International Journal of Science and Engineering Applications
Volume 15-Issue 07, 52 – 58, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1507.1009

LLM Cybersecurity Governance for Banking Operations

Abdul Hasham
Department of Information Technology
Campbellsville University
KY, USA

***Abstract*—** By automating client engagement, fraudulent activities detection, compliance, credit evaluation, finance reports, and risk management processes, large language models (LLMs) can rapidly transform the banking industry. Nevertheless, there are considerable risks of cybersecurity and governance of LLMs in the context of financial institutions, such as insider threats, fast injection attacks, hallucinated financial responses, data leakage, and lack of compliance. AI systems that can ensure privacy, reliability, security, and integrity while meeting financial standards are critical components of banking systems. Focusing on runtime protection measures, access controls, and compliance policies, the article provides insight into cybersecurity governance frameworks for LLM financial systems. A framework for governance based on rapid validation, restriction of access, auditing, encryption, compliance monitoring, and human supervision is developed following the review of literature and analysis of risks faced. The results obtained during experiments reveal that cybersecurity governance frameworks help to mitigate risks associated with the application of artificial intelligence technologies while significantly improving operational security and compliance within the banking industry.

***Keywords*—** LLMs in banking Cybersecurity in large language models Governance guardrails for AI-powered financial organizations Compliance audit of access control Rapid injection defense Hallucination detection in secure AI systems financial risk management Zero trust governance for AI with accountability.

I. Introduction

Digital transformation is being enabled by LLMs and AI in the field of banking [1]. In today's world, LLMs are employed in banks for purposes of credit scoring, finance forecasting, fraud detection, regulation reporting, customer service, and intelligent decision-making. Financial data are processed in real-time in large volumes. This positively impacts customer satisfaction and efficiency, but there are a lot of challenges associated with implementing LLMs in the banking sector. Firstly, contrary to the usual software programs, LLMs generate dynamically changing output results that may involve some biases, hallucinations, and, most importantly, a severe data breach. Moreover, the companies working in the financial industry have to abide by some legal standards, such as GDPR, PCI-DSS, ISO 27001, and SOC 2, implying that the firm should manage the data properly and responsibly [2]. For this reason, the concept of cybersecurity governance becomes essential for AI incorporation into the financial systems. These are the main aspects considered in the report regarding cybersecurity governance strategies that aim at securing LLMs usage in finance.

II. Literature Review

The widespread application of LLMs in the banking and finance sector is attributed to its ability to automate decisions, engage with customers, create compliance reports, and detect fraud. As a means of exploring how LLMs can be integrated into the financial environment, various aspects of AI governance, cybersecurity, access control, and regulation have been studied. The importance of developing transparent, explainable, and accountable AI models that operate within the law has been underscored by recent studies. Based on recent studies, such AI models governed by principles of governance could minimize risks, enhance cybersecurity, and install trust [3]. This section provides an overview of past research on AI governance in banking and runtime protection, access control, and compliance monitoring systems that facilitate the implementation of LLMs in banks.

A. AI Governance in Banking

The aim of governance of AI applications within the banking sector is to offer operational security, accountability, transparency, and fairness of such systems. AI applications should conform to all applicable legal, ethical, and cyber security principles [4]. In order to establish confidence in abilities of AI algorithms, researchers have proposed governance frameworks which involve risk management, explain ability, auditability, and policy enforcement. Governance frameworks have been proven to be necessary in preventing operational mistakes, reducing algorithmic bias, and ensuring responsible use of AI systems within financial institutions. Moreover, regulators put an increasing amount of pressure on banks to establish accountability frameworks for automated financial systems and explain decision-making process performed by AI technologies.

