistan?

RQ2: How do legal and regulatory gaps associated with AI affect cyber governance capacity in Pakistan?

RQ3: What is the relationship between cyber governance capacity and national security resilience?

RQ4: Does cyber governance capacity mediate the relationship between AI-driven cyber threats and national security resilience?

RQ5: What major legal and international governance challenges arise from AI-driven cyber threats to Pakistan's cyber sovereignty?

4. Research Objectives

The study aims to:

1. **Examine** the relationship between AI-driven cyber threats and national security resilience in Pakistan.
2. **Assess** the effect of legal and regulatory gaps on Pakistan's cyber governance capacity.
3. **Determine** the relationship between cyber governance capacity and national security resilience.
4. **Evaluate** the mediating role of cyber governance capacity in the relationship between AI-driven cyber threats and national security resilience.
5. **Identify** major legal, institutional, and international governance challenges associated with AI-driven cyber sovereignty in Pakistan.
6. **Develop** evidence-informed recommendations for strengthening Pakistan's legal and institutional capacity to manage AI-related cybersecurity risks.

5. Significance of the Study

5.1 Theoretical Significance

The study contributes to the emerging literature on AI governance and cybersecurity by integrating technological threats, institutional capacity, legal regulation, and national security into a single analytical framework. Rather than conceptualizing AI as exclusively a technological phenomenon, the study treats AI-driven cybersecurity as a multidimensional governance issue involving law, institutions, accountability, sovereignty, and international cooperation.

The framework also contributes to the literature by positioning cyber governance capacity as a potential mediating mechanism between AI-driven cyber threats and national security resilience. This provides a basis for examining how institutional capabilities may translate technological risks into different national security outcomes.

5.2 Practical Significance

The study can assist cybersecurity professionals, government institutions, technology organizations, financial institutions, and critical-infrastructure operators in identifying areas where conventional cybersecurity mechanisms may require adaptation to AI-enabled threats. Particular attention can be directed toward AI-assisted social engineering, automated cyberattacks, deepfakes, data poisoning, AI supply-chain risks, and AI-enabled vulnerability exploitation.

The findings can also support organizations in developing AI-specific risk assessments, incident-response procedures, cybersecurity training, human oversight mechanisms, and accountability arrangements. Recent research emphasizes that the rapid adoption of generative AI without adequate security and governance controls can introduce significant organizational vulnerabilities (Humphreys et al., 2024).

5.3 Legal and Regulatory Significance

The study highlights the need to examine whether existing cybercrime and digital-governance legislation adequately addresses AI-assisted and AI-generated cyber activities. Key areas include attribution, liability, evidence, privacy, data protection, synthetic media, automated decision-making, AI system security,

and responsibilities across AI developers, deployers, users, and service providers.

This is particularly relevant because recent scholarship on Pakistan has identified the need for a more comprehensive legal framework capable of addressing AI-related privacy, ethical, regulatory, and societal challenges (Ahmad et al., 2024).

5.4 Policy Significance

The study has direct relevance to Pakistan's evolving digital policy environment. The National Cyber Security Policy 2021 provides a cybersecurity foundation, while the National AI Policy approved in 2025 provides a broader framework for responsible AI development and adoption (MoITT, 2021, 2025). The study can contribute to future policy development by identifying areas where these policy domains need greater integration.

At the international level, the study emphasizes the importance of cross-border cooperation because AI-driven cyber threats cannot always be contained within national jurisdictions. International cooperation concerning cyber incident response, digital evidence, technical standards, information sharing, and AI governance can therefore complement national efforts to strengthen cyber sovereignty.

5.5 National Security Significance

The study is significant for national security because the protection of modern states increasingly depends on resilient digital infrastructure. Strengthening cyber sovereignty does not necessarily require technological isolation; rather, it involves developing sufficient national capacity to protect critical digital assets, regulate high-risk technologies, respond to cyber incidents, and participate effectively in international digital governance.

The central premise of this study is therefore that effective cyber sovereignty depends on the interaction of technological capability, legal authority, institutional capacity, and international cooperation. Understanding these relationships can help provide a more comprehensive foundation for strengthening Pakistan's national security resilience in an increasingly AI-driven cyber environment.

2. Literature Review

2.1 AI-Driven Cyber Threats and the Changing Cybersecurity Environment

Artificial intelligence is transforming cybersecurity from a predominantly rule-based and human-intensive activity toward increasingly automated, adaptive, and data-driven security operations. AI can improve anomaly detection, malware classification, vulnerability assessment, threat intelligence, and incident response. At the same time, malicious actors can exploit similar capabilities to automate reconnaissance, generate convincing phishing messages, identify vulnerabilities, manipulate digital identities, and develop more adaptive attack strategies. Consequently, AI represents a dual-use technology in which defensive and offensive capabilities develop simultaneously (Hashmi et al., 2024; Wen et al., 2025).

Recent research emphasizes that the emergence of large language models (LLMs) has further changed the threat landscape. LLMs can support cybersecurity professionals in security analysis and threat detection, but they can also lower barriers to sophisticated social engineering, malicious-code generation, information gathering, and automated attack workflows. Zhang et al. (2025), in a systematic review of the intersection between LLMs and cybersecurity, demonstrate that the relationship is multidimensional: LLMs can function as defensive instruments while simultaneously creating new attack surfaces and security vulnerabilities.

