International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

"Cyber Security Threat Intelligence Using Artificial Intelligence"

Rohit V Kadu[1], Bhavika J Surode[2], Samruddhi Tayade[3]
Student, MCA II-year Sem IV
[3] Assistant Professor, P.G. Dept of Computer Applications,
PRMITR Badnera, Maharashtra, India.

***Abstract*-- Threat intelligence is the procurement of evidence-based knowledge regarding current or potential threats. The interest in threat intelligence comprises advancements in efficiency and boosting effectiveness in terms of analytical and prevention capabilities. Cybersecurity represents serious interest for numerous organizations because most of them use Internet-connected data devices, which open doors for cyber attackers. Outstanding threat intelligence within the cybersphere requires a knowledge base of threat information and a thoughtful way to represent this knowledge. This study proposes a clear rationale for the significant artificial intelligence (AI) techniques used to recognize cyber-attacks. Data analysis can be formulated to guide industries and Internet-connected systems, such as smartphones or robotic factories, on what to do in the event of an incident. AI techniques analyze past incidents and summarize knowledge from experts and will continue to adapt or reform new branches as they review new incidents. In addition, various data mining approaches used to boost threat truthfulness in cybersecurity data are studied. In conclusion, we discussed that AI will robotize the collation of machine-readable external threats and improve the efficiency and accuracy of the data for each smart organization's specific framework.**

***Keywords*-- Artificial intelligence, Cybersecurity, Intrusion detection, Threat intelligence.**

I. INTRODUCTION

In today's digital world, every organization is connected to different technologies, work cultures, and processes. This allows cyber attackers to stroll freely in the working environment. Every organization is likely to encounter a stream of attackers and intruders. They target both big organizations as well as small organizations in the public and private sectors. Therefore, unveiling cyber attackers and threat conditions requires cybersecurity defense tools, processes, and algorithms.

Cybersecurity refers to the set of algorithms and techniques used to preserve the integrity of nodes, networks, and data from damage, attacks, and illegal access. Large organizations have petabytes of data, including the most important and sensitive information; hence, it is important to protect the data from malicious access and threats.

As cyber threats continue to evolve in complexity and frequency, traditional security measures struggle to keep pace with rapidly changing threat landscapes. Cybercriminals employ increasingly sophisticated techniques, including advanced persistent threats (APTs), ransomware, phishing attacks, and zero-day exploits. These advanced attack strategies make it difficult for conventional security solutions to detect and mitigate risks effectively [1]. This leaves organizations vulnerable to breaches, data theft, and operational disruption. The growing interconnectivity of digital infrastructure further exacerbates these challenges, as organizations across industries face an ongoing battle to secure their digital assets against adversaries who exploit vulnerabilities at an unprecedented rate [2]

Machine learning has many potential applications in threat intelligence. A major unsolved issue in cyber threat intelligence is attacking attribution. Due to the asymmetric and remote nature of cyberattacks, it is very difficult to conclusively determine the

AI-driven cybersecurity not only strengthens an organization's ability to defend against cyber threats but also reduces the burden on human analysts. The sheer volume of security alerts and data logs generated daily can overwhelm even the most experienced security teams, leading to alert fatigue and decreased response times. AI assists in filtering out false positives, prioritizing genuine threats, and automating routine security tasks,

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

Figure 2: AI-Driven Data Processing Workflow.

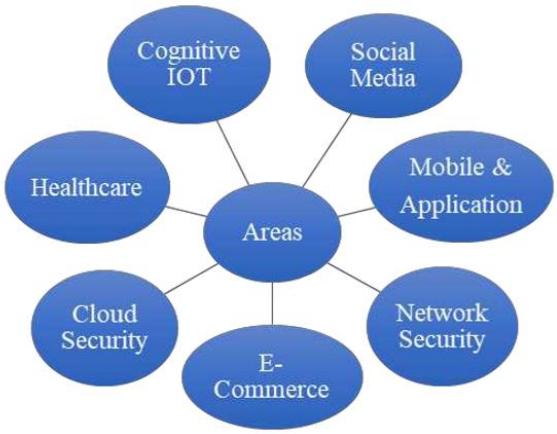

Figure 1: Research Areas of Cyber Security.