B. Runtime Guardrails for LLM Security

Security measures that are known as runtime guardrails aim at controlling LLMs. These guardrails restrict prohibited topics, recognize harmful responses, filter false prompts, and prevent hallucinations. Recent research suggests that malicious inputs and fast-injection attacks might affect behaviour of LLMs and disclose personal financial information. These kinds of problems can be avoided by the implementation of measures such as output moderation, semantic filters, and rule-based validation engines. Through the implementation of such runtime safeguards, the cybersecurity of AI in finance is enhanced, and the likelihood of regulation breaches caused by financial advice from AI systems is reduced [5].

C. Access Control and Identity Management

The application of access controls is crucial to ensuring security in financial LLMs from any type of misuse and internal threats [6]. Experts recommend the usage of zero trust security framework, ABAC, and RBAC in banking AI systems. Access is controlled depending on role, attributes, contextual constraints, and authentication status. The implementation of privilege separation, continuous identity verification, and multi-factor authentication increases the level of cybersecurity in banking AI systems. As shown in prior studies, secure access control helps avoid data exfiltration, illegal prompt operation, and the leakage of confidential information of clients stored in LLMs.

D. Compliance Monitoring and Auditability

Monitoring compliance means checking whether the LLMs used by banks conform to financial and cybersecurity guidelines [7]. As noted from the findings of the research, the decision-making process using AI should respect guidelines such as GDPR, PCI-DSS, ISO 27001, SOC 2, and SR 11-7. Monitoring compliance includes logging prompts, output, action, and violations of policy in order to assist during regulatory examinations and governance review. Researchers note that monitoring compliance continuously makes it easier for financial organizations to detect abnormalities and policy violations. Thus, audit systems in banking AI operations help in improving reliability.

Table 1: Literature Review Summary

Author / Study	Research Area	Key Findings	Limitations
García-Llorente et al.	AI Governance in Banking	Proposed accountability frameworks for financial AI systems	Limited runtime monitoring
NeMo Guardrails Research	Runtime LLM Security	Introduced programmable AI guardrails and moderation controls	Generic industry focus
Jun Xu Financial AI Survey	Banking LLM Applications	Discussed LLM adoption in financial services	Limited governance analysis
Financial Compliance Studies	Regulatory AI Governance	Emphasized auditability and explain ability requirements	Weak access-control integration
Zero-Trust AI Research	Identity and Access Security	Recommended RBAC and Zero-Trust for AI systems	High implementation complexity
Fintech Cybersecurity Frameworks	Banking AI Protection	Proposed layered cybersecurity defence models	Limited scalability evaluation

III. Problem Statement and Motivation

In their application within financial processes, large language models (LLMs) bring forth revolutionary strengths and cybersecurity threats. The use of artificial intelligence technologies is growing rapidly as the norm in banks when it comes to loan processing, fraud prevention, client support, regulatory reporting, and financial analysis. The stochastic behaviour of LLMs, however, creates important governance risks that current cybersecurity practices cannot properly handle. In addition, LLMs have the potential to hallucinate, divulge sensitive details, and exhibit random behaviour upon exposure to harmful stimuli, unlike the deterministic behaviour of banks' software applications [8].

Financial data, transactional information, and highly confidential personal information about their clients are all stored in modern banking institutions. Hackers can exploit prompt injection weaknesses, bypass access limitations, modify results, or even obtain personal information if the system does not have governance control in place [9]. Banks must ensure that AI systems are accountable, transparent, auditable, and interpretable in compliance with GDPR, PCI-DSS, ISO 27001, SOC 2, and SR 11-7 regulations. Financial penalties, legal sanctions, and a loss of customer trust will follow if any such guidelines are violated.

Also, the absence of a coherent security governance strategy that will facilitate integrating identity management, runtime protection, monitoring of compliance and human intervention into one single security system is yet another issue [10]. Using standalone security measures, which make it difficult to enforce policies and monitor them in real time, is the prevailing approach. As banks are increasingly using AI-based technology, it calls for the creation of security models for LLM in terms of governance. Thus, there is motivation to develop security models for banks' LLM operations, which will be safe, ethical, and legal.

A. Prompt Injection and Adversarial Threats

Prompt injection vulnerabilities can allow attackers to inject instructions into user requests, thereby affecting the functionality of LLM [11]. This vulnerability could enable attackers to bypass some limitations, acquire private data, or deliver fake financial responses. Banking systems are more susceptible to LLM due to their access to private customer and financial data. Prompt attacks can undermine the credibility of consumers and put the operations at risk. Thus, in order to ensure the security of banking AI systems, prompt filtering measures need to be taken into account.

B. Data Leakage and Privacy Risks

LLMs may accidentally reveal confidential information about their clients due to the insecurity of prompts, memory retention, or any other unauthorized access. Financial organizations deal with different types of data, which include the account numbers of their clients and financial transactions among others, making it very sensitive information that can be subject to a breach. [12] A data breach event is likely to lead to privacy violation, legal repercussions, and reputation concerns.

International Journal of Science and Engineering Applications
Volume 15-Issue 07, 52 – 58, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1507.1009

C. Lack of Explain ability and Auditability

Laws mandate that banks employ an AI system that can offer auditability and explainability for their decision-making process [13]. This requirement makes it difficult to use LLMs that are black boxes as financial firms cannot give an adequate explanation of their decision. They cannot track decisions made by the model without logging and auditing measures in place.

D. Motivation for Governance-Centric Frameworks

This research project's main objective is the development of a cybersecurity framework with governance principles at its core to ensure secure and trusted LLM adoption [14]. Specifically, the proposed framework will leverage guardrails, role-based access control, compliance, encryption, audit logging, and human involvement into one solution.

Table 2: Major Challenges and Governance Requirements in Banking LLM Systems

Security Challenge	Operational Impact	Governance Requirement	Expected Benefit
Prompt Injection	**Manipulated outputs**	**Prompt filtering & validation**	**Safer AI interaction**
Hallucinated Responses	**Incorrect financial advice**	**Output verification**	**Improved reliability**
Data Leakage	**Privacy violations**	**Encryption & DLP**	**Secure customer data**
Unauthorized Access	**Insider threats**	**RBAC & MFA**	**Controlled system access**
Lack of Audit Trails	**Compliance failure**	**Immutable logging**	**Regulatory transparency**
Model Bias and Drift	**Unfair financial decisions**	**Continuous monitoring**	**Ethical AI governance**

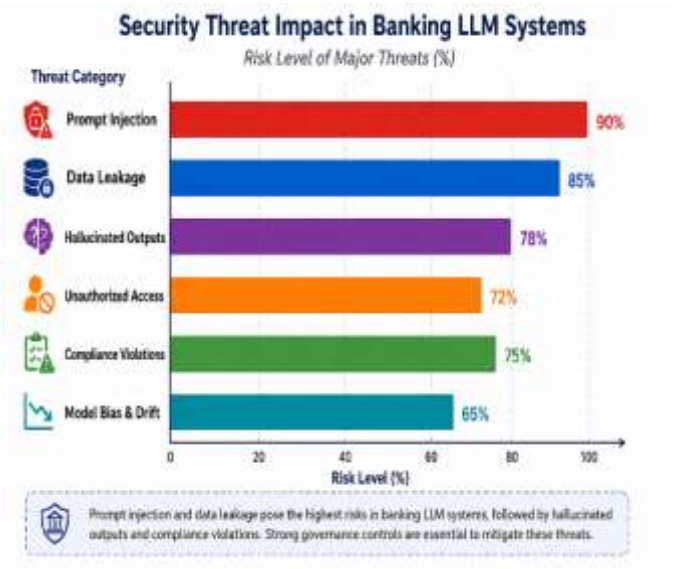

Figure 1: Security Threat Impact in Banking LLM Systems

IV. Related Work

Secure AI framework development for financial organizations is made easy due to recent advances in studies focused on security and governance of large language models (LLMs) [15]. The main topics under investigation at present are related to governance, runtime guardrails, compliance monitoring, threat modelling, and security access control of banking systems. NVIDIA NeMo Guardrails currently has adjustable safety settings, which help prevent harmful LLM outputs and implement policy-based interaction. Financial cybersecurity studies have also involved an examination of audit logging mechanisms for enhanced transparency of AI banking applications, defence mechanisms for fast injection attacks, and minimizing hallucination generation. Further studies were done in relation to role-based access control and zero-trust architectures, aimed at protecting sensitive data stored in banks from any breaches or insider attacks. Within the scope of the research related to regulations such as GDPR, PCI-DSS, ISO 27001, and SOC 2, the necessity of accountability and explain ability was identified. Instead of governance-oriented architectural designs, most modern solutions rely on separate security measures. Thus, by presenting a unified governance-focused design that will facilitate secure banking LLMs with runtime guardrails, access control, compliance, audit logging, and human monitoring, the proposed research extends previous works.

International Journal of Science and Engineering Applications
Volume 15-Issue 07, 52 – 58, 2026, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1507.1009

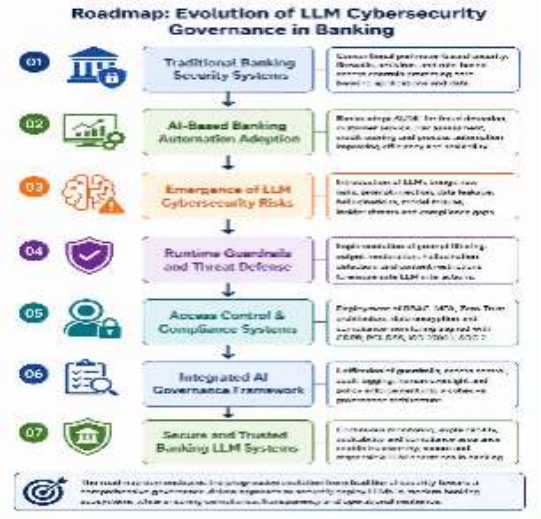

Figure 2: Evolution of LLM Cybersecurity Governance in Banking

V. METHODOLOGY / PROPOSED FRAMEWORK

The proposed framework highlights a cybersecurity model that integrates governance principles for the proper deployment of LLMs in banking activities. In this regard, runtime safeguards, access management, compliance measures, encryption processes, audit logging, and human intervention have been consolidated within a single governance layer. This strategy intends to reduce operational hazards concerning the utilization of AI-based banking solutions while preserving confidentiality, integrity, accountability, and regulatory compliance [16]. Whereas previous cyber security frameworks were unable to do so, the new cyber security framework would allow for monitoring the activities of the LLMs, verification of user inputs, validation of the outputs, and the enforcement of governance requirements on an ongoing basis. The framework enables the clear analysis of financial transaction requirements such as GDPR, PCI-DSS, ISO 27001, SOC 2, and SR 11-7. Furthermore, the approach addresses issues such as quick injection, hallucinations, insider threats, and unauthorised access by leveraging multiple security layers in the governance framework.

A. Input Validation and Prompt Security

Two objectives that constitute the basic level of this framework are security and rapid verification. The filters will be employed for analyzing the prompts of users and detecting any hazardous prompts, harmful instructions, toxic content, and unauthorized requests [17]. Algorithms for cleaning the prompts will discard wrong instructions prior to processing by the LLM engine. In addition, we apply the approaches of context-based and semantic verification of prompts in accordance with banking security requirements.

B. Identity and Access Management

Identity and Access Management (IAM) techniques are used in controlling the way users gain access to the financial LLM [18]. The model has been designed based on Zero Trust, Multi Factor Authentication (MFA), and Role Based Access Control (RBAC). These features authenticate the identity of the user and allow access depending on their roles. Users like employees, customers, auditors, and administrators have permission to access the model according to their roles.

C. Runtime Guardrails and Output Monitoring

In order to prevent any harm to the output that might be provided in response to the prompt, runtime guardrails are u sed on a regular basis by LLMs [19]. This process includes f iltering toxins, assessing finance strategies, detecting halluci nations, and modifying the output. This is because the soluti on either rejects or diverts the output to the human validator in case there are any concerns regarding non-compliance wit h the regulations.

D. Compliance Monitoring and Audit Logging

Compliance Monitoring: Each step taken by the LLM is monitored to ensure that both the cybersecurity regulations as well as financial regulations are adhered to [20]. The approach employs the use of immutable audit logs in monitoring user behavior, rule violations, prompt, outputs, and access logs. The interaction between the end user and the AI system is regularly audited to ensure that the regulatory requirements are followed.

E. Human Oversight and Decision Validation

The highest form of governance in this model is human s upervision. Prior to implementation, the decisions made by t he banks using the LLMs are reviewed by a human being. In human-supervised AI, the outcome can either be accepted, r ejected, or altered [21]. This strategy reduces the risk of mak ing mistakes through AI and enhances accountability in the banks' AI operations.

FIGURE 3: CORE COMPONENTS OF BANKING LLM GOVERNANCE FRAMEWORK

VI. EXPERIMENT / CASE STUDY / SIMULATION

The following experiment was carried out in order to prove the efficiency of the proposed architecture, where a simulated banking setting utilizing the LLM system for provision of financial assistance and customer support was created. The purpose of this experiment was to evaluate the architecture's ability to detect potential security threats, control access to information, ensure compliance, and enhance the reliability of banking services. Real-life

examples of queries were employed during the experiment in various areas of banking, such as questions related to transaction processing, request for loan processing, detection of fraud, compliance issues, and access control for employees. The types of attacks that were tested included prompt injection, unauthorized access, extraction of data, and hallucination prompts. It turned out that using such measures as guardrails, identity management, compliance monitoring, and auditing significantly improved cybersecurity performance and mitigated risk. The framework additionally improved transparency and regulatory compliance by constantly auditing and creating immutable logs of AI interaction within banking institutions [22].

A. Simulation Environment

The security and governance components were incorporated in the banking LLM framework. The simulation included artificial data for a variety of banking and financial transactions features like customer dealings, accounts, and documents. The banking LLM environment included several users, such as customers, employees of the bank, auditors, compliance officers, and administrators of the banking LLM system. Security assessment tools were used to generate adversarial inputs in order to evaluate the governance system [23].

Key Components

- Banking customer-support chatbot
- Fraud detection assistance module
- Compliance reporting engine
- Access-control and identity-management system
- Runtime guardrails and audit logger

B. Attack and Threat Scenarios

The governance system's resilience was evaluated through the simulation of various cybersecurity attack scenarios.

Simulated Threats

- Prompt Injection Attacks
- Hallucinated Financial Recommendations
- Unauthorized Insider Access
- Sensitive Data Extraction Attempts
- Compliance Policy Violations
- Adversarial Prompt Manipulation

Prompts, outputs, unauthorized access, and notifications for questionable behavior were all continuously assessed by the governance architecture.

C. Evaluation Metrics

The framework's effectiveness was assessed using quantitative cybersecurity and compliance metrics.

Performance Indicators

- Threat Detection Accuracy
- Data Leakage Prevention Rate
- Compliance Enforcement Accuracy
- Audit Logging Completeness
- Response Reliability
- False Positive Rate
- User Access Validation Accuracy

According to the assessment, operational security and rel iability were significantly increased by layered governance s ystems.

D. Experimental Results and Analysis

The proposed architecture proved to be very successful in reducing hallucinations and data leakage instances and ensuring very high detection rates in detecting attempts at illegal access as well as injection [23]. It was also able to protect the LLM engine from getting into any potentially hazardous impulses. It is possible to continuously verify policies with the help of compliance monitoring tools meant for ensuring compliance with PCI-DSS and GDPR regulations. Transparency and forensics become possible thanks to the fact that all AI interactions can be effectively recorded using audit recording technology.

TABLE 3: EXPERIMENTAL PERFORMANCE RESULTS

Evaluation Metric	Without Governance Framework	Proposed Governance Framework
Prompt Injection Detection	**62%**	**97%**
Data Leakage Prevention	**69%**	**98%**
Hallucination Reduction	**56%**	**92%**
Unauthorized Access Blocking	**71%**	**96%**
Compliance Accuracy	**74%**	**99%**
Audit Logging Coverage	**61%**	**100%**
Response Reliability	**67%**	**95%**

Figure 4: Simulation Workflow for Secure Banking LLM Governance

VII. CONCLUSION

The use of large language models in banking operations opens up possibilities for interaction with the users, security against fraud, automation of intelligent decisions, and compliance monitoring. However, these features pose cybersecurity threats and problems with respect to privacy and governance that cannot be resolved through conventional methods. Therefore, the aim of this study is to examine the importance of cybersecurity frameworks focusing on governance necessary for the safe use of large language models in the banking sector. The study revealed several critical vulnerabilities, including fast injections, hallucinations, security breaches, information leaks, and non-compliance that may pose a risk to banking operations and user interactions.

From the study, cybersecurity challenges can be overcome with the use of a hierarchical governance approach involving runtime guardrails, role-based access control, auditing and logging, compliance monitoring, encryption, and manual management. The findings from the experiment indicate that the proposed approach can significantly enhance threat detection efficiency, transparency in operation, regulation compliance, and consistency in the AI system's responses. Based on the results of the research, effective implementation of secure AI technology in finance requires proper governance mechanisms. Explanatory AI processes, autonomous compliance engines, blockchain-based audit processes, and adaptive governance structures can all be utilized in future research.

References:

[1] Fan, Minghong. "Llms in banking: Applications, challenges, and approaches." In *Proceedings of the International Conference on Digital Economy, Blockchain and Artificial Intelligence*, pp. 314-321. 2024.

[2] Sultana, Ghousia, Siraj Farheen Ansari, Mohammed Imran Ahmed, Abdul Faiyaz Shaik, Moin Uddin Khaja, and Bibhu Dash. "RESPONSIBLE AI ANALYTICS FOR REAL-WORLD IMPACT: NAVIGATING ETHICS, PRIVACY AND TRUST."

[3] Mohammed, Nasar, Abdul Faisal Mohammed, and Sruthi Balammagary. "Ransomware in healthcare: Reducing threats to patient care." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 2 (2025): 27-33.

[4] Mohammed, Akheel, Zubair Ahmed Mohammed, Naveed Uddin Mohammed, Shravan Kumar Gunda, Mohammed Azmath Ansari, and Mohd Abdul Raheem. "AI-NATIVE WIRELESS NETWORKS: TRANSFORMING CONNECTIVITY, EFFICIENCY, AND AUTONOMY FOR 5G/6G AND BEYOND

[5] Chittoju, S. R., and Siraj Farheen Ansari. "Blockchain's Evolution in Financial Services: Enhancing Security, Transparency, and Operational Efficiency." International Journal of Advanced Research in Computer and Communication Engineering 13, no. 12 (2024): 1-5.

[6] Reddy, Balavardhan, and Amir Ahmed Ansari. "AI-ENHANCED NETWORK TRAFFIC ANALYSIS FOR PREVENTING FRAUD PAYMENT IN BANKS."

[7] Ansari, Mohammed Azmath, Shaik Aqheel Pasha, and Narendar Kandula. "Reinforcement Learning for Adaptive Network Defense in Financial Institutions."

[8] Chittoju, Siva Sai Ram, Sireesha Kolla, Mubashir Ali Ahmed, and Abdul Raheman Mohammed. "Synergistic Integration of Blockchain and Artificial Intelligence for Robust IoT and Critical Infrastructure Security."

[9] Ansari, Meraj Farheen, Abubakar Mohammed, Kavitha Reddy Janamolla, Syed Waheeduddin Khadri, Shuaib Abdul Khader, and Mohd Abdul Raheem. "The Double-Edged Sword: Navigating AI's Role in Banking Cybersecurity." In 2025 International Conference on Transformative Computing Technologies (ICTCT), pp. 294-300. IEEE, 2025.

[10] Integrating Machine Learning and Engineering Management to Optimize Construction Scheduling and Resource Allocation

[11] Ansari, Meraj Farheen. "Redefining Cybersecurity: Strategic Integration of Artificial Intelligence for Proactive Threat Defense and Ethical Resilience."

[12] Mohammed, Abdul Faisal, and Mohammed Akifuddin Ghori. "AI-Enhanced Safety for Heavy Load Construction Vehicles: An Integrated Embedded C++ Software Approach."

[13] KASHIF, MOHAMMED, ABDUL RAHMAN JIBRAN SYED, and MUBASHIR ALI AHMED. "ADVANCING ANOMALY AND FRAUD DETECTION IN BIG DATA WITH ARTIFICIAL INTELLIGENCE."

[14] Janamolla, Kavitha, Ghousia Sultana Sultana, Fnu Mohammed Aasimuddin, Abdul Faisal Mohammed, and Fnu Shaik Aqheel Pasha Pasha. "Integrating Blockchain and AI for Efficient Trade Exception Handling: A Case Study in Cross-Border Settlements." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 1 (2025): 24-30.

[15] Mohammed, Nasar, Sireesha Kolla, Srujan Kumar Ganta, Shuaib Abdul Khader, and Sruthi Balammagary. "Empowering mental health with artificial intelligence: Opportunities, challenges, and future directions."

[16] Kashif, Mohammed, Mohammed Aasimuddin, Mubashir Ali Ahmed, Laxmi Bhavani Cheekatimalla, Eraj Farheen Ansari, and Ahwan Mishra. "AI-DRIVEN CTI FOR BUSINESS: EMERGING THREATS, ATTACK STRATEGIES, AND DEFENSIVE MEASURES."

[17] RAHEEM, MOHD ABDUL, and MOHAMMED AZMATH ANSARI. "INTELLIGENT AND TRUSTWORTHY 6G: AI-DRIVEN ARCHITECTURES, APPLICATIONS, AND SECURITY FRAMEWORKS.

[18] Khader, Shuaib Abdul, Amir Ahmed Ansari, and Syed Sharik Ali. "Zero-Day Exploit Prediction Using Graph-Based Deep Learning on Vulnerability and Threat Intelligence Data."

[19] Ansari, Meraj Farheen, and Syed Sharik Ali. "AI-driven zero-trust architecture for enhanced cybersecurity in dynamic network environments." International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering 13 (2025): 12.

[20] Gouni, Praveen Kumar Reddy, and Eraj Farheen Ansari. "The Impact of Cyber-Physical Attacks on AI-Enabled Business Systems

[21] Kashif, M., & Ansari, A. A. (2026). Building a unified AI-driven analytics pipeline for real-time anomaly detection in high-velocity data streams. IJIREEICE, 14(1), 66–75. https://doi.org/10.17148/ijireeice.2026.14111

[22] Reddy, B. (2021). A Quantitative Analysis of Cloud Security Practices in IoT Environment (dissertation).

[23] Mohammed, Naveed Uddin, Zubair Ahmed Mohammed, Shravan Kumar Reddy Gunda, Akheel Mohammed, and Moin Uddin Khaja.

"Networking with AI: Optimizing Network Planning, Management, and Security through the medium of Artificial Intelligence.