The literature therefore challenges the traditional assumption that cybersecurity can be managed primarily through perimeter protection and conventional technical controls. AI-enabled threats require continuous monitoring, adaptive risk assessment, model security, data protection, human oversight, and lifecycle governance. Wen et al. (2025) similarly argue that AI can contribute to system-security assurance but that AI-based security mechanisms themselves require systematic evaluation against security standards, legal requirements, and organizational controls. (

This dual-use character has important implications for national security. AI can improve the resilience of national systems, but it can also amplify the speed and scale of attacks

against critical infrastructure. The strategic challenge is consequently not simply whether governments should adopt AI for cybersecurity, but how they can govern its deployment while controlling associated risks.

2.2 AI, Cyber Sovereignty, and National Security

The growing dependence of governments and economies on digital infrastructure has expanded the concept of sovereignty into cyberspace. Cyber sovereignty involves a state's ability to protect its digital infrastructure, regulate activities within its digital domain, safeguard strategic data, and exercise legitimate authority over digital systems. However, cyberspace does not conform neatly to territorial boundaries because data, infrastructure, software, cloud services, and cyber actors can be distributed across multiple jurisdictions.

Recent international legal scholarship argues that existing international legal frameworks remain fragmented in addressing the relationship between cybersecurity and digital sovereignty. Questions surrounding state responsibility, jurisdiction, cyber warfare, attribution, and enforcement remain particularly difficult because cyber operations can be transnational and technically complex. A 2025 analysis in *Computer Law & Security Review* argues that international law requires more adaptable normative approaches for addressing digital sovereignty and cybersecurity.

AI intensifies these challenges. An AI-assisted cyberattack may involve an attacker in one jurisdiction, an AI service in another, cloud infrastructure in a third, and a victim organization in a fourth. Determining responsibility and jurisdiction can therefore become considerably more difficult than in conventional cybercrime. This creates a governance problem at the intersection of national sovereignty and international digital interdependence.

For national security, cyber sovereignty should therefore not be interpreted simply as technological isolation. Rather, it involves the capacity to protect critical digital assets, establish enforceable governance mechanisms, develop national technical capabilities, and participate effectively in international cooperation. This interpretation is particularly relevant to developing countries such as Pakistan, where digital transformation increasingly intersects with national security and economic development.

2.3 AI Governance and Regulatory Capacity

AI governance has developed rapidly in response to concerns about accountability, transparency, privacy, security, fairness, and human oversight. However, the literature indicates that AI governance remains fragmented across organizational, sectoral, national, and international levels.

Batool et al. (2025), through a systematic review of AI governance research, identify governance mechanisms operating at team, organizational, industry, national, and international levels. Their analysis demonstrates that AI governance concerns not only the technology itself but also the actors responsible for governing it, the stages of the AI lifecycle at which governance occurs, and the mechanisms through which accountability is implemented.

This multi-level perspective is particularly relevant to cybersecurity because AI systems can create risks throughout their lifecycle. Governance cannot therefore be limited to the final deployment stage. Data collection, model development, training, testing, deployment, monitoring, updating, and retirement can all create security vulnerabilities.

The literature also identifies a persistent implementation gap between responsible-AI principles and actual organizational practice. A systematic review of empirical studies found that responsible-AI discussions frequently emphasize principles such as accountability, privacy, security, transparency, explainability, and human oversight, while practical implementation remains less developed (Batool et al., 2025). Similar concerns have been identified in research examining real-world responsible-AI practices, where formal principles do not always translate into operational governance mechanisms.

This distinction is important for the present study because the existence of laws or national policies does not necessarily demonstrate effective cyber governance capacity. Effective governance requires institutional resources,

technical expertise, regulatory enforcement, coordination, monitoring, and mechanisms for assigning responsibility.

2.4 Legal Accountability and AI-Enabled Cyber Operations

Accountability is one of the most important challenges in AI governance. Traditional legal systems generally assign responsibility to identifiable human or organizational actors. AI systems, however, may involve complex relationships among developers, model providers, deployers, users, data providers, infrastructure operators, and third-party service providers.

Novelli et al. (2024) conceptualize AI accountability as a multidimensional phenomenon involving responsibility for system development, deployment, operation, and consequences. They emphasize that accountability becomes particularly challenging because AI systems operate within complex sociotechnical environments rather than as isolated technological artifacts.

Recent bibliometric research further demonstrates the rapid growth of scholarship concerning accountability for AI systems and highlights continuing ambiguity concerning who should explain, justify, and assume responsibility for AI-related outcomes (Benlian et al., 2025).

These concerns become more significant in cybersecurity. If an AI-enabled system contributes to a cyber incident, several questions emerge: Who is responsible for the resulting damage? Is responsibility assigned to the developer, system operator, organization, user, or another actor? What evidence is required to establish causation? How should liability be determined when an AI system changes its behavior in response to evolving data?

These questions indicate that conventional cybercrime legislation may require adaptation to AI-enabled forms of cyber activity. Legal frameworks need to address not only unauthorized access and data misuse but also AI-assisted attack mechanisms, synthetic identities, deepfake-enabled fraud, automated exploitation, model manipulation, and AI supply-chain vulnerabilities.

2.5 AI Governance, Privacy, and Data Sovereignty

Data constitute the foundation of contemporary AI systems and therefore represent an important dimension of cyber sovereignty. AI systems require large volumes of data for training, validation, operation, and continuous improvement. Consequently, questions concerning data ownership, cross-border data flows, privacy, security, access, and localization become increasingly important.

The literature on responsible AI consistently identifies privacy and security as core governance requirements. Batool et al. (2025) demonstrate that privacy, accountability, transparency, and security are recurring dimensions of AI governance across different governance levels. Similarly, research on responsible-AI implementation identifies privacy and security as major areas requiring organizational controls rather than purely aspirational principles.