II. Literature Survey

Cyber Security Threat Intelligence using Artificial Intelligence

CTI highlights the creation of a multilayered threat intelligence system. A search robot was developed that analyzes the false activities of Internet resources by self-learning patterns based on the workings of neural networks. The goal of pattern generation technique is to train the system to learn newly generated threats and generate an alert against new

Processes, whether they are safe or not. Information foraging (IF) addresses issues related to the volume and velocity of data generated by stabilizing human intuition with automation. This addresses that information foraging is helpful in expansion of tools to anticipate cyber intrusion using publicly attainable data.

Sr. No.	Author(s) & Year	Title of Paper	Methodology / Techniques	Key Findings	Limitations
1	S. Dua & X. Du (2020)	Data Mining and Machine Learning in Cybersecurity	Machine Learning, Data Mining	ML assists detect cyber threats efficiently in large datasets	High false positives in some cases
2	A. Apruzzese et al. (2019)	Machine Learning for Cybersecurity: A Survey	Supervised & Unsupervised Learning	AI improves intrusion detection systems	Requires high-quality labeled data
3	M. Conti et al. (2018)	A Survey on Security and Privacy Issues of IoT	AI-based anomaly detection	AI assists secure IOT devices from attacks	Scalability issues in large networks
4	R. Sommer & V. Paxson (2010)	Outside the Closed World: ML for Network Intrusion Detection	Statistical ML models	ML models can detect unknown attacks	Difficult to generalize models
5	I. Good fellow et al. (2016)	Deep Learning for Cyber Security	Deep Neural Networks	Deep learning improves threat detection accuracy	High computational cost
6	Y. Xin et al. (2018)	Machine Learning and Deep Learning Methods for Cybersecurity	Deep Learning, SVM, Random Forest	AI techniques outperform traditional security methods	Training complexity and time
7	B. Miller et al. (2020)	AI-based Cyber Threat Intelligence	Natural Language Processing (NLP)	NLP extracts threat intelligence from reports	Limited contextual understanding
8	K. Kim et al. (2019)	Deep Learning in Intrusion Detection Systems	CNN, RNN models	Deep learning detects complex attack patterns	Needs large datasets
9	J. Zhang et al. (2021)	AI-driven Threat Intelligence Systems	AI automation, predictive analysis	Predictive models can prevent future attacks	Data privacy concerns

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

The Weighted Random Forest (WRF) technique and cluster-based Weighted Random Forest approach were used to reform the robustness of the random forest. These approaches used bagging to boost the uncertainty of the models. In addition, the sparse computation that generates a similarity matrix based on a pair of objects that share the same neighborhood was projected efficiently for massively large

Accuracy Comparison

Computational Cost

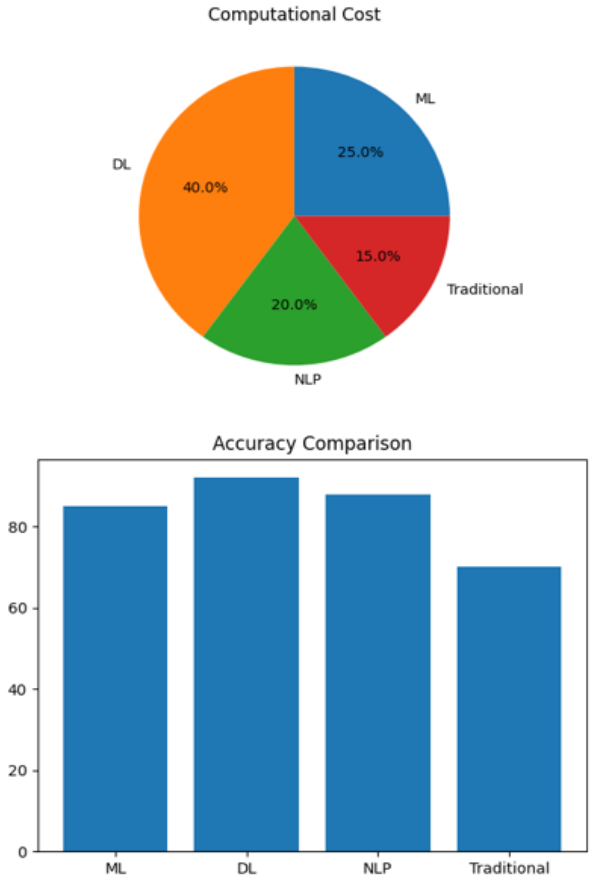

Figure 3: Comparison Graphs.

This technique resulted in a reduction in the testing time with minimal effect on accuracy. With the advent of the Internet, criminal activities have also accelerated; therefore, a solution was proposed to detect unauthorized activity based on anomaly detection and database signatures using data mining techniques. This infers that the intrusion detection system helped recognize security threats in the system.

III. Methodology

AI-powered threat intelligence uses ML algorithms to analyze vast amounts of data from multiple sources, including network traffic, system logs, and external threat databases. The methodology involves several key steps that contribute to a comprehensive cyber security framework.

The data collection phase serves as a foundational step in AI-driven cybersecurity, where structured and unstructured data are gathered from diverse sources to enhance threat detection and analysis. AI systems acquire data from network endpoints, cloud environments, security feeds, and external threat intelligence platforms to ensure comprehensive coverage of the cybersecurity landscape. Network endpoints, including computers, routers, and IoT devices, generate logs and activity records that provide insights into potential security threats to the network. Cloud environments contribute authentication logs, access patterns, and telemetry data from services such as AWS, Azure and Google Cloud. Additionally, security feeds from intrusion detection systems (IDS), firewalls, and antivirus tools offer real-time security event data, whereas external intelligence platforms provide updates on emerging threats, vulnerabilities, and adversarial tactics.

Once collected, the data are aggregated and centralized into repositories, such as security data platforms or data lakes, allowing for efficient storage, normalization, and standardization across multiple formats. This process ensures consistency, reduces redundancy, and facilitates large-scale data processing. The integration of multiple data sources enhances the accuracy and reliability of AI-driven security analyses by creating a holistic cybersecurity view. By leveraging vast and diverse datasets, AI can improve threat detection capabilities, reduce false positives, and identify anomalies with high precision. The comprehensive data collection process enables AI-driven cybersecurity systems to proactively detect, analyze, and mitigate security threats, enhancing overall cyber resilience

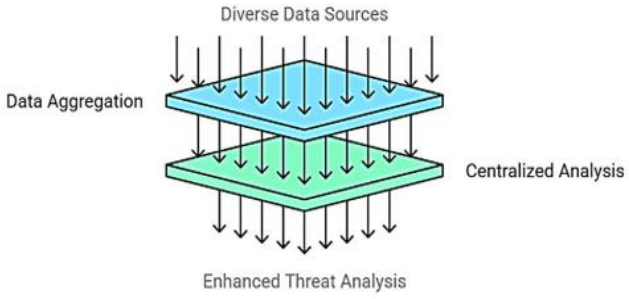

Figure 4: Cybersecurity Data Processing Architecture.

Raw cybersecurity data often contain inconsistencies, missing values, or duplicate records that can hinder effective analysis. Data cleaning ensures the removal of incomplete, irrelevant, or corrupted data points, thereby improving the integrity of the dataset. Normalization standardizes the data by converting different formats, units, or scales into a uniform structure, thereby allowing AI algorithms to process and compare the data efficiently. This step is particularly important when integrating data from multiple sources, such as security logs, cloud telemetry, and external threat intelligence feeds [15].

After normalization, AI systems apply data transformation techniques to convert raw data into structured formats that are suitable for analysis. This process may include encoding categorical variables, aggregating event logs, or structuring the free-text data for further processing. One of the most crucial aspects of data processing is feature extraction, in which AI identifies key attributes, patterns, and relationships within the dataset. Feature extraction assists recognize suspicious behaviors, such as unusual login patterns, lateral movement within networks, or deviations from normal user activity. Machine learning models rely on these extracted features to classify threats and predict cyberattacks with high accuracy.

Cybersecurity datasets are often noisy and contain large volumes of irrelevant data, such as harmless system alerts, routine network traffic, or false-positive security events. AI-driven systems employ anomaly detection algorithms, clustering techniques, and statistical filtering to differentiate between normal and harmful activities. This step ensures that AI focuses on high-priority threats, minimizing the risk of alert fatigue for cybersecurity analysts

Once the data processing is complete, the refined and structured dataset serves as the foundation for AI-driven threat intelligence. By reducing noise, extracting critical features, and applying advanced analytical techniques, AI can prioritize cybersecurity threats, detect zero-day vulnerabilities, and provide actionable insights for security teams to take appropriate measures. The processed data are then fed into machine learning models, behavior analytics engines, and security automation frameworks to enable real-time threat mitigation and proactive defense mechanisms.

AI-driven cybersecurity systems utilize multiple ML techniques to improve threat detection capabilities.

Anomaly Detection: AI identifies deviations from normal behavior by analyzing historical data and establishing a baseline of expected system activities.

Any significant deviation, such as unusual login times, irregular data transfers, or unexpected network access, is flagged as a potential security threat to the system. This approach is particularly useful for detecting insider and advanced persistent threats (APTs).

AI-driven threat detection systems operate in real time, ensuring that security teams receive immediate alerts upon the detection of suspicious activities. Automated Security Orchestration, Automation, and Response (SOAR) platforms integrate AI-driven detection with automated response mechanisms, allowing for the rapid containment of threats. AI can trigger automated actions, such as blocking malicious IP addresses, isolating compromised endpoints, or enforcing additional authentication measures when a security risk is identified. To improve accuracy and reduce false positives, AI integrates external threat intelligence feeds that provide updated information on emerging attack techniques, malware signatures, and threat-actor behaviors. By continuously learning from global cybersecurity data, AI enhances its ability to detect new and evolving threats with greater precision

Data Collection (Cyber Attack Datasets)

Data collection is the first and foremost step while building AI-based cyber threat intelligence system. This includes collecting vast amounts of data related to cybersecurity from various sources. Some of these sources are network traffic logs, system logs, firewall logs, intrusion detection systems (IDS), antivirus reports and external threat intelligence feeds. Some of the most popular datasets that are widely used for research included KDD Cup 99 NSL-KDD CICIDS and UNSW-NB15. These datasets are necessary for training machine learning models and include normal (benign) and malicious (attack) data. Data can either be structured, semi-structured or unstructured. For example, structured data can be logs or any other numerical value while unstructured data can text from reports or emails. Gathering varied data ensures security teams can identify various types of cyber threats, including malware, phishing (to obtain sensitive information), DDoS attacks and ransomware. On the other hand, real-time data collection is critical for identifying ongoing attacks.

The performance of models depends a lot upon the quality of collected data. Data that is incomplete or biased can result in inaccurate predictions. Hence the need for data to be accurate, relevant and complete. Privacy and security regulations are also necessary for data collection systems to ensure sensitive information is protected. Handling large datasets effectively typically involves using data lakes or cloud storage.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

- *Data Preprocessing*

Data preprocessing: Data preprocessing is the method of cleaning and preparing raw data for analysis. Cybersecurity-related data is often noisy, has missing or duplicate values, and irrelevant information. These issues can hinder the efficiency of machine learning models hence preprocessing is vital.

Removing or correcting inaccurate and incomplete data: Data cleaning is the initial step of preprocessing. Missing values can be filled with mean, median or given default value. Redundant records are removed from Table I. Next comes data normalization that assists in scaling the values into a common range to be easily processed by algorithms.

Label encoding or one-hot encoding is used to convert categorical data like attack types into numerical format. For text-based data, you might consider tokenization. Model accuracy is further improved by applying feature scaling techniques such as Min-Max scaling or standardization. The next important step is data reduction, which simplifies the complexity by eliminating unnecessary features. Preprocessing also includes dividing the dataset into training and testing sets, e.g., 80% training and 20% testing. Preprocessing assists to make sure that the data is consistent, clean and format ready for machine learning analysis.

- *Feature Extraction*

Feature Extraction: It involves selecting significant features from the dataset essential to detect the cyber threat. Features of the input variables used by a machine learning model to make predictions. Examples of such features could be IP address, packet size, login attempts, and time interval and protocol type in cyber security. Good feature extraction will help improve model performance, and also reduce cost. Methods for feature extraction include statistical analysis, frequency analysis and pattern recognition. In case of text-based data TF-IDF (Term Frequency-Inverse Document Frequency), Count Victimizer and Word Embedding's (Word2Vec) are some of the techniques used.

Then comes the part of feature selection itself, where we only choose relevant features. It keeps noise down and fits your case better. We may apply dimensionality reduction methods such as PCA (Principal Component Analysis) to high dimensional data.

Feature extraction in cybersecurity aids in detecting anomalous behavior like login at odd hours, unordinary traffic patterns, or frequent failed login attempts.

These characteristics help AI models to identify normal and harmful behaviors. Selecting all the most relevant features can greatly improve threat detection capabilities.

- *Model Training (AI/ML Algorithms)*

Model Training in Machine Learning In this process, model training occurs on labeled data with known inputs (features) and outputs (attack or normal). You want to train a model which will be capable of classifying cyber threats accurately.

Support Vector Machine (SVM), Decision Tree, Random Forest, K-Nearest Neighbors (KNN) and Neural Networks are some of the commonly used algorithms in Cybersecurity. Higher-level deep learning models like Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are utilized for complex pattern recognition.

So, you feed data, adjust parameters, and try to minimize errors using some optimization techniques. Tune hyper parameters to improve the performance of the model. The model is trained techniques to make sure it generalizes well on unseen data.

Training — particularly on large datasets — takes huge computational resources. You need to balance the model, cover overfitting and under fitting. Under fitting and Overfitting: Under fitting refers to the model not learning patterns correctly, while overfitting when it knows the training data but performs poorly on new unseen data.

A good model trained on relevant data can be both precise in detecting attacks as well as increasing its capabilities with each new threat, making it a key component of AI-based cybersecurity systems.

- *Threat Detection*

Threat detection: Based on trained AI models, the detecting process identifies malicious actions. After training the model, it is deployed and utilized in real-time systems enhancing the monitoring of network traffic and behavior on systems. It does so by analyzing incoming data and comparing it with previously learned patterns to identify anomalies or known attack signatures.

AI-powered threat detection systems can detect different kinds of cyber threats like malware, phishing attacks, and insider threats. They are used to identify unusual behavior in deviation from common behaviors, a task typical of anomaly detection methods. However, signature-based detection is only effective against known attacks.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

Timely detection is identified as key to reducing damage. Threat detection by AI systems can provide alerts in real-time. More advanced systems can take automated actions like block IP addresses, quarantine infected systems or trigger security protocols.

Threat detection systems need high accuracy to avoid false positives (normal activity interpreted as attack) and false negatives (attack not detected). These models must be constantly monitored and updated as threats evolve, however.

Evaluation: Process to measure the accuracy of trained model. It serves as an indicator of the model's capability to detect cyber threats. In order to measure model accuracy and effectiveness, there are multiple evaluation metrics used.

Some of the common metrics include Accuracy, Precision, Recall, and F1-Score. Accuracy: This is the overall correctness of the model. Precision tells you how many of the threats detected were correct. Recall is how well the model can identify all existing threats. F1 Score is the harmonic mean of precision and recall.

A confusion matrix is usually used to visualize model performance with contrasting true positives, true negatives, false positives and false negatives. In the field of cybersecurity, it is preferred to have high accuracy coupled with low false positives.

Test data which is not used while training is utilized to evaluate. Cross-validation techniques ensure reliable results. And if it does a poor job, the tuning of parameters, addition of more data or selection of features assists improve performance.

Cyber threat landscape changes with time, so evaluation has to be continuous. That's why models are regularly retrained and updated.

IV. Approach

Research Questions

We address the following research questions (RQs) in this study through extensive experiments and analyses:

1) RQ1 (Few-Shot): How effective are LLMs at extracting CTI information in a few-shot setting?

-For this, we investigate prompt engineering & guidance frameworks

2) RQ2 (Fine-tuning): How much does fine-tuning LLMs does labelled data improve CTI information extraction?

-For this, we employ the parameter-efficient QLORA

Method.

3) RQ3 (Knowledge Graph Quality): What is the quality of the knowledge graph of triples generated from a large CTI corpus using LLMs, and how can we improve them?

-For this, we identify shortcomings and propose error reduction methods.

4) RQ4 (Link Prediction): How does CTI-derived and how does a knowledge graph perform link prediction?

-For this, we considered both transudative and inductive settings.

Artificial intelligence is the ability of machines and software to mimic human behavior. It assists to determine the problem in a manner similar to humans. The presence of AI is invariably expanding in the cyber world because of its tireless performance of tasks. The main motivation for intelligent machines is that decisions made are more logical and appropriate because of the absence of emotions. In contrast, humans make decisions after considering everything, including their emotions. Some ethical decisions are being encountered, such as supporting or opposing the development of lethal autonomous weapons systems (LAWS). Ethical issues are developing as we give more importance and power to robots. Ethical issues can be addressed by writing appropriate codes and testing issues properly. A hybrid system combining AI and humans is an intelligent solution for implementing machine intelligence. Knowledge Engineering (KE) reviews the structure of a decision to recognize how a conclusion is attained. It is widely practiced by engineers, who integrate KE into decision support software to gain the ability to identify a face or parse what a person says for meaning. Soon, