CyberInfo: A Comprehensive Review of AI-Driven Cyber Threat Intelligence, LLM Based Analysis, MITRE ATT&CK Mapping, and Incident Correlation Systems

Sanika Dhongade[1], Rajat Bhade[1], Ritesh Katare[1], Ashwin Dhanvijay[1], Rupesh Samarth[1], Saurabh Vyas[2]

[1]Student, Dept. of Computer Science and Engineering (Cyber Security), G H Raisoni College of Engineering and Management, Nagpur, India

[2]Asst. Professor, Dept. of Computer Science and Engineering (Cyber Security), G H Raisoni College of Engineering and Management, Nagpur, India

DOI: https://doi.org/10.47772/IJRISS.2026.101000222

Received: 14 October 2026; Accepted: 19 October 2026; Published: 01 November 2026

ABSTRACT

The increasing complexity of cyberattacks has created a growing demand for intelligent Cyber Threat Intelligence (CTI) systems capable of processing large volumes of security data and generating actionable insights. Traditional threat analysis methods rely heavily on manual investigation, resulting in delayed detection and response to emerging threats. Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) have significantly enhanced the automation of threat intelligence workflows. This review paper examines existing research on AI-driven cyber threat intelligence, automated threat extraction, MITRE ATT&CK mapping, anomaly detection, and incident correlation techniques. Various methodologies, architectures, and frameworks proposed by researchers are analyzed and compared to identify their strengths, limitations, and practical applications. The study further highlights existing research gaps, including limited integration of intelligence sources, lack of contextual threat correlation, and challenges in real-time threat analysis. Based on the reviewed literature, the paper discusses the development of CyberInfo, an AI-powered threat intelligence and incident correlation platform designed to improve threat visibility, reduce analyst workload, and support proactive cybersecurity operations.

Keywords: Cyber Threat Intelligence, Artificial Intelligence, Machine Learning, Natural Language Processing, Large Language Models, Retrieval-Augmented Generation, MITRE ATT&CK, Incident Correlation, Threat Intelligence Platform, Cybersecurity.

INTRODUCTION

The rapid growth of digital technologies, cloud computing, and interconnected information systems has significantly transformed modern organizations. While these advancements improve efficiency and accessibility, they also expose organizations to increasingly sophisticated cyber threats. Attackers continuously employ techniques such as ransomware, phishing campaigns, advanced persistent threats (APTs), and zero-day exploits to compromise critical systems and sensitive information. As a result, cybersecurity has become a major concern across both public and private sectors [7]–[10].

Cyber Threat Intelligence (CTI) has emerged as an essential component of modern cybersecurity strategies. CTI enables organizations to collect, analyze, and utilize threat-related information to understand adversarial activities and strengthen defensive capabilities. By transforming raw security data into actionable intelligence, CTI supports proactive threat detection, risk assessment, and incident response [2], [5], [6]. However, the growing volume of threat reports, security alerts, vulnerability disclosures, and open-source intelligence sources creates significant challenges for manual analysis.

To address these challenges, researchers have explored the use of Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), and Large Language Models (LLMs) to automate threat intelligence workflows [3], [13]–[15]. These technologies improve the efficiency of threat extraction, classification, anomaly detection, and cybersecurity decision-making. Furthermore, frameworks such as MITRE ATT&CK provide a standardized approach for understanding adversarial tactics, techniques, and procedures (TTPs), enabling security teams to better analyze and respond to cyber incidents [1], [4].

Recent developments in Retrieval-Augmented Generation (RAG) and incident correlation techniques have further enhanced the ability of cybersecurity platforms to provide context-aware intelligence [12], [23], [29]. By integrating information retrieval, knowledge representation, and AI-driven reasoning, these technologies enable more accurate threat analysis and reduce the burden on security analysts.

This review paper examines recent advancements in AI-driven Cyber Threat Intelligence systems, automated MITRE ATT&CK mapping approaches, NLP-based threat intelligence extraction techniques, Large Language Models, Retrieval-Augmented Generation frameworks, and incident correlation mechanisms. The paper analyzes existing research contributions, identifies current limitations and research gaps, and discusses the development of CyberInfo, an AI-powered threat intelligence and incident correlation platform designed to enhance threat visibility, improve operational efficiency, and support proactive cybersecurity operations [11], [12], [14].