From a national-security perspective, data sovereignty becomes particularly important when strategic or sensitive datasets are processed through foreign cloud infrastructure or AI platforms. Dependence on external technological infrastructure may create additional security and governance concerns, particularly where national authorities have limited control over data storage, processing, model architecture, or system updates.

However, complete data localization is not necessarily synonymous with stronger security. Excessive localization may reduce access to international expertise, technologies, and collaborative threat intelligence. The relevant policy challenge is therefore to balance data protection and national control with interoperability, innovation, and international cooperation.

2.6 Critical Infrastructure and National Security Resilience

AI-driven cyber risks have particular significance for critical information infrastructure. Energy, telecommunications, financial services, transportation, healthcare, government services, and other strategic sectors increasingly rely on interconnected digital systems. An AI-enabled cyberattack against these systems may produce

cascading consequences beyond the directly targeted organization.

Research on AI and information security emphasizes that AI can strengthen real-time threat detection and automated response but also creates new vulnerabilities that require continuous security assurance (Hashmi et al., 2024; Wen et al., 2025). The security of AI systems must therefore be considered alongside the security provided by AI systems.

For Pakistan, this issue is reflected in the National Cyber Security Policy 2021, which places considerable emphasis on critical information infrastructure, incident response, cybersecurity capacity, and institutional coordination. Pakistan's National AI Policy subsequently strengthened this direction by explicitly identifying an AI-based cybersecurity pillar involving AI-enabled threat detection, collaborative threat intelligence, secure data practices, lifecycle security, and human oversight (Ministry of Information Technology and Telecommunication [MoITT], 2025).

The policy evolution indicates that Pakistan recognizes the growing connection between AI and cybersecurity. However, a policy framework does not automatically establish operational resilience. Effective implementation requires skilled personnel, technological infrastructure, regulatory coordination, continuous monitoring, incident-response mechanisms, and measurable security standards.

2.7 International Governance of AI and Cybersecurity

AI-driven cybersecurity challenges increasingly exceed the capacity of individual states. Cyber threats can cross borders rapidly, while AI technologies and digital services are developed through globally distributed supply chains. International governance is therefore necessary for information sharing, technical standards, digital evidence, cybercrime cooperation, responsible AI development, and incident response.

Recent AI governance literature increasingly distinguishes among organizational, national, and international governance mechanisms. Batool et al. (2025) argue that effective AI governance requires coordination across multiple governance levels rather than isolated organizational rules.

At the international level, the challenge is complicated by different regulatory philosophies and national interests. States may prioritize sovereignty, security, innovation, privacy, human rights, or economic competitiveness differently. These differences can make harmonization difficult.

The international governance problem is particularly important for cyber sovereignty because stronger national control must coexist with the need for international cooperation. A state cannot effectively combat transnational AI-enabled cybercrime without mechanisms for cross-border investigation, information sharing, attribution, and technical cooperation. Thus, cyber sovereignty and international cooperation should not necessarily be viewed as contradictory objectives; they can be mutually reinforcing when international cooperation strengthens a state's ability to protect its digital domain.

2.8 Pakistan's Emerging AI and Cyber Governance Environment

Pakistan provides a particularly relevant context for examining AI-driven cyber sovereignty because its policy architecture is developing simultaneously in the cybersecurity and AI domains.

The National Cyber Security Policy 2021 established a national cybersecurity framework emphasizing critical infrastructure protection, institutional coordination, capacity building, incident response, and international cooperation. The policy reflects an understanding that cybersecurity is a national strategic concern rather than solely a technical or organizational issue.

The subsequent National AI Policy represents a further institutional development. Its Secure AI Ecosystem pillar specifically addresses AI-based cybersecurity, threat detection, collaborative defence mechanisms, secure data storage and transmission, lifecycle security, and human oversight.

Nevertheless, important research questions remain. Policy development does not by itself establish whether public and private institutions possess the technical and organizational capacity

required to implement these objectives. Nor does the existence of policy necessarily resolve questions of legal liability, cross-border jurisdiction, AI-assisted cybercrime, data sovereignty, or international cyber governance.

The Pakistani literature remains comparatively limited when viewed against the rapidly expanding international scholarship on AI governance and cybersecurity. Existing studies tend to focus on specific dimensions such as AI regulation, privacy, ethics, cybersecurity, or technological adoption. There is comparatively less empirical work integrating AI-driven cyber threats, legal-regulatory gaps, cyber governance capacity, cyber sovereignty, and national security resilience within one framework.

This represents an important research gap and provides the basis for the present study.

2.9 Critical Synthesis and Research Gap

The literature reveals four important conclusions.

First, AI has created a dual-use cybersecurity environment. The same technologies can strengthen national cyber defence and facilitate increasingly sophisticated attacks. Therefore, cybersecurity strategies must address both the defensive use of AI and the malicious use of AI.

Second, effective AI governance is inherently multilevel. Governance responsibilities extend from developers and organizations to national regulators and international institutions. The literature increasingly supports lifecycle-based governance involving accountability, security, privacy, transparency, and human oversight (Batool et al., 2025; Novelli et al., 2024).

Third, cyber sovereignty increasingly depends on institutional capacity rather than legal authority alone. A state may possess cybersecurity laws and AI policies but remain vulnerable if it lacks technical expertise, enforcement capacity, incident-response capability, critical-infrastructure protection, and international cooperation.

Fourth, there is a contextual and empirical gap concerning Pakistan. International literature provides substantial conceptual discussion of AI governance and cybersecurity, but relatively limited empirical research examines how AI-driven cyber threats interact with legal-regulatory gaps and governance capacity to influence national security resilience in Pakistan.

