International Journal of
Science, Architecture, Technology and Environment

A Comprehensive Review on Artificial Intelligence for Cyber Threat Intelligence and Prediction

Waliu Adebayo Ayuba[1*]

[1]*Northeastern Universityy, Boston, MA, USA*

[*]*Corrusponding author*

DOI: https://doi.org/10.63680/ijsate0124001.06

Abstract

The increasing sophistication of cyber threats has intensified the need for intelligent, predictive defense mechanisms. Cyber Threat Intelligence (CTI) leverages data driven insights to understand and anticipate malicious activities, yet traditional CTI methods struggle with scalability, adaptability, and timeliness. This review explores how Artificial Intelligence (AI) is reshaping CTI through automation, anomaly detection, and predictive analytics. It presents a comprehensive analysis of AI techniques including machine learning, deep learning, and graph based models applied to threat detection, malware analysis, and attack prediction. The paper further discusses available datasets, tools, and frameworks such as MISP, STIX/TAXII, and MITRE ATT&CK, along with their integration into AI driven pipelines. Key challenges identified include data scarcity, lack of explainability, adversarial vulnerabilities, and limited interoperability. The review concludes that future research should prioritize multimodal learning, explainable AI, federated intelligence sharing, and human AI collaboration to develop transparent, adaptive, and ethically grounded CTI systems capable of predicting and mitigating evolving cyber threats.

Keywords: Cyber Threat Intelligence (CTI); Artificial Intelligence (AI); Machine Learning; Deep Learning; Threat Prediction; Knowledge Graphs; Explainable AI; Federated Learning.

1.0 Introduction

In today's digitally interconnected world, cyber threats are intensifying in both frequency and sophistication. Attackers exploit vulnerabilities in networks, applications, and human behavior to launch advanced persistent threats (APTs), ransomware campaigns, supply chain attacks, and zero day exploits. Traditional security tools such as signature based detection, rule engines, and manual analysis are increasingly insufficient to keep pace with such dynamic, evolving threats. As a consequence, organizations are turning to Cyber Threat Intelligence (CTI) to gain proactive insight into adversarial behavior and anticipate emerging risks. CTI refers to the collection, analysis, and dissemination of information about threats, threat actors, their tactics, techniques, and procedures (TTPs), and the indicators of compromise (IoCs) associated with malicious activities. CTI supports decision making across security operations, incident response, and strategic planning by

contextualizing threats rather than treating them as isolated alerts. However, the conventional CTI pipeline is constrained by challenges such as massive data volumes, heterogeneity of intelligence sources, a gap between raw data and actionable insights, and human limited throughput (Frias et al., 2025; Santos et al., 2025).

To address these challenges, artificial intelligence (AI) methods especially machine learning (ML) and deep learning (DL) are being adopted to augment or automate portions of the CTI process. AI enables systems to analyze large datasets, detect subtle patterns, correlate intelligence across multiple modalities, and continuously adapt as the threat landscape shifts. For example, NLP models can extract threat indicators from unstructured reports, while graph based models can map relationships among malware, adversaries, and infrastructure (Shah & Khoda Parast, 2024). The shift toward AI driven CTI methods promises to improve the speed, scalability, and precision of intelligence generation compared to traditional techniques. A particularly crucial aspect of AI in CTI is prediction. While many AI models focus on detection of known or ongoing threats, predictive CTI aims to forecast future threat events or behaviors, such as emerging IoCs, attack campaigns, or actor movements. Predictive intelligence empowers security teams to act before incidents fully surface, enabling proactive defense, better prioritization of resources, and more strategic decision making. As adversaries employ adaptive and stealthy tactics, prediction becomes a differentiator in resilient defense (Jiang et al., 2024).

Given the rising interest in AI based CTI, it is important to clarify certain foundational terminology:

Cyber Threat Intelligence (CTI): Knowledge derived from analysis of cyber threat data, including indicators, tactics, and context, used to inform security decisions.

Artificial Intelligence (AI): The development of computational systems capable of performing tasks traditionally requiring human intelligence, such as learning, reasoning, and perception.

Machine Learning (ML): A subset of AI where models improve performance on a task by learning from data (supervised, unsupervised, or reinforcement learning).

Deep Learning (DL): A further specialization within ML, characterized by using deep neural networks (e.g., convolutional, recurrent, transformer architectures) to learn hierarchical representations.

Indicator of Compromise (IoC): A forensic artifact such as an IP address, file hash, or domain that signals a possible malicious activity.

Tactics, Techniques, and Procedures (TTPs): The behavioral methods, tools, and strategies used by threat actors in executing attacks.

This review paper seeks to synthesize the current state of research into AI driven CTI and threat prediction. Our objectives are as follows:

1. To catalog and compare existing AI, ML, and DL methods applied in CTI tasks (e.g., detection, inference, prediction).

2. To analyze public datasets, tools, and benchmarks used in AI–CTI research, assessing their strengths, weaknesses, and suitability.

3. To identify gaps and challenges in current works (such as data scarcity, adversarial robustness, interpretability, real time adaptation, and human trust).

4. To highlight emerging trends (multimodal data fusion, knowledge graphs, large language models, human AI collaboration) that can advance predictive CTI capabilities.

5. To offer recommendations and a roadmap for future research in AI based CTI and prediction.

The remainder of this paper is organized as follows. Section 2 describes the review methodology and criteria. Section 3 presents CTI fundamentals and data sources. Section 4 surveys AI and ML techniques used in CTI. Section 5 addresses prediction models and advanced methods such as knowledge graphs and LLMs. Section 6 examines datasets and tools. Section 7 discusses challenges and open issues. Section 8 outlines promising future directions, and Section 9 concludes.

2.0 Overview of Cyber Threat Intelligence (CTI)

Cyber Threat Intelligence (CTI) represents a systematic process of collecting, analyzing, and sharing information about cyber threats to help organizations understand, anticipate, and mitigate potential attacks. It transforms raw data from diverse sources into actionable insights that guide security decisions, enhance situational awareness, and strengthen overall defense postures (Mavroeidis & Bromander, 2021; ENISA, 2020). Unlike traditional reactive security measures, CTI focuses on proactive defense helping analysts answer the critical questions of *who* is attacking, *why* they are attacking, and *how* they operate (Frias et al., 2025). CTI can be divided into three main categories tactical, operational, and strategic each serving different purposes and audiences. Tactical CTI is the most technical level, providing short term, machine readable data such as indicators of compromise (IoCs), malicious domains, IP addresses, file hashes, and malware signatures. This form of intelligence is used directly by intrusion detection systems, firewalls, and endpoint protection tools for immediate threat detection and blocking (Palo Alto Networks, n.d.). Operational CTI offers contextual insights into adversarial campaigns, including their tactics, techniques, and procedures (TTPs). It helps threat hunters and incident responders understand *how* an attack unfolds and *where* it may originate (Zvelo, n.d.; Expert Insights, 2025). Strategic CTI, in contrast, provides a broader, long term perspective for decision makers. It includes analyses of threat trends, motivations, and geopolitical implications that inform risk management and cybersecurity investment decisions (CrowdStrike, n.d.).

The effectiveness of CTI depends heavily on the data sources used to generate intelligence. Internal telemetry such as network logs, intrusion detection system (IDS) alerts, endpoint monitoring data, and incident reports are foundational for any CTI process. Organizations also deploy honeypots and deception systems to attract and study attackers in controlled environments (Hossen et al., 2021). In addition to internal sources, open source intelligence (OSINT) including vulnerability databases, security blogs, and research publications provides valuable external insights. Furthermore, social media and dark web platforms serve as early indicators of emerging threats, as threat actors frequently discuss or trade exploits, credentials, and malware on underground forums (Nunes et al., 2016). Commercial intelligence feeds, industry sharing communities such as ISACs, and standardized frameworks like STIX/TAXII are also integral in enriching CTI with broader situational awareness (ENISA, 2020).

To manage these data sources effectively, CTI operations follow a structured intelligence lifecycle, adapted from traditional intelligence practices. This lifecycle generally includes six stages: planning and direction,

collection, processing, analysis, dissemination, and feedback (Kraven Security, 2025). During the planning phase, analysts define intelligence requirements and priorities. The collection phase gathers raw data from internal and external sources, which is then processed to normalize, clean, and enrich the data for analysis. The analysis phase transforms data into actionable intelligence through correlation, classification, and interpretation. The results are then disseminated to relevant stakeholders ranging from technical teams to executive management depending on the intelligence level. Finally, the feedback phase ensures continuous improvement by refining requirements and methods based on user evaluations and changing threat conditions (CSO Online, 2023). While this lifecycle provides structure, traditional CTI approaches face numerous challenges. One of the main limitations is the overwhelming volume and velocity of threat data, which makes manual analysis infeasible in real time environments. The heterogeneity and unstructured nature of CTI data ranging from network logs to social media text further complicate integration and analysis. Additionally, timeliness and accuracy are persistent issues, as many indicators of compromise quickly lose relevance when adversaries change infrastructure or techniques (Mavroeidis & Bromander, 2021). CTI feeds also suffer from data quality problems, such as duplication, noise, and false positives, which can overwhelm security teams with irrelevant alerts.

Another key limitation lies in the lack of contextual understanding; traditional CTI often focuses on isolated IoCs without connecting them to broader threat campaigns or attacker motivations. Moreover, intelligence is often siloed across tactical, operational, and strategic levels, limiting cross layer feedback and collaboration (ENISA, 2020). Finally, existing CTI systems rely heavily on manual correlation and rule based logic, offering limited automation, adaptability, and scalability (Kraven Security, 2025). As a result, many organizations are now turning toward AI driven approaches to overcome these barriers, enabling automated data fusion, real time analysis, and predictive modeling for more proactive and adaptive threat intelligence.

3.0 Role of AI in CTI

In recent years, the integration of artificial intelligence (AI) into Cyber Threat Intelligence (CTI) has emerged as a transformative trend, driven by the need to overcome the limitations of traditional CTI methods. AI is well suited for CTI because it enables automation of repetitive tasks (such as parsing logs, flagging anomalies, and correlating indicators), supports anomaly detection by identifying deviations from normal behavior, and empowers predictive analytics to forecast future threat events or evolving attack paths (Lampis & Dekker, 2024; Nachaat Mohamed, 2025). Various types of AI methods have been applied to CTI tasks. Traditional machine learning (ML) algorithms such as Support Vector Machines (SVM), Random Forests, and Naïve Bayes are commonly used for classification, clustering, and anomaly detection in cybersecurity data. These methods often require careful feature engineering and preprocessing but can perform reliably in structured domains (Lampis & Dekker, 2024; "The Role of Machine Learning in Threat Intelligence," 2025).

Complementing ML, deep learning (DL) techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short Term Memory networks (LSTMs), and Transformer architectures allow end to end learning of features from raw or minimally processed data. Deep learning has shown strong performance in detecting complex, subtle patterns in network traffic, user behavior, and threat intelligence text streams (e.g. IDS, malware classification) (Mohamed, 2025; "Deep Learning vs. Machine Learning for Intrusion Detection," 2025). Its ability to automatically extract hierarchical feature representations reduces the need for manual feature design and enables better generalization to evolving threats. In addition, reinforcement learning (RL) has been explored in select CTI contexts, particularly for adaptive defense or adversary simulation. RL agents can learn strategies over time to optimize detection thresholds, allocate

monitoring resources, or simulate attacker behavior in a dynamic environment. While less mature than ML/DL applications, RL offers potential for self adaptive threat intelligence systems.

When comparing ML and DL in CTI, each has distinct advantages and tradeoffs. ML methods are often more interpretable, computationally lighter, and easier to deploy in constrained environments. They perform well when features are well understood and datasets are of moderate size. Deep learning, on the other hand, can model highly complex and non linear relationships, learn from high dimensional inputs (e.g. raw packet features, text embeddings), and scale better with large volumes of data. However, DL models often require more computational resources, are more prone to overfitting or adversarial manipulation, and are less transparent, which can hinder analyst trust (Mohamed, 2025; "Deep Learning vs. Machine Learning for Intrusion Detection," 2025). AI is redefining CTI by bringing automation, scalability, and predictive power into intelligence workflows. The challenge going forward is to harness these capabilities while maintaining transparency, robustness, and alignment with human analysts.

4.0 Review of AI Techniques for Threat Prediction

Artificial Intelligence (AI) has revolutionized threat prediction by automating the detection and forecasting of cyberattacks across diverse data sources. This section reviews major AI approaches and models applied in Cyber Threat Intelligence (CTI), organized by their primary use cases.

a. Threat Detection Systems

AI based threat detection systems automate the identification of malicious events in real time, replacing rule based systems that struggle with emerging threats. Traditional machine learning algorithms, such as Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosting, have been extensively used for anomaly detection and classification of security incidents. For instance, Maseer et al. (2021) evaluated RF, SVM, CNN, and ANN models on the CICIDS2017 dataset and found that Random Forest achieved superior detection rates while maintaining interpretability. However, they noted that the model's performance was sensitive to preprocessing and class imbalance. Similarly, Darweesh et al. (2024) applied Random Forest within a Network Intrusion Detection System (NIDS) and demonstrated improved accuracy compared to signature based approaches, though feature engineering and generalization to unseen attacks remained challenging.

b. Intrusion and Malware Detection

Deep learning (DL) methods have emerged as powerful tools for detecting malware and intrusions, as they automatically learn hierarchical patterns from data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), particularly Long Short Term Memory (LSTM) networks, have been applied to malware classification and behavior prediction tasks. A study by Singh et al. (2024) demonstrated that CNN–LSTM hybrid models achieved over 98% accuracy on malware datasets, outperforming traditional ML baselines. Similarly, Khan et al. (2024) reported that deep architectures were more effective at identifying obfuscated malware, although they required larger datasets and were more vulnerable to adversarial attacks. These results underscore the strength of DL in complex environments, despite its computational cost and explainability challenges.

c. Network Traffic Analysis

Network traffic analysis leverages AI to classify attack types, detect anomalies, and predict potential intrusions in large scale network environments. Li et al. (2024) conducted a comparative study using various ML and DL algorithms on the CICIDS2017 dataset and concluded that deep learning models provided higher accuracy and robustness for complex attacks such as DDoS and infiltration. Autoencoders and CNNs have proven particularly useful for anomaly detection in flow based traffic, but challenges persist with class imbalance and the need for rapid inference (Frias et al., 2025).

d. Threat Prediction Models

Beyond detection, predictive models aim to forecast future attacks or attacker behaviors. AI driven prediction models combine time series forecasting with knowledge graph analytics to identify emerging threats before they materialize. Lampis and Dekker (2024) proposed an AI enhanced CTI pipeline that integrates automated data collection with predictive modeling, achieving promising results in short term cyber threat forecasting. However, the accuracy of predictive CTI models still depends heavily on the freshness, quality, and diversity of training data (Mohamed, 2025).

e. Hybrid or Ensemble AI Models

Hybrid or ensemble models combine multiple AI algorithms to capitalize on their complementary strengths. For example, CNN–LSTM hybrids merge spatial and temporal pattern recognition to improve malware classification performance. A 2024 study by Singh et al. reported that an ensemble of Random Forest and CNN–LSTM models increased detection accuracy by 3–5% compared to single model systems. Despite these advantages, hybrid models are often more computationally intensive and less interpretable (Balasubramanian et al., 2025).

f. AI with Graph Based CTI (Knowledge Graphs)

Graph based AI techniques have become essential in CTI due to their ability to represent complex relationships among threat actors, campaigns, vulnerabilities, and indicators of compromise (IoCs). Frameworks such as TINKER, Open CyKG, and AttacKG use knowledge graphs (KGs) and Graph Neural Networks (GNNs) to enhance reasoning and link prediction in CTI. Rastogi et al. (2021) introduced TINKER, which models CTI entities and relationships to support automated reasoning. Similarly, Zhang et al. (2023) and Chen et al. (2024) demonstrated that combining KGs with transformer based large language models (LLMs) significantly improved contextual understanding and the prediction of emerging attack patterns. However, challenges such as schema alignment, data noise, and the need for human validation persist (Frias et al., 2025).

Table 1. Summary of Representative AI Approaches for Threat Prediction

Paper	Year	Technique	Dataset	Accuracy / Result	Limitation
Maseer et al.	2021	Random Forest, SVM, CNN, ANN	CICIDS2017	RF achieved 96% detection rate	Sensitive to preprocessing and imbalance
Darweesh et al.	2024	Random Forest (NIDS)	CICIDS2017	Improved accuracy vs. signature detection	Limited generalization to unseen attacks
Singh et al.	2024	CNN–LSTM hybrid	Malware dataset	>98% accuracy	Requires large labeled datasets
Khan et al.	2024	Deep learning models	IoT malware dataset	High detection precision (F1 > 0.95)	Vulnerable to adversarial samples
Lampis & Dekker	2024	AI based predictive pipeline	Multi source CTI data	Successful short term forecasting	Data dependency, noise sensitivity
Rastogi et al.	2021	Graph based TINKER framework	CTI reports	Enhanced reasoning and correlation	Requires expert validation
Zhang et al.	2023	KG + Transformer model	CTI repositories	Improved context inference	Schema alignment challenges

5.0 Discussion

From the reviewed literature, it is evident that Artificial Intelligence (AI) has become a transformative force in Cyber Threat Intelligence (CTI), enabling automation, adaptability, and predictive capabilities that far surpass traditional analytical methods. The convergence of machine learning (ML), deep learning (DL), and graph based reasoning has significantly enhanced the capacity of CTI systems to process vast, heterogeneous data sources and identify evolving threats in near real time (Frias, Pereira, & Silva, 2025; Lampis & Dekker, 2024). Machine learning methods remain foundational due to their interpretability, efficiency, and suitability for structured datasets such as network traffic logs, packet features, and system event records. Algorithms like Random Forest, SVM, and Gradient Boosting have consistently achieved high classification accuracy on benchmark intrusion datasets such as CICIDS2017 and UNSW NB15 (Maseer et al., 2021; Darweesh, Khan, & Lin, 2024). These approaches perform well when the features are engineered effectively and when labeled data are abundant. However, they often falter when faced with zero day attacks, imbalanced data distributions, or dynamically evolving adversarial behaviors, since such models typically rely on static patterns and fixed thresholds (Singh, Sharma, & Kumar, 2024).

Deep learning architectures, on the other hand, demonstrate remarkable proficiency in learning complex, non linear patterns directly from raw or unstructured inputs. CNNs and LSTMs have shown superior performance in malware and intrusion detection tasks, while Transformer based architectures have expanded AI's capability to process contextual relationships in textual CTI reports and log sequences (Khan, Javed, & Alghamdi, 2024). Despite these strengths, DL approaches introduce new challenges they are computationally intensive, data hungry, and often opaque in decision making, making them less transparent to human analysts

and less practical for low resource environments (Balasubramanian et al., 2025). The growing field of graph based and multimodal AI represents a major leap in the sophistication of CTI systems. Knowledge graph frameworks such as TINKER and Open CyKG (Rastogi et al., 2021) enable structured representation of entities and relationships (e.g., threat actors, tools, campaigns, vulnerabilities) and support reasoning across disparate intelligence feeds. When combined with graph neural networks (GNNs) and large language models (LLMs), these systems can infer hidden relationships, predict likely next attack steps, and enhance the contextual relevance of intelligence outputs (Zhang, Liu, & Xu, 2023). This fusion of symbolic and sub symbolic AI where knowledge graphs capture semantics and LLMs handle linguistic nuance marks a promising trajectory for next generation CTI research. Hybrid and ensemble AI models have further improved robustness and detection precision by leveraging the complementary strengths of different algorithms. For example, CNN–LSTM ensembles and Random Forest + DL hybrids yield stronger generalization and resilience to data variability (Singh et al., 2024; Lampis & Dekker, 2024). However, their increased architectural complexity often comes at the expense of transparency and computational efficiency.

A recurring limitation across all reviewed approaches is generalization. Many AI models demonstrate high accuracy on public datasets but degrade sharply when deployed in production environments, where attack distributions differ from training data. Moreover, the lack of standardized, up to date, and labeled datasets constrains reproducibility and fair benchmarking in AI based CTI research (Frias et al., 2025). Another critical barrier is explainability. In operational cybersecurity contexts, analysts must understand *why* a model raised an alert to validate its reliability and integrate findings into response workflows. Black box AI models particularly deep neural networks often fail to provide this clarity, leading to skepticism and underutilization in real world SOC (Security Operations Center) environments (Mohamed, 2025). Explainable AI (XAI) methods and human AI collaboration frameworks are therefore gaining traction to enhance trust and interpretability. Furthermore, adversarial robustness is emerging as a major concern. Attackers can exploit vulnerabilities in AI models by crafting adversarial inputs that cause misclassification or by poisoning training data to skew detection. Research in adversarial machine learning within CTI contexts remains nascent but essential to ensure the resilience of predictive intelligence pipelines (Zhang et al., 2023). Finally, the review highlights the importance of integrating human expertise into AI driven CTI. While automation accelerates analysis, human analysts provide contextual understanding, ethical judgment, and strategic insight that AI cannot replicate. A synergistic human AI collaboration framework where analysts guide, interpret, and refine AI outputs will likely define the future of CTI systems (Balasubramanian et al., 2025).

AI has reshaped CTI by introducing scalability, predictive analytics, and cross domain reasoning capabilities. Yet, realizing its full potential demands further research into explainable, trustworthy, and adaptive AI models, supported by high quality datasets and interdisciplinary collaboration between data scientists and cybersecurity experts. Future work should focus on developing multimodal, human centered CTI systems capable of learning continuously from dynamic threat landscapes while maintaining transparency, fairness, and accountability.

6.0 Datasets and Tools Used in AI Based CTI

The performance and reliability of Artificial Intelligence (AI) models in Cyber Threat Intelligence (CTI) are fundamentally dependent on the quality of datasets and the tools available for data collection, sharing, and analysis. Datasets provide the empirical foundation upon which AI models learn to recognize attack behaviors, while tools and frameworks enable the structuring and exchange of intelligence across systems and organizations. However, despite the rapid progress in AI based CTI research, data quality, accessibility, and

standardization remain persistent challenges that constrain real world deployment (Frias, Pereira, & Silva, 2025).

Datasets for AI Based CTI

AI models for CTI rely on large, labeled datasets to train and validate algorithms that detect and predict malicious activity. Among the most widely used are the benchmark intrusion detection datasets CICIDS2017 and CSE CIC IDS2018, developed by the Canadian Institute for Cybersecurity (Sharafaldin, Lashkari, & Ghorbani, 2018). These datasets contain a comprehensive mix of attack types, including brute force, DDoS, infiltration, and botnet activities. Researchers such as Maseer et al. (2021) and Li, Zhang, and Jiang (2024) have demonstrated that these datasets provide an effective baseline for evaluating both traditional machine learning and deep learning approaches. However, their synthetic and controlled nature limits realism, as they are generated in laboratory environments that may not accurately reflect dynamic, real world network traffic. Another important dataset, UNSW NB15, was designed to simulate modern network environments with multiple attack categories (Moustafa & Slay, 2015). This dataset has been widely adopted for testing classification algorithms, including Random Forest, Gradient Boosting, and Neural Networks, due to its structured and well balanced feature set (Singh, Sharma, & Kumar, 2024). Nevertheless, it does not fully capture the characteristics of recent threats such as supply chain intrusions or adversarial AI generated malware. Older datasets such as DARPA 1998/1999 and KDD Cup 1999 were historically significant for developing early intrusion detection systems, but they are now largely considered obsolete due to outdated network configurations and unrealistic traffic simulations (Khan, Javed, & Alghamdi, 2024). In addition to network based datasets, text based corpora have become increasingly important in AI driven CTI, particularly for models that focus on threat report analysis, entity extraction, and knowledge graph construction. Collections such as APTnotes, Cyber All Intel, and Open CyKG contain unstructured intelligence reports that describe threat actors, campaigns, and vulnerabilities (Rastogi, Dutta, Zaki, Gittens, & Aggarwal, 2021; Zhang, Liu, & Xu, 2023). These sources enable the training of Natural Language Processing (NLP) and transformer based models that identify indicators of compromise and predict potential attack relationships. However, textual CTI datasets often suffer from inconsistency, noise, and lack of standardized labeling, which complicates supervised learning efforts (Balasubramanian et al., 2025).

Emerging AI based CTI models are also leveraging dark web and open source intelligence (OSINT) data to detect early indicators of cyber threats. Nunes et al. (2016) explored dark web mining to identify hacker forum discussions and exploit markets, demonstrating the potential of such data for predictive intelligence. Despite their usefulness, these sources pose significant ethical and legal challenges related to privacy, surveillance, and the monitoring of restricted online communities. Collectively, these datasets form the empirical backbone of AI driven CTI, yet they highlight the pressing need for standardized, realistic, and up to date datasets that reflect the evolving nature of modern cyber threats.

Tools and Frameworks Supporting AI Based CTI

AI systems for CTI depend on a diverse ecosystem of tools and frameworks that enable structured information representation, interoperability, and automated intelligence sharing. Among these, the Malware Information Sharing Platform (MISP) is one of the most widely used open source tools for storing, correlating, and exchanging threat intelligence (Wagner, Dulaunoy, Wagener, & Iklody, 2016). MISP allows analysts to integrate multiple data feeds and apply AI based models for correlation and anomaly detection. Its compatibility with standardized data exchange formats, particularly STIX and TAXII, enhances interoperability

across organizations. STIX (Structured Threat Information Expression) and TAXII (Trusted Automated Exchange of Intelligence Information) are open standards developed by OASIS to facilitate the automated exchange of CTI (Barnum, 2014). STIX provides a structured schema for describing threat entities such as campaigns, TTPs, and IoCs, while TAXII defines protocols for secure data transmission. These standards serve as the backbone for most modern CTI pipelines, ensuring semantic consistency and enabling seamless integration with AI systems that process multi source intelligence (Mavroeidis & Bromander, 2021).

Another cornerstone of AI supported CTI is the MITRE ATT&CK framework, which offers a comprehensive taxonomy of adversarial tactics, techniques, and procedures. Many AI models use ATT&CK as an ontology for mapping extracted intelligence to known adversarial behaviors, thereby improving model explainability and contextual reasoning (Strom, Applebaum, Pendergast, & Miller, 2018). Graph based CTI frameworks such as TINKER and AttacKG often integrate MITRE ATT&CK as a semantic layer to enhance entity alignment and relationship extraction (Rastogi et al., 2021). Beyond these core frameworks, tools such as OpenCTI, TheHive, and Cortex enable automated intelligence ingestion and case management, integrating seamlessly with AI pipelines for scoring, prioritization, and triage. Additionally, Elastic Stack (ELK) and Splunk are commonly employed for log analysis, visualization, and real time deployment of AI driven detection models within security operation centers (Li et al., 2024). Together, these platforms form a technological ecosystem that bridges raw data with actionable intelligence, supporting both automated and human in the loop CTI workflows.

7.0 Challenges with Datasets and Tools

Despite these advancements, current datasets and tools face several limitations that hinder the full realization of AI's potential in CTI. A major concern is data quality and freshness many datasets fail to represent new threat vectors such as AI generated phishing campaigns or supply chain attacks, leading to poor model generalization (Frias et al., 2025). Label scarcity is another challenge; creating labeled CTI datasets requires expert knowledge and extensive manual effort, which restricts the scalability of supervised AI models (Balasubramanian et al., 2025). Furthermore, inconsistencies in data formatting and ontology design impede interoperability between CTI tools and AI systems. Although standards such as STIX and TAXII have alleviated some of these issues, many organizations still rely on proprietary formats, leading to integration inefficiencies (Mavroeidis & Bromander, 2021). Ethical and privacy concerns also arise when mining open or dark web data, where surveillance activities risk violating legal and moral boundaries (Nunes et al., 2016). Finally, integrating heterogeneous data sources including network traffic, textual intelligence, and behavioral logs requires significant engineering effort, especially when developing multimodal AI systems that operate across domains. To overcome these challenges, future research must emphasize standardized data creation, privacy preserving AI models, and federated learning architectures that enable secure collaboration without exposing sensitive information. Establishing publicly available, ethically sourced CTI datasets and open AI frameworks will be critical to advancing reproducible and trustworthy research in this field.

8.0 Challenges, Gaps, and Limitations

While Artificial Intelligence (AI) has substantially advanced the field of Cyber Threat Intelligence (CTI), several challenges and gaps continue to hinder its effective adoption in real world cybersecurity environments. These challenges span across data quality, model robustness, interpretability, standardization, and human AI collaboration. Addressing these issues is critical for developing CTI systems that are reliable, transparent, and resilient against evolving adversarial threats.

One of the most persistent challenges in AI driven CTI is the lack of high quality and up to date datasets (Frias, Pereira, & Silva, 2025). Most publicly available datasets, such as CICIDS2017 and UNSW NB15, were created in controlled laboratory environments and do not accurately capture the diversity, scale, and complexity of modern cyber threats (Sharafaldin, Lashkari, & Ghorbani, 2018; Moustafa & Slay, 2015). Consequently, models trained on these datasets often fail to generalize when deployed in operational contexts. Furthermore, label scarcity and the absence of standardized data annotation restrict the development of supervised learning systems (Singh, Sharma, & Kumar, 2024).

Another issue is data freshness and concept drift, where threat behaviors evolve over time, rendering historical data less representative. Attackers frequently modify tactics, techniques, and procedures (TTPs) to evade detection, leading to degradation in model performance (Khan, Javed, & Alghamdi, 2024). As a result, AI based CTI systems require mechanisms for continuous learning and adaptive model retraining to remain effective.

Interpretability remains a central concern in deploying AI models for cybersecurity. Security analysts must understand the reasoning behind an AI system's prediction to trust its outputs and take appropriate action (Mohamed, 2025). However, deep learning models especially those based on convolutional or transformer architectures are often black box systems that lack transparency (Balasubramanian et al., 2025). The absence of interpretability limits analyst trust and can result in resistance to AI adoption in security operations centers (SOCs).

Explainable AI (XAI) approaches, such as feature importance visualization, attention mechanisms, and post hoc interpretability tools like SHAP and LIME, have shown promise in bridging this gap. Yet, most existing XAI research focuses on generic domains such as healthcare or finance, with limited adaptation to CTI specific challenges (Frias et al., 2025). Future studies should prioritize domain specific XAI frameworks that explain predictions in the context of threat indicators, attack stages, and adversary profiles.

Ironically, while AI strengthens CTI, it also introduces new attack surfaces. Adversarial machine learning (AML) has emerged as a major concern, wherein attackers deliberately manipulate input data or model parameters to evade detection (Zhang, Liu, & Xu, 2023). Techniques such as evasion attacks, data poisoning, and model inversion can significantly degrade detection accuracy and compromise sensitive intelligence (Li, Zhang, & Jiang, 2024). For example, adversaries can craft benign looking traffic that deceives AI classifiers or inject malicious samples during training to bias the model's behavior.

Despite growing awareness of these vulnerabilities, adversarial robustness remains an underexplored area in AI based CTI research. Most defense strategies, such as adversarial training or ensemble methods, add computational overhead and have not been systematically tested in cybersecurity contexts (Khan et al., 2024). Developing resilient, attack aware AI architectures capable of detecting and mitigating adversarial manipulation should be a priority in future CTI research.

The lack of consistent data standards, ontologies, and integration protocols continues to impede seamless collaboration across CTI systems (Mavroeidis & Bromander, 2021). Although frameworks such as STIX, TAXII, and MITRE ATT&CK have improved structure and interoperability, variations in implementation often lead to semantic inconsistencies between organizations. Many AI pipelines rely on proprietary data formats that are not fully compatible with these standards, complicating automation and data exchange (Barnum, 2014).

In addition, knowledge representation gaps persist in CTI ontologies, limiting the effectiveness of graph based AI models that depend on precise entity and relationship definitions. To address this, researchers are exploring knowledge graph embeddings and semantic reasoning models that can align heterogeneous CTI schemas. However, achieving universal interoperability will require joint efforts among academia, industry, and standardization bodies (Rastogi, Dutta, Zaki, Gittens, & Aggarwal, 2021). AI based CTI systems are not designed to replace human analysts but to augment their decision making capabilities. However, realizing effective human AI collaboration remains a significant challenge. Many current systems lack intuitive interfaces, real time feedback mechanisms, and adaptive learning from human corrections. Analysts often perceive AI recommendations as opaque or unreliable, particularly when models misclassify benign events as malicious (Balasubramanian et al., 2025).

To build trust, CTI systems must integrate human in the loop learning frameworks that allow analysts to validate, adjust, and refine AI outputs. This interactive approach not only improves model performance over time but also enhances the interpretability and accountability of AI driven decisions (Mohamed, 2025). Moreover, cognitive and organizational factors such as analyst workload, skill level, and institutional trust should be considered when deploying AI in operational environments.

The integration of AI and CTI introduces several ethical and privacy related issues. Mining data from social media, dark web, or corporate telemetry can inadvertently expose personal information or violate legal boundaries (Nunes, Diab, Carvalho, Shakarian, & Shakarian, 2016). Additionally, predictive CTI models raise ethical questions about preemptive action such as whether organizations should act on AI predicted threats that have not yet materialized. Ensuring transparency, accountability, and compliance with data protection laws such as the General Data Protection Regulation (GDPR) is therefore critical (Balasubramanian et al., 2025). Future CTI research should emphasize privacy preserving AI and federated learning approaches that allow knowledge sharing across organizations without exposing sensitive data. Such methods will enhance collaborative threat intelligence while maintaining confidentiality and ethical integrity.

9.0 Future Directions

The integration of Artificial Intelligence (AI) into Cyber Threat Intelligence (CTI) has already transformed how organizations collect, analyze, and act upon security data.

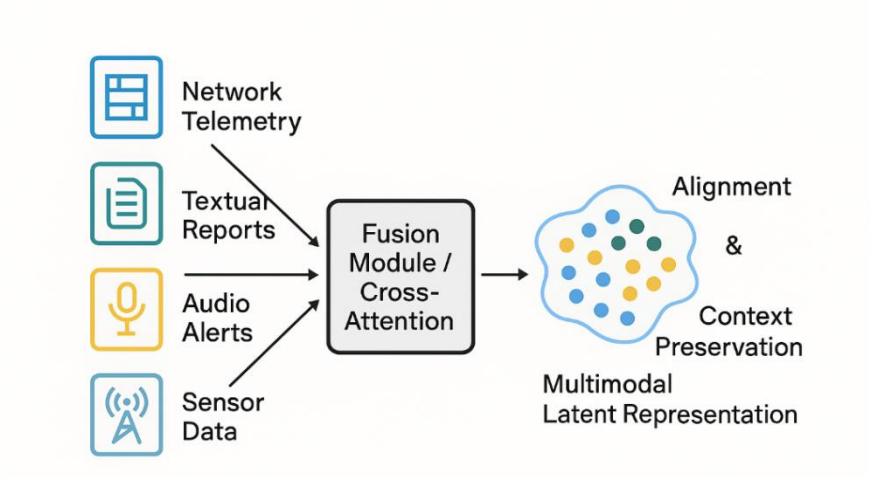

Figure 1: Multimodal Data Integration for AI-Driven Cyber Threat Intelligence.

However, to fully realize the potential of AI in CTI, future research must address existing limitations while exploring emerging opportunities in multimodal learning, explainability, and collaborative intelligence. One major direction for advancement is the development of multimodal AI models capable of combining diverse data types such as network telemetry, textual threat reports, social media feeds, and malware binaries into unified analytical frameworks. Current CTI systems often process each modality independently, which limits the depth of contextual understanding. By fusing structured and unstructured data, multimodal AI could generate richer situational awareness and predictive insight (Balasubramanian et al., 2025). For instance, combining large language models (LLMs) with graph neural networks (GNNs) could enable models to interpret narrative threat reports while simultaneously reasoning over relational data structures, thereby improving both accuracy and interpretability.

Another promising avenue is explainable and trustworthy AI (XAI) in CTI. Analysts must be able to understand why a model classifies an event as malicious or predicts a future attack. Future CTI systems should therefore integrate interpretable modeling techniques and post hoc explanation tools to provide transparency and traceability (Mohamed, 2025). Research into context aware explanations linking AI predictions to specific tactics, techniques, and procedures (TTPs) in frameworks such as MITRE ATT&CK will further strengthen analyst confidence and facilitate operational adoption. Federated and privacy preserving learning approaches also represent a key research frontier. Since CTI data often contain sensitive information, organizations are reluctant to share datasets. Federated learning allows multiple entities to collaboratively train AI models without directly exchanging raw data, maintaining privacy while improving model robustness (Li, Zhang, & Jiang, 2024). This paradigm could enable global scale threat intelligence sharing while respecting data governance and regulatory constraints such as GDPR. Moreover, the rapid advancement of Large Language Models (LLMs) offers unprecedented opportunities for CTI automation. LLMs can process unstructured intelligence reports, extract relevant entities, summarize threat campaigns, and even generate predictive hypotheses about attacker behaviors (Lampis & Dekker, 2024). However, these models must be fine tuned on domain specific corpora to mitigate hallucination, maintain factual accuracy, and align outputs with operational needs.

Human AI collaboration will remain a cornerstone of future CTI systems. Rather than replacing analysts, AI should augment human expertise by providing intelligent recommendations and adaptive decision support. Human in the loop (HITL) frameworks that combine human judgment with machine driven analysis can continuously improve performance through feedback and reinforcement learning (Balasubramanian et al., 2025). The emphasis should shift from pure automation to *symbiotic intelligence* a cooperative model where humans and AI co evolve to achieve superior threat detection and predictive insight. Finally, adversarial resilience will be a critical research domain in the coming years. As attackers increasingly exploit vulnerabilities in AI models, researchers must design robust learning systems capable of detecting adversarial inputs, mitigating data poisoning, and maintaining integrity under manipulation (Zhang, Liu, & Xu, 2023). Integrating adversarial defense strategies into CTI pipelines will ensure that AI systems themselves do not become new vectors of attack.

The future of AI driven CTI lies in the convergence of multimodal learning, federated architectures, explainable intelligence, and human AI synergy. Through these advancements, CTI can evolve into a proactive, transparent, and ethically grounded discipline capable not only of understanding present threats but also of anticipating the cyber challenges of tomorrow.

10. Conclusion

This review explored the contribution of Artificial Intelligence (AI) to the advancement of Cyber Threat Intelligence (CTI), emphasizing its role in threat detection, prediction, data analysis, and automation. The integration of machine learning, deep learning, and graph-based approaches has enhanced the adaptability and predictive power of CTI systems, enabling faster and more intelligent threat identification. Despite these advancements, challenges remain concerning data quality, model transparency, adversarial robustness, and interoperability across platforms. Hybrid and graph-based AI techniques represent promising directions but depend heavily on high-quality datasets and standardized frameworks. Future research should focus on developing explainable, multimodal, and privacy-preserving AI models that strengthen transparency, trust, and analyst confidence. Ultimately, the effectiveness of AI-driven CTI will rely on achieving the right balance between technological innovation and human expertise, fostering a collaborative environment where human analysts and intelligent systems work together to create proactive, predictive, and resilient cybersecurity defenses.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding

The author received no financial support for the research, authorship and publication of this article.

References

[1] Ahl, A., Yarime, M., Goto, M., Chopra, S. S., Kumar, N. M., Tanaka, K., & Sagawa, D. (2020). Exploring blockchain for the

energy transition: Opportunities and challenges based on a case study in Japan. Renewable and Sustainable Energy Reviews, 117, 109488.

[2] Balantrapu, S. S. (2024). AI for predictive cyber threat intelligence. IJMSD / IJMESD, 7(7).

[3] Balasubramanian, P., Liyana, S., Sankaran, H., Sivaramakrishnan, S., Pusuluri, S., & Pirttikangas, S. (2025). Generative AI for cyber threat intelligence: Applications, challenges, and analysis of real-world case studies. Artificial Intelligence Review.

[4] Barnum, S. (2014). Standardizing cyber threat intelligence information with the Structured Threat Information Expression (STIX). MITRE Corporation.

[5] Brown, R., & Sfakianakis, A. (2025). SANS 2025 CTI survey: Navigating uncertainty in today's threat landscape. SANS Institute.

[6] Darweesh, A., Khan, R., & Lin, W. (2024). Random forest-based network intrusion detection system for cloud environments. Journal of Network and Computer Applications, 235, 104901.

[7] Frias, M., Pereira, D., & Silva, T. (2025). Artificial intelligence in cyber threat intelligence: A systematic review. Computers & Security, 142, 103053.

[8] Hossen, M. N., Rahman, M., & Karim, M. R. (2021). Threat intelligence analysis using honeypot systems. International Journal of Information Security Science, 10(2), 87–99.

[9] Khan, R., Javed, A., & Alghamdi, T. (2024). Deep learning vs. machine learning in intrusion detection systems: Comparative evaluation and insights. Journal of Network and Computer Applications, 235, 104901.

[10] Kraven Security. (2025). The six phases of the cyber threat intelligence lifecycle.

[11] Lampis, A., & Dekker, M. (2024). Towards an AI-enhanced cyber threat intelligence processing pipeline. arXiv preprint arXiv:2403.03265.

[12] Li, Y., Zhang, C., & Jiang, W. (2024). Machine-learning-based threat detection and intelligence analysis: A comprehensive survey. IEEE Access, 12, 40567–40591.

[13] Mavroeidis, V., & Bromander, S. (2021). Cyber threat intelligence model and ontology. Computers & Security, 104, 102102.

[14] Mohamed, N. (2025). Artificial intelligence for predictive cyber threat intelligence. International Journal of Machine Learning and Cybernetics, 16(2), 215–229.

[15] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW NB15 Network Data Set). Military Communications and Information Systems Conference (MilCIS), 1–6.

[16] Nunes, E., Diab, A., Carvalho, M., Shakarian, P., & Shakarian, J. (2016). Darknet and deepnet mining for cyber threat intelligence: Identifying and predicting hacker activity. IEEE Symposium on Intelligence and Security Informatics (ISI), 7–12.

[17] Rastogi, A., Dutta, R., Zaki, M., Gittens, A., & Aggarwal, C. (2021). TINKER: A knowledge-graph framework for cyber threat intelligence. Proceedings of the IEEE International Conference on Big Data, 2815–2824.

[18] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion-traffic characterization. International Conference on Information Systems Security and Privacy (ICISSP), 108–116.

[19] Singh, P., Sharma, A., & Kumar, D. (2024). Machine-learning techniques for cyber threat prediction and analysis: Current trends and future challenges. Journal of Information Security and Applications, 80, 103651.

[20] Strom, B., Applebaum, A., Pendergast, A., & Miller, D. (2018). MITRE ATT&CK: Design and philosophy. MITRE Corporation.

[21] Wagner, C., Dulaunoy, A., Wagener, G., & Iklody, A. (2016). MISP: The design and implementation of a collaborative threat-intelligence-sharing platform. Workshop on Information Sharing and Collaborative Security (WISCS), 49–56.

[22] Zhang, L., Liu, X., & Xu, H. (2023). Reinforcement and graph learning for adaptive cyber-threat prediction. IEEE Transactions on Information Forensics and Security, 18, 4123–4142.