Contribution of this paper. Existing surveys usually cover one aspect of the field, such as CTI frameworks, ATT&CK mapping, or LLM-based analysis. This paper differs in three ways: (i) it reviews CTI extraction, ATT&CK mapping, LLM/RAG analysis, and incident correlation together; (ii) it compares the capability coverage of existing system categories against a single integrated design; and (iii) it presents the CyberInfo prototype as a concrete case that addresses the gaps identified.

METHODOLOGY

This paper follows a structured literature review protocol. It is not a PRISMA-compliant systematic review, and the protocol is reported here so that the selection can be reproduced and its limits understood.

Sources and period. Research papers were collected mainly from IEEE Xplore, supplemented by conference proceedings (NeurIPS, NAACL-HLT) and arXiv for foundational language-model work. Standards, threat reports, and the ATT&CK knowledge base were taken from the publishing organizations (MITRE, NIST, CISA, ENISA, IBM, CrowdStrike, Verizon). Applied CTI studies cover 2018–2025, and foundational NLP/LLM works cover 2019–2023.

Search terms. Boolean combinations of the following were used: ("cyber threat intelligence" OR CTI) AND ("MITRE ATT&CK" OR "ATT&CK mapping" OR "natural language processing" OR NER OR "machine learning" OR "large language model" OR LLM OR "retrieval-augmented generation" OR RAG OR "knowledge graph" OR "incident correlation" OR "anomaly detection").

Inclusion criteria:

- Peer-reviewed or officially published work written in English.
- Addresses CTI extraction, ATT&CK mapping, AI/ML/NLP/LLM/RAG-based threat analysis, or incident correlation.
- Describes a method, architecture, framework, or dataset in enough detail to be compared.

Exclusion criteria:

- Vendor marketing material, blog posts, and non-reviewed web articles.
- Studies that mention CTI only in passing, or that do not describe their method.

- Duplicate or superseded versions of the same work.

Selection and quality assessment. Titles and abstracts were screened first, followed by full-text reading of the candidates. Each retained study was checked against five criteria: relevance to CTI, clarity of the method, venue credibility, recency, and whether its limitations or results were reported. The final corpus contains 33 sources: 21 CTI research papers [1]–[3], [16]–[33]; 4 foundational NLP/LLM works [11]–[14]; 1 reference book [15]; and 7 standards and industry reports [4]–[10]. The number of records initially retrieved was not logged.

Each study was then analyzed for its objective, technique, contribution, and limitation, and grouped into five domains: CTI frameworks, AI/ML analytics, NLP-based extraction, LLM/RAG systems, and incident correlation. The synthesis of these groups produced the comparison in Section 3 and the design of CyberInfo.

LITERATURE REVIEW AND COMPARATIVE ANALYSIS

Numerous studies have explored the application of AI, NLP, MITRE ATT&CK mapping, and Cyber Threat Intelligence to enhance cybersecurity operations. These works focus on threat intelligence extraction, attack detection, incident correlation, and automated security analysis. Table 1 presents a summary of the selected research papers and their key findings relevant to the proposed CyberInfo platform.

Table 1: Literature Review of Existing Research

Sr.No.	Paper Title & Author(s)	Publication Details	Findings	Ref.
1	Automated Cyber Threat Intelligence Mapping to the MITRE ATT&CK Framework – A. M. Alshamrani et al.	IEEE Access, 2025	Uses NLP techniques to automatically map threat intelligence reports to MITRE ATT&CK tactics and techniques, reducing manual analysis effort.	[1]
2	Taxonomy of Cyber Threat Intelligence Frameworks – M. Husák et al.	IEEE Access, 2022	Provides a comprehensive classification of CTI frameworks and standardizes threat intelligence processes for organizations.	[2]
3	AI-driven Data Analytics for Cyber Threat Intelligence and Anomaly Detection – S. Kumar et al.	IEEE Access, 2025	Applies AI and ML models for anomaly detection and threat prediction, improving cyberattack identification accuracy.	[3]
4	Automatic Tagging of Cyber Threat Intelligence Unstructured Data using Semantics Extraction – T. Wang and K. P. Chow	IEEE, 2019	Uses semantic extraction and NLP to identify indicators of compromise (IoCs) and threat entities from unstructured reports.	[16]
5	A Novel System Model for Managing Cyber Threat Intelligence	IEEE, 2022	Proposes a structured CTI management architecture that improves threat collection, storage, and dissemination.	[17]
6	A Study on a Cyber Threat Intelligence Analysis Platform for Proactive Detection of Cyber Attacks Based on Automated Analysis	IEEE, 2018	Demonstrates automated CTI analysis for early threat detection and proactive cyber defense mechanisms.	[18]
7	AI-Powered Cyber Threat Intelligence Analyst Platform for Automated Vulnerability Assessment and Threat Analysis	IEEE, 2025	Integrates AI-driven vulnerability assessment and threat analysis to support security analysts in decision-making.	[19]
8	LLM and Knowledge Graphs for Automated Cyber Threat Intelligence and Incident Response	IEEE, 2025	Combines LLMs and knowledge graphs to improve contextual threat understanding and automated incident response.	[20]

Table 2 compares existing Cyber Threat Intelligence (CTI) systems, MITRE ATT&CK mapping systems, AI analytics systems, and the proposed CyberInfo platform across thirteen capabilities identified in the reviewed literature. Support is rated as Full, Partial, or None for the existing categories, using the key in Table 3.

Table 2: Qualitative Comparison of Existing Approaches and the Proposed CyberInfo Platform

Feature	Existing CTI Systems	ATT&CK Mapping Systems	AI Analytics Systems	CyberInfo (proposed)
Threat Intelligence Processing	Full	Full	Full	Designed
NLP-Based Extraction	Partial	Partial	Partial	Designed
AI Integration	Partial	None	Full	Designed
MITRE ATT&CK Mapping	Partial	Full	None	Designed
Incident Correlation	None	None	Partial	Designed
LLM Integration	None	None	Partial	Designed
RAG-Based Retrieval	None	None	None	Designed
Risk Scoring	Partial	None	Full	Designed
Vulnerability Analysis	Partial	None	Full	Designed
Historical Incident Linking	None	None	Partial	Designed
Context-Aware Threat Analysis	Partial	Partial	Full	Designed
Automated Intelligence Generation	None	None	Partial	Designed
Unified Platform Capability	None	None	Partial	Designed

Table 3: Rating Key for Table 2

Rating	Meaning
Full	Core, well-documented capability of the category in the reviewed studies
Partial	Supported only by some studies, or in limited form
None	Not reported in the reviewed studies
Designed	Part of the planned CyberInfo design; not yet empirically validated

Basis of the ratings. The ratings in Table 2 are qualitative judgements by the authors, derived from how fully each capability is described in the studies of Table 1 and the related literature [1]–[3], [16]–[33]. They are not benchmark results and do not measure performance. The "Designed" entries describe the intended scope of CyberInfo, not demonstrated capability. The table therefore shows design coverage only, and the comparison should be read as indicative until CyberInfo is empirically evaluated.

Existing CTI systems demonstrate strong capabilities in threat intelligence processing but provide limited support for advanced functionalities such as incident correlation, contextual analysis, and automated intelligence generation. MITRE ATT&CK mapping solutions excel in adversary tactic and technique mapping; however, they generally lack AI-driven reasoning and incident correlation capabilities. AI Analytics Systems offer strong performance in anomaly detection, risk assessment, and predictive analysis but often operate independently without standardized threat intelligence integration.

In contrast, CyberInfo is designed to cover all evaluated parameters through a unified architecture that combines Natural Language Processing (NLP), Artificial Intelligence (AI), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), MITRE ATT&CK mapping, vulnerability assessment, and incident correlation. This integrated approach is intended to support comprehensive threat analysis, automated intelligence generation, and enhanced cybersecurity decision support.

Findings from the comparative analysis:

- Existing solutions primarily focus on individual cybersecurity functions rather than end-to-end threat intelligence workflows.

- ATT&CK mapping systems provide excellent threat classification but limited contextual threat analysis.
- AI analytics systems improve detection and prediction capabilities but lack comprehensive threat intelligence correlation.
- CyberInfo integrates multiple technologies into a single platform, enabling automated threat extraction, ATT&CK mapping, incident correlation, and intelligent threat analysis.
- The analysis highlights the need for unified cybersecurity platforms capable of combining AI-driven intelligence with contextual threat correlation and automated decision support.

Technologies and Techniques Used In Existing Research

The reviewed literature employs various technologies and analytical techniques to improve Cyber Threat Intelligence processes. Natural Language Processing (NLP) is widely used for threat extraction from unstructured reports [16], [25], while Artificial Intelligence and Machine Learning support anomaly detection and threat classification [3], [33]. Recent studies have incorporated Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and MITRE ATT&CK mapping to enhance contextual understanding and automated threat analysis [1], [20], [29]. Table 4 summarizes the major technologies identified in the reviewed research.

Table 4: Major Technologies Identified in the Reviewed Research

Technology / Tool	Purpose in CTI Research	Ref.
NLP	Threat extraction from unstructured reports	[16], [25]
Named Entity Recognition (NER)	Identification of IoCs, malware, IPs, vulnerabilities	[25]
Machine Learning	Anomaly detection and attack prediction	[3], [15], [33]
MITRE ATT&CK	Standardized threat mapping	[1], [4], [26], [27]
LLMs	Threat analysis and report summarization	[11], [14]
RAG	Context-aware intelligence generation	[12], [29], [30]
Knowledge Graphs	Relationship discovery and incident correlation	[20], [23]

Existing Work

Recent advancements in Cyber Threat Intelligence (CTI) have focused on improving the collection, processing, and utilization of threat information through automation and artificial intelligence. Early CTI frameworks primarily emphasized structured threat sharing, intelligence lifecycle management, and knowledge organization [5], [17], [24]. Taxonomy-based studies established foundational classifications of threat intelligence, enabling organizations to understand strategic, operational, tactical, and technical intelligence more effectively [2].

Subsequent research introduced Natural Language Processing (NLP) techniques to address the challenge of extracting actionable intelligence from unstructured cybersecurity reports. Semantic extraction and automated tagging approaches demonstrated the ability to identify threat indicators, malware families, vulnerabilities, and attack-related entities from textual sources with reduced manual effort [16], [25]. Transformer-based language models such as BERT [13] have further advanced such text-understanding tasks.

Researchers further explored the application of Artificial Intelligence and Machine Learning for anomaly detection, attack classification, and predictive threat analysis. These approaches improved the identification of suspicious activities by learning behavioural patterns from large-scale security datasets [3], [15], [28], [33]. Automated ATT&CK mapping frameworks also emerged, enabling threat information to be systematically aligned with MITRE ATT&CK tactics and techniques for improved threat understanding [1], [26], [27].

More recently, Large Language Models (LLMs) [11], [14], Retrieval-Augmented Generation (RAG) [12], [29], [30], and knowledge graph technologies [20], [23] have been investigated to enhance contextual reasoning and cybersecurity decision support. Related work has also explored automated threat advisory and intelligence report generation [22], [31], explainable AI for real-time analysis [21], and the integration of AIOps with threat intelligence databases [32]. These solutions have shown promise in automating intelligence analysis and supporting incident response. However, most existing studies address specific cybersecurity functions independently, resulting in fragmented workflows and limited end-to-end intelligence correlation.

Innovation

The primary innovation identified through the reviewed literature is the gradual evolution from isolated cybersecurity analysis tools toward intelligent and context-aware threat intelligence systems. While existing solutions successfully address individual challenges such as anomaly detection, threat extraction, ATT&CK mapping, or automated intelligence generation, limited research focuses on integrating these capabilities into a unified analytical environment.

CyberInfo extends the concepts presented in previous studies by combining multiple intelligence functions within a single platform. The proposed approach integrates Natural Language Processing, Artificial Intelligence, Large Language Models, Retrieval-Augmented Generation, MITRE ATT&CK mapping, and incident correlation to create a comprehensive threat intelligence workflow.

Another innovative aspect is the incorporation of contextual incident correlation, which establishes relationships between current threat observations and historical cybersecurity events. This capability is expected to improve situational awareness and help analysts understand attack progression.

By combining automated intelligence extraction, knowledge-assisted reasoning, ATT&CK-based threat representation, and correlation-driven analysis, the proposed platform is intended to provide more actionable intelligence, with the expected benefits of lower analyst workload and more proactive operations. These benefits are design goals and have not yet been measured. Figures 1–4 show the main screens of the CyberInfo prototype.

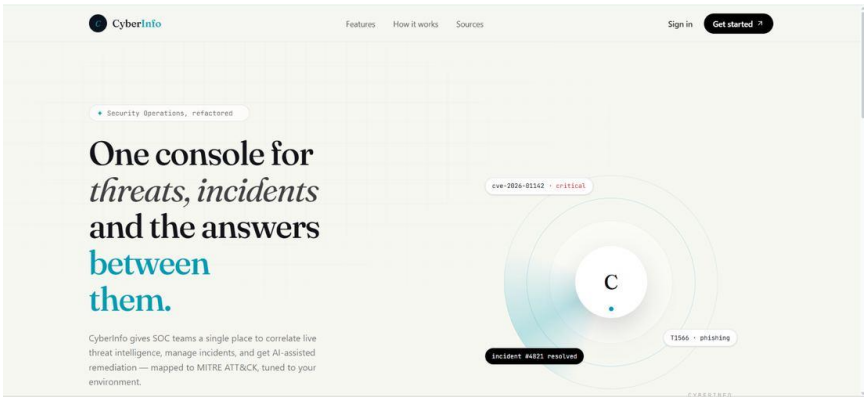


Figure 1: CyberInfo Landing Page

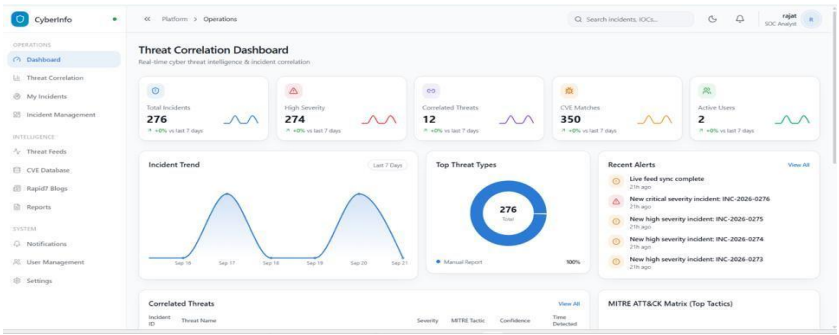


Figure 2: CyberInfo Dashboard

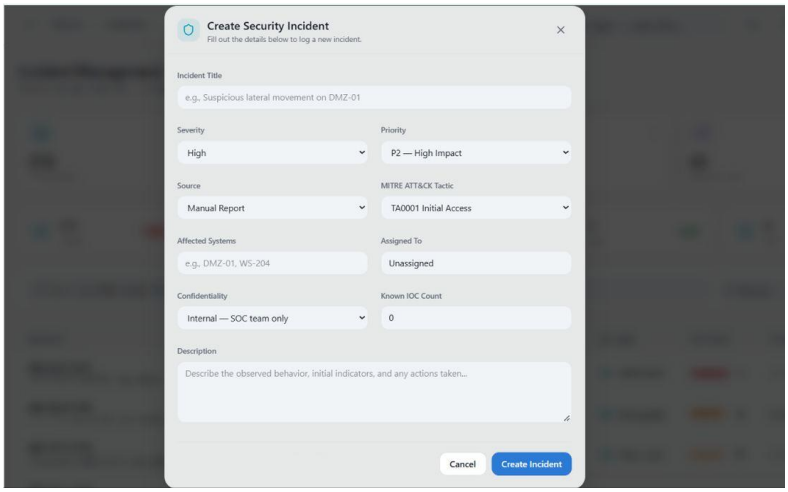


Figure 3: CyberInfo Create Incident Page

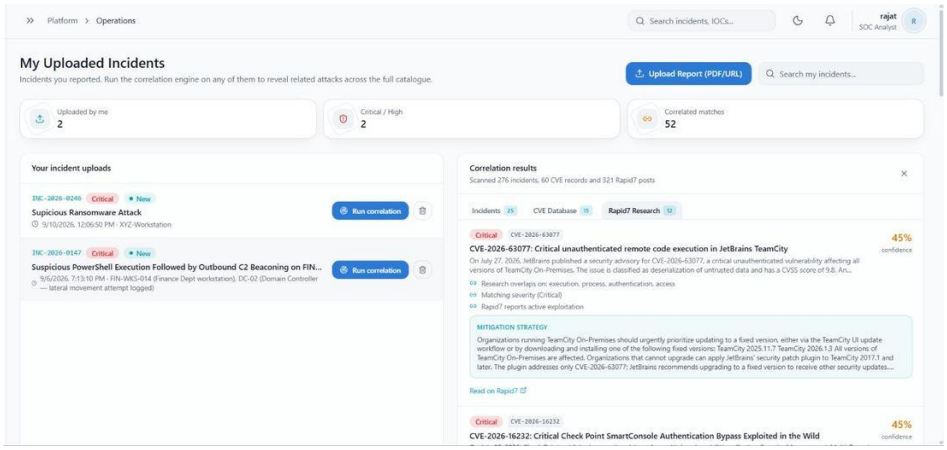


Figure 4: CyberInfo Incident Correlation Page

Limitations. This work has three main limitations. First, the review is a structured, not fully systematic, review, and the record counts at the search stage were not logged. Second, the comparison in Table 2 is qualitative. Third, CyberInfo has not yet been evaluated empirically, so claims about accuracy, efficiency, scalability, and operational benefit remain prospective.

Future Scope

Cyber Threat Intelligence systems must continuously evolve to address increasingly sophisticated cyber threats. Future enhancements to CyberInfo may include the integration of real-time threat intelligence feeds to improve the detection of emerging attacks and provide up-to-date security insights.

Advanced Machine Learning and Deep Learning techniques can be incorporated to support predictive threat analysis by identifying potential attack patterns before they cause significant damage. The platform may also be enhanced through graph-based analytics and knowledge graph technologies to improve incident correlation and contextual threat understanding.

Another important direction is the integration of additional cybersecurity standards and frameworks such as STIX, TAXII, CVE, and CVSS to enable better threat representation and information sharing. Furthermore, future advancements in Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) can improve automated report generation, contextual reasoning, and cybersecurity decision support.

Overall, these enhancements can strengthen CyberInfo's ability to provide intelligent, scalable, and proactive cyber defense capabilities while improving operational efficiency and threat visibility.

CONCLUSION

Cyber Threat Intelligence (CTI) plays an important role in identifying emerging cyber threats and improving incident response. This review examined existing research on AI-driven threat analysis, Natural Language Processing, automated MITRE ATT&CK mapping, Large Language Models, Retrieval-Augmented Generation, and incident correlation. The literature indicates that these technologies can reduce manual effort and improve the extraction and interpretation of threat intelligence. However, challenges remain in integrating diverse intelligence sources, maintaining contextual accuracy, and correlating current threats with historical incidents.

To address these limitations, CyberInfo is proposed as an integrated threat intelligence platform that combines automated intelligence extraction, contextual analysis, ATT&CK mapping, and incident correlation. By bringing these capabilities together, the platform aims to support faster threat investigation, improve security visibility, and assist analysts in making informed decisions. Planned evaluation includes testing on representative CTI reports, measuring ATT&CK mapping precision and recall, assessing incident-correlation cases, and running a user study with security analysts. Until this is completed, the benefits of CyberInfo should be regarded as design goals rather than established results.

REFERENCES

1. A. M. Alshamrani, A. Myneni, A. Chowdhary, and D. Huang, "Automated Cyber Threat Intelligence Mapping to the MITRE ATT&CK Framework," IEEE Access, 2025.
2. M. Husák, J. Komárková, E. Bou-Harb, and P. Čeleda, "Taxonomy of Cyber Threat Intelligence Frameworks," IEEE Access, vol. 10, 2022.
3. S. Kumar, R. Sharma, and A. Gupta, "AI-driven Data Analytics for Cyber Threat Intelligence and Anomaly Detection," IEEE Access, 2025.
4. MITRE Corporation, "MITRE ATT&CK Framework." [Online]. Available: https://attack.mitre.org
5. National Institute of Standards and Technology (NIST), "The NIST Cybersecurity Framework (CSF) 2.0," 2024.
6. Cybersecurity and Infrastructure Security Agency (CISA), "Threat Intelligence Resources," 2024.
7. European Union Agency for Cybersecurity (ENISA), "ENISA Threat Landscape 2024," 2024.
8. IBM Security, "Cost of a Data Breach Report 2024," IBM Corporation, 2024.
9. CrowdStrike, "2024 Global Threat Report," CrowdStrike Inc., 2024.
10. Verizon, "2024 Data Breach Investigations Report (DBIR)," Verizon Enterprise, 2024.
11. OpenAI, "GPT-4 Technical Report," arXiv:2303.08774, 2023.
12. P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems (NeurIPS), 2020.
13. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.
14. T. Brown et al., "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems (NeurIPS), 2020.
15. C. C. Aggarwal, Machine Learning for Cybersecurity. Springer, 2022.
16. T. Wang and K. P. Chow, "Automatic Tagging of Cyber Threat Intelligence Unstructured Data using Semantics Extraction," in Proc. IEEE Int. Conf. Intelligence and Security Informatics (ISI), 2019, pp. 197–199, doi: 10.1109/ISI.2019.8823252.
17. "A Novel System Model for Managing Cyber Threat Intelligence," IEEE, 2022.
18. "A Study on a Cyber Threat Intelligence Analysis (CTI) Platform for the Proactive Detection of Cyber Attacks Based on Automated Analysis," IEEE, 2018.
19. "AI-Powered Cyber Threat Intelligence Analyst Platform for Automated Vulnerability Assessment and Threat Analysis," IEEE, 2025.
20. "LLM and Knowledge Graphs for Automated Cyber Threat Intelligence and Incident Response," IEEE, 2025.
21. "Explainable AI-Based Cyber Threat Intelligence for Real-Time Security Analysis," IEEE, 2025.
22. "Automated Cyber Threat Intelligence Generation on Multi-Host Network Incidents," IEEE, 2024.

23. "Research on Network Security Event Correlation Analysis Method Based on Knowledge Graph," IEEE, 2024.
24. "Design of a Cyber Threat Information Collection System for Cyber Attack Correlation," IEEE, 2018.
25. "Research on Named Entity Recognition Models for Cybersecurity," IEEE, 2024.
26. "Mapping Zero-Click Attack Behavior into MITRE ATT&CK Mobile: A Systematic Process," IEEE, 2023.
27. "Aligning Security Operations with the MITRE ATT&CK Framework: Level up your Security Operations Center for Better Security," IEEE, 2023.
28. "GraphWatch: A Novel Threat Hunting Approach for APT Activities Based on Anomaly Detection," IEEE, 2024.
29. "RAGSec: Retrieval-Augmented Generation for Cybersecurity Threat Intelligence in Enterprise Networks," IEEE, 2025.
30. "Integrating Knowledge Graphs with Retrieval-Augmented Generation to Automate IoT Device Security Compliance," IEEE, 2025.
31. "Automating Threat Advisory Report Generation Using Threat Intelligence," IEEE, 2024.
32. "Network Connection Management Based on the Integration of Artificial Intelligence for IT Operations and Threat Intelligence Databases," IEEE, 2025.
33. "Evolving ML-Based Intrusion Detection: Cyber Threat Intelligence for Dynamic Model Updates," IEEE, 2025.