The present study addresses this gap by proposing an integrated relationship:

AI-Driven Cyber Threats → Cyber Governance Capacity → National Security Resilience

while incorporating legal and regulatory gaps as an additional institutional determinant of cyber governance capacity.

This framework moves beyond viewing AI cybersecurity exclusively as a technological problem and conceptualizes it as a combined technological, legal, institutional, sovereignty, and national-security challenge.

3. Underpinning Theory

3.1 Socio-Technical Systems Theory

Socio-Technical Systems (STS) Theory provides the principal theoretical foundation for this study. STS Theory conceptualizes organizations and complex systems as interactions between technical components and social components, emphasizing that system performance and security depend on the alignment between technology, people, organizational structures, rules, and institutional environments.

The theory is particularly appropriate for AI-driven cybersecurity because AI systems do not operate independently of human and institutional structures. Their security consequences emerge from the interaction among algorithms, data, infrastructure, users, organizations, legal rules, governance institutions, and broader social environments.

In the context of this study, AI-driven cyber threats represent the technical dimension, whereas legal and regulatory arrangements, cybersecurity institutions, professionals, organizational processes, and international governance mechanisms represent the social and institutional dimensions. National security resilience is consequently treated as an outcome of the interaction between these technical and institutional components.

3.2 Applicability of STS Theory

STS Theory provides four important explanations for the proposed research model.

First, technology and institutions are interdependent.

AI-enabled cyber threats cannot be addressed exclusively through technical solutions. Even highly sophisticated detection systems require appropriate institutional structures, skilled personnel, legal authority, response procedures, and accountability mechanisms. Therefore, the impact of AI threats depends partly on the institutional environment in which cybersecurity technologies operate.

Second, regulatory gaps can create sociotechnical vulnerabilities.
Where legal responsibilities are unclear, organizations may face uncertainty concerning data protection, AI deployment, incident reporting, liability, and cyber incident response. From an STS perspective, such weaknesses are not merely legal problems; they become system-level vulnerabilities because technical capabilities are embedded within organizational and regulatory structures.

Third, governance capacity can strengthen resilience
Strong cyber governance capacity can connect technical capabilities with institutional mechanisms. Effective governance can facilitate threat intelligence, incident response, cybersecurity standards, skilled human resources, regulatory enforcement, and coordination among public and private actors. Thus, governance capacity can potentially weaken the adverse effects of AI-driven cyber threats on national security resilience.

Fourth, national security is an emergent sociotechnical outcome.
National security resilience cannot be produced by technology alone. It emerges from the interaction of secure technologies, competent institutions, effective laws, organizational preparedness, skilled professionals, and international cooperation. This makes STS Theory particularly suitable for explaining why technological innovation without corresponding institutional adaptation can generate new vulnerabilities.

Hypotheses

H1: AI-driven cyber threats have a significant negative effect on national security resilience in Pakistan.
H2: Legal and regulatory gaps have a significant negative effect on cyber governance capacity in Pakistan.
H3: Cyber governance capacity has a significant positive effect on national security resilience in Pakistan.
H4: Cyber governance capacity significantly mediates the negative relationship between AI-driven cyber threats and national security resilience in Pakistan.

3. Methodology

3.1 Research Design

The study employed a quantitative, cross-sectional research design to examine the relationships among AI-driven cyber threats, legal and regulatory gaps, cyber governance capacity, and national security resilience in Pakistan. A structured survey approach was used because the study focused on measuring perceptions and relationships among predefined constructs. The analysis was based on illustrative/simulated data and therefore did not represent actual field observations.

3.2 Population

The target population comprised professionals with relevant knowledge of cybersecurity, artificial intelligence, information technology, law, digital governance, and national security in Pakistan. The population included cybersecurity professionals, IT specialists, legal practitioners, university academics, researchers, and public-sector officials involved in digital governance or security-related functions.

3.3 Sampling Technique and Sample Size

A **purposive sampling technique** was employed to select respondents with professional or academic knowledge relevant to AI and cybersecurity governance. A total sample of **300 respondents** was considered for the illustrative analysis. The sample size was regarded as adequate for conducting correlation, multiple regression, and mediation analyses involving the proposed constructs.

3.4 Data Collection Procedures

Data were assumed to have been collected through a structured self-administered questionnaire. Respondents were informed about the academic purpose of the study and were asked to provide responses based on their professional understanding of AI-enabled cybersecurity and governance challenges. Participation was treated as voluntary, and responses were assumed to have been handled confidentially. For this illustrative study, the dataset was simulated to demonstrate the proposed analytical procedures.

3.5 Instruments and Measures

The questionnaire consisted of two sections. The first section collected demographic and professional information, including institutional sector, organizational size, cybersecurity experience, and professional education. The second section measured the four major constructs using multiple items on a five-point Likert scale, ranging from 1 = *Strongly Disagree* to 5 = *Strongly Agree*.

- **AI-Driven Cyber Threats (AICT):** measured through five items assessing AI-enabled phishing, automated attacks, deepfakes, intelligent malware, and other emerging cyber risks.
- **Legal and Regulatory Gaps (LRG):** measured through five items assessing deficiencies in cybersecurity legislation, AI regulation, accountability mechanisms, data governance, and enforcement.
- **Cyber Governance Capacity (CGC):** measured through six items assessing institutional coordination, incident-response capability, regulatory capacity, technical preparedness, human expertise, and governance mechanisms.
- **National Security Resilience (NSR):** measured through six items assessing the ability of institutions and national systems to prevent, withstand, respond to, and recover from AI-enabled cyber threats.

The measurement items were developed from relevant cybersecurity, AI governance, and national security literature and were conceptually aligned with the study's theoretical framework.

3.6 Reliability and Validity

The reliability of the measurement instruments was assessed using Cronbach's alpha. The illustrative results indicated satisfactory internal consistency, with alpha coefficients of 0.89 for AICT, 0.87 for LRG, 0.91 for CGC, and 0.92 for NSR. All coefficients exceeded the commonly accepted threshold of 0.70, indicating adequate internal reliability.

Content validity was established by aligning the measurement items with established concepts in the AI governance, cybersecurity, and national security literature. The questionnaire was also assumed to have been reviewed by subject-matter experts to assess the clarity, relevance, and conceptual coverage of the items. Construct validity was further assessed through the expected relationships among the study variables. Because the dataset was illustrative, these reliability and validity statistics were presented solely to demonstrate the proposed analytical framework and should be recalculated using actual data in an empirical study.

4. Data Analysis

The data were analyzed using descriptive statistics, reliability analysis, Pearson correlation, multiple regression, and mediation analysis. The analysis examined the relationships among AI-driven cyber threats (AICT), legal and regulatory gaps (LRG), cyber governance capacity (CGC), and national security resilience (NSR). The results presented below were based on an illustrative dataset of 300 respondents and were intended solely to demonstrate the proposed empirical analysis.

4.1 Descriptive Statistics and Reliability

Table 1 presents the descriptive statistics and internal consistency estimates for the study variables.

Table 1. Descriptive Statistics and Reliability

Variable	Items	Mean	SD	Cronbach's α
AI-Driven Cyber Threats (AICT)	5	3.71	0.78	0.89

Variable	Items	Mean	SD	Cronbach's α
Legal and Regulatory Gaps (LRG)	5	3.64	0.81	0.87
Cyber Governance Capacity (CGC)	6	3.18	0.84	0.91
National Security Resilience (NSR)	6	3.27	0.80	0.92

The mean scores indicated that respondents perceived AI-driven cyber threats and legal-regulatory gaps relatively strongly, with mean values of 3.71 and 3.64, respectively. In contrast, cyber governance capacity recorded a lower mean of 3.18, suggesting comparatively moderate perceived institutional capacity to manage emerging cyber risks. National security resilience had a mean of 3.27, indicating a moderate level of perceived resilience. The Cronbach's alpha coefficients ranged from 0.87 to 0.92, exceeding the commonly accepted threshold of 0.70 and indicating strong internal consistency among the measurement items.

4.2 Correlation Analysis

Pearson correlation analysis was conducted to determine the direction and strength of associations among the principal variables.

Table 2. Pearson Correlation Matrix

Variable	AICT	LRG	CGC	NSR
AICT	1			
LRG	.48**	1		
CGC	-.46**	-.55**	1	
NSR	-.52**	-.43**	.61**	1

Note. p < .01 (two-tailed).

AI-driven cyber threats were negatively associated with cyber governance capacity ($r = -.46$, $p < .01$) and national security resilience ($r = -.52$, $p < .01$). This indicated that higher perceived exposure to AI-enabled cyber threats was associated with lower perceived governance capacity and resilience. Legal and regulatory gaps also demonstrated significant negative relationships with cyber governance capacity ($r = -.55$, $p < .01$) and national security resilience ($r = -.43$, $p < .01$). Conversely, cyber governance capacity had a relatively strong positive relationship with national security resilience ($r = .61$, $p < .01$). Overall, the correlation pattern was consistent with the proposed theoretical model and provided preliminary support for the hypothesized relationships.

4.3 Multiple Regression Analysis

Multiple regression analysis was conducted to assess the effects of AI-driven cyber threats and cyber governance capacity on national security resilience.

Table 3. Regression Results for National Security Resilience

Predictor	β	SE	t	p
AI-Driven Cyber Threats (AICT)	−0.34	0.061	−5.57	< .001
Cyber Governance Capacity (CGC)	0.48	0.058	8.28	< .001

Model statistics: $R^2 = .47$; Adjusted $R^2 = .46$; $F = 65.40$, $p < .001$.

The regression model explained approximately 47% of the variance in national security resilience, indicating substantial explanatory power for the illustrative model. AI-driven cyber threats had a statistically significant negative effect on national security resilience ($\beta = -0.34$, $p < .001$). Thus, greater exposure to AI-enabled cyber threats was associated with lower national security resilience. Therefore, H1 was supported in the illustrative analysis.

Cyber governance capacity demonstrated a statistically significant positive effect on national security resilience ($\beta = 0.48$, $p < .001$). The positive coefficient indicated that stronger governance capacity was associated with greater national security resilience. The magnitude of the coefficient also suggested that governance capacity represented an important explanatory factor in the model. Accordingly, H3 was supported.

The overall regression model was statistically significant, $F = 65.40$, $p < .001$, indicating that AI-driven cyber threats and cyber governance capacity jointly explained a significant proportion of the variation in national security resilience.

4.4 Effect of Legal and Regulatory Gaps on Cyber Governance Capacity

A separate regression analysis was used to examine whether legal and regulatory gaps predicted cyber governance capacity.

Table 4. Regression Result for Cyber Governance Capacity

Predictor	β	Direction	p
Legal and Regulatory Gaps (LRG)	−0.55	Negative	< .001

Legal and regulatory gaps demonstrated a significant negative relationship with cyber governance capacity ($\beta = -0.55$, $p < .001$). The finding indicated that weaknesses in legislation, regulatory clarity, accountability mechanisms, data governance, and enforcement were associated with reduced capacity to manage AI-related cyber risks. Consequently, H2 was supported in the illustrative analysis.

The result also highlighted the institutional dimension of cybersecurity governance. Technical capabilities alone may not provide sufficient national protection when legal responsibilities, regulatory authority, enforcement mechanisms, and accountability arrangements remain fragmented.

4.5 Mediation Analysis

A mediation analysis was conducted to examine whether cyber governance capacity mediated the relationship between AI-driven cyber threats and national security resilience.

Table 5. Mediation Analysis

Relationship	β	p	Interpretation
AICT → CGC	−0.46	< .001	Significant negative
CGC → NSR	0.48	< .001	Significant positive
AICT → NSR (direct effect)	−0.34	< .001	Significant negative
AICT → CGC → NSR (indirect effect)	−0.22	< .001*	Significant indirect effect

Note. *Mediation result; an actual study should report bootstrapped confidence intervals for the indirect effect.*

AI-driven cyber threats had a significant negative relationship with cyber governance capacity ($\beta = -0.46$, $p < .001$), while cyber governance capacity had a significant positive relationship with national security resilience ($\beta = 0.48$, $p < .001$). The direct effect of AI-driven cyber threats on national security resilience remained negative and significant after the mediator was included ($\beta = -0.34$, $p < .001$). The indirect effect through cyber governance capacity was approximately −0.22.

The simultaneous presence of a significant direct effect and a significant indirect effect suggested partial mediation. In substantive terms, AI-driven cyber threats may reduce national security resilience both directly and indirectly by placing pressure on existing cyber governance capabilities. Thus, H4 was supported in the analysis.

4.6 Hypotheses Summary

Table 6. Summary of Hypothesis Testing

Hypothesis	Relationship	Result
H1	AICT → NSR (−)	Supported
H2	LRG → CGC (−)	Supported
H3	CGC → NSR (+)	Supported
H4	AICT → CGC → NSR (− indirect effect)	Supported

4.7 Overall Findings

The overall findings demonstrated a coherent pattern in which technological threats, institutional weaknesses, and national security resilience were interconnected. Higher levels of AI-driven cyber threats were associated with weaker national security resilience, whereas stronger cyber governance capacity was associated with greater resilience. At the same time, legal and regulatory gaps were negatively associated with governance capacity, indicating that cybersecurity resilience depended not only on technological capabilities but also on the adequacy of institutional and regulatory arrangements.

The mediation results further indicated that cyber governance capacity represented an important mechanism linking emerging AI-driven cyber threats with national security resilience. This finding was consistent with the **Socio-Technical Systems perspective**, under which cybersecurity outcomes emerge from interactions among technological systems, human actors, organizational structures, institutional arrangements, and regulatory frameworks. The results therefore suggested that strengthening technical cybersecurity capabilities without simultaneously improving governance, legal clarity, institutional coordination, and accountability may provide an incomplete response to AI-enabled national security risks.

5. Discussion

The illustrative findings provided consistent support for the proposed relationships among AI-driven cyber threats, legal and regulatory gaps, cyber governance capacity, and national security resilience. The results indicated that AI-driven cyber threats were negatively associated with national security resilience, while cyber governance capacity demonstrated a positive relationship with resilience. Legal and regulatory gaps were also negatively associated with cyber governance capacity. These findings suggested that the security implications of AI were shaped not only by technological capabilities but also by the institutional and regulatory environment within which those technologies were developed and deployed.

The negative relationship between AI-driven cyber threats and national security resilience was theoretically consistent with recent cybersecurity literature emphasizing the dual-use nature of AI. Hashmi et al. (2025) highlighted that AI can strengthen information-security capabilities while simultaneously creating new vulnerabilities and attack possibilities. The illustrative negative coefficient between AI-driven cyber threats and resilience therefore supported the broader proposition that increasing AI-enabled threat exposure can place additional pressure on national cybersecurity systems.

The finding concerning legal and regulatory gaps was particularly important. The illustrative results showed a significant negative relationship between legal and regulatory gaps and cyber governance capacity. This corresponds with Batool et al. (2025), whose systematic review demonstrated that AI governance requires coordinated arrangements concerning who governs AI, what is governed, when governance occurs, and how governance is implemented

across organizational, national, and international levels. The present model extended this perspective by treating regulatory adequacy as an institutional condition that can influence the effectiveness of cyber governance rather than as an isolated legal issue.

The positive relationship between cyber governance capacity and national security resilience further suggested that technological preparedness needs to be supported by institutional capacity. The finding is consistent with the wider AI-governance literature, which increasingly emphasizes accountability, organizational responsibility, transparency, and appropriate governance mechanisms. Novelli et al. (2024), for example, conceptualized AI accountability as requiring mechanisms through which responsibility can be assigned and implemented across AI-related activities. Thus, stronger governance capacity may help organizations and public institutions translate technological capabilities into more resilient cybersecurity outcomes.

The mediation result provided an additional theoretical contribution. The illustrative analysis indicated that cyber governance capacity partially mediated the relationship between AI-driven cyber threats and national security resilience. This implies that the consequences of AI-enabled cyber threats may depend partly on the capacity of institutions to identify risks, coordinate responses, establish accountability, and implement appropriate controls. Rather than viewing AI-driven cyber risk as a purely technological problem, the model positioned governance capacity as a mechanism through which technological threats can influence broader security outcomes.

5.1 Theoretical Implications

The findings supported the application of Socio-Technical Systems (STS) Theory to AI-driven cybersecurity and national security. STS theory proposes that organizational and system outcomes emerge from interactions between technical and social components. In the present framework, AI-driven cyber threats represented the technological dimension, while legal and regulatory gaps and cyber governance capacity represented institutional and organizational dimensions. National security resilience consequently emerged as a broader system-level outcome.

The mediation finding strengthened this theoretical interpretation by demonstrating, within the illustrative model, that the relationship between technological threats and security outcomes was not entirely direct. Governance capacity provided an institutional pathway through which technological risks could be translated into resilience outcomes. This complements recent AI-governance research emphasizing that effective AI governance operates across multiple levels rather than through technical controls alone.

The study therefore extended the application of STS thinking to the intersection of AI, cybersecurity, and cyber sovereignty by emphasizing the interaction of technology, law, institutions, organizational capability, and national security.

6. Conclusion

The study examined the emerging relationship between AI-driven cyber threats, legal and regulatory gaps, cyber governance capacity, and national security resilience in Pakistan. The illustrative analysis indicated that AI-driven cyber threats were negatively associated with national security resilience, while stronger cyber governance capacity was positively associated with resilience. Legal and regulatory gaps were negatively associated with governance capacity, and governance capacity partially mediated the relationship between AI-driven cyber threats and national security resilience.

Overall, the findings suggested that AI-driven cyber sovereignty should not be approached solely as a matter of acquiring advanced technologies. Effective national cybersecurity also requires coherent legislation, institutional coordination, accountability mechanisms, skilled human resources, incident-response capacity, and appropriate national and international governance arrangements. Recent AI-governance research similarly emphasizes the need to address governance across organizational, national, and international levels.

The study consequently positioned cyber governance capacity as an important bridge between emerging technological threats and

national security resilience. However, because the numerical analysis was based on simulated data, the findings should be regarded as an illustrative demonstration of the proposed framework rather than as verified empirical evidence concerning Pakistan.

7. Implications

7.1 Theoretical Implications

The study contributes to theory by applying Socio-Technical Systems Theory to AI-driven cybersecurity and cyber sovereignty. It demonstrates conceptually that national security outcomes may emerge from interactions between technological risks and institutional arrangements. The mediation framework further positions governance capacity as a mechanism connecting AI-related threats with security resilience.

7.2 Managerial Implications

For organizational managers and cybersecurity leaders, the findings highlight the importance of integrating AI risk management into broader enterprise cybersecurity governance. Organizations should establish clear responsibilities for AI-related risks, conduct regular AI-enabled threat assessments, strengthen incident-response mechanisms, and ensure coordination among IT, cybersecurity, legal, compliance, and senior management functions.

7.3 Practical Implications

Cybersecurity institutions should move beyond conventional perimeter-based security approaches and incorporate AI-specific threat intelligence, automated monitoring, adversarial testing, deepfake detection, identity protection, and continuous vulnerability assessment. Human expertise should remain central because automated systems require appropriate oversight, validation, and escalation mechanisms.

7.4 Policy Implications

Policymakers should develop a coordinated governance framework addressing AI-enabled cyber risks, data protection, accountability, critical infrastructure, incident reporting, cross-border cybercrime, and responsible use of AI in national-security contexts. Governance arrangements should also clarify institutional responsibilities and mechanisms for cooperation between government agencies, private-sector organizations, academia, and international partners. The international and multi-level nature of AI governance identified in recent literature reinforces the importance of coordination beyond individual organizations or jurisdictions.

8. Recommendations

1. **Develop an integrated AI-cybersecurity governance framework:** National institutions should align AI governance, cybersecurity, data protection, and national-security policies within a coherent institutional framework.
2. **Strengthen legal accountability:** Laws and regulations should clearly define responsibilities for AI developers, deployers, organizations, and relevant public authorities, particularly where AI systems create or amplify cybersecurity risks.
3. **Establish continuous AI threat-intelligence systems:** Public and critical-infrastructure organizations should continuously monitor AI-enabled phishing, automated attacks, synthetic media, intelligent malware, and other emerging threats.
4. **Strengthen institutional cyber capacity:** Investment should focus not only on cybersecurity technologies but also on specialized personnel, training, incident-response teams, forensic capabilities, and inter-agency coordination.
5. **Conduct regular AI security audits:** Organizations using high-impact AI systems should periodically evaluate vulnerabilities, model behavior, data security, access controls, and potential misuse.
6. **Promote public-private collaboration:** Government agencies should establish structured information-sharing mechanisms with banks, telecommunications companies, technology firms, universities, and critical-infrastructure operators.
7. **Develop AI-specific incident-response protocols:** National and organizational response plans should explicitly address AI-enabled attacks, synthetic identity manipulation,

deepfakes, automated social engineering, and other emerging attack vectors.

8. **Strengthen international cooperation:** Pakistan should continue developing mechanisms for information sharing, cybercrime cooperation, technical capacity building, and responsible AI governance with relevant international partners.

9. **Promote professional education and research:** Universities and professional institutions should expand interdisciplinary programs combining AI, cybersecurity, law, governance, and national-security studies.

10. **Introduce governance evaluation mechanisms:** Cyber governance capacity should be periodically assessed using measurable indicators covering institutional coordination, regulatory effectiveness, technical preparedness, human resources, and incident-response performance.

9. Limitations and Future Directions

The study had several limitations. The proposed design was cross-sectional, which would limit the ability of a future empirical study to establish temporal or causal relationships among the variables. Third, purposive sampling could introduce selection bias and limit generalizability across Pakistan's diverse public and private sectors.

The study relied primarily on perceptual measures. Future research should incorporate objective cybersecurity indicators, such as incident frequency, response time, security investment, vulnerability rates, and recovery performance. Fifth, the model focused on a limited number of constructs. Future studies could incorporate variables such as data sovereignty, AI literacy, organizational cybersecurity maturity, institutional trust, international cyber cooperation, privacy protection, and critical-infrastructure dependence.

Future empirical research should collect data from cybersecurity professionals, regulators, technology firms, critical-infrastructure organizations, legal experts, and public-sector institutions across different regions of Pakistan. Structural equation modeling could be used to test the complete theoretical model, while bootstrapping should be employed to establish confidence intervals for mediation effects. Longitudinal and mixed-method studies could further examine how AI-driven threats and governance capacity evolve over time.

Future research could also conduct comparative studies between Pakistan and other developing economies to identify similarities and differences in AI governance, cyber sovereignty, regulatory capacity, and national security resilience. Such research would provide stronger empirical foundations for developing context-sensitive AI cybersecurity and governance frameworks.

References

Ali, S., Wang, J., & Leung, V. C. M. (2025). AI-driven fusion with cybersecurity: Exploring current trends, advanced techniques, future directions, and policy implications for evolving paradigms—A comprehensive review. *Information Fusion, 110*, 102922.

Bartsch, S. C., Nguyen, L. H., Schmidt, J.-H., Du, G., Adam, M., Benlian, A., & Sunyaev, A. (2025). The present and future of accountability for AI systems: A bibliometric analysis. *Information Systems Frontiers, 27*, 2463–2484. https://doi.org/10.1007/s10796-025-10636-9

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics, 5*, 3265–3279. https://doi.org/10.1007/s43681-024-00653-w

Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial Intelligence Trust, Risk and Security Management (AI TRiSM): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications, 240*, 122442. https://doi.org/10.1016/j.eswa.2023.122442

Hashmi, E., Yamin, M. M., & Yayilgan, S. Y. (2025). Securing tomorrow: A comprehensive survey on the synergy of artificial intelligence and information security. *AI and Ethics, 5*, 1911–1929. https://doi.org/10.1007/s43681-024-00529-z

Knight, S., Shibani, A., & Vincent, N. (2025). Ethical AI governance: Mapping a research ecosystem. *AI and Ethics, 5*, 841–862. https://doi.org/10.1007/s43681-023-00416-z

Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: What it is and how it works. *AI & Society, 39*, 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

Okdem, S., & Okdem, S. (2024). Artificial intelligence in cybersecurity: A review and a case study. *Applied Sciences, 14*(22), 10487. https://doi.org/10.3390/app142210487

Ortega-Bolaños, R., Bernal-Salcedo, J., Germán Ortiz, M., Galeano Sarmiento, J., Ruz, G. A., & Tabares-Soto, R. (2024). Applying the ethics of AI: A systematic review of tools for developing and assessing AI-based systems. *Artificial Intelligence Review, 57*, 110. https://doi.org/10.1007/s10462-024-10740-3

Pant, A., Hoda, R., Tantithamthavorn, C., & Turhan, B. (2024). Ethics in AI through the practitioner's view: A grounded theory literature review. *Empirical Software Engineering, 29*, 67.

Roberts, H. (2024). Digital sovereignty and artificial intelligence: A normative approach. *Ethics and Information Technology, 26*, 70. https://doi.org/10.1007/s10676-024-09810-5

Schuett, J., Reuel, A.-K., & Carlier, A. (2025). How to design an AI ethics board. *AI and Ethics, 5*, 863–881. https://doi.org/10.1007/s43681-023-00409-y

Szadeczky, T., & Bederna, Z. (2025). Risk, regulation, and governance: Evaluating artificial intelligence across diverse application scenarios. *Security Journal, 38*, 35. https://doi.org/10.1057/s41284-025-00495-z

Vallor, S., & Vierkant, T. (2024). Find the gap: AI, responsible agency and vulnerability. *Minds and Machines, 34*, 20. https://doi.org/10.1007/s11023-024-09674-0

Wen, S.-F., Shukla, A., & Katt, B. (2025). Artificial intelligence for system security assurance: A systematic literature review. *International Journal of Information Security, 24*, 43. https://doi.org/10.1007/s10207-024-00959-0

Zhang, J., Bu, H., Wen, H., Liu, Y., Fei, H., Xi, R., Li, L., Yang, Y., Zhu, H., & Meng, D. (2025). When LLMs meet cybersecurity: A systematic literature review. *Cybersecurity, 8*, 55. https://doi.org/10.1186/s42400-025-00361-w

Abdullah, M., Khan, M. A., & colleagues. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management, 8*(2), 100063. https://doi.org/10.1016/j.dim.2023.100063

Lezzi, M., Montefusco, P., Lazoi, M., & Corallo, A. (2025). AI-based cybersecurity for a sustainable digital industry: Systematic literature review and future research directions. *Journal of Industrial Information Integration, 48*, 100980. https://doi.org/10.1016/j.jii.2025.100980

Wang, Y., Zowghi, D., & Bano, M. (2024). Responsible artificial intelligence governance: Emerging challenges and research directions. *AI and Ethics, 4*, 1–15.

Zhao, Y., Liu, Y., & Chen, X. (2024). The weaponization of artificial intelligence in cybersecurity: A systematic review. *Procedia Computer Science, 239*, 547–555. https://doi.org/10.1016/j.procs.2024.06.206

IEER Institute for Excellence in Education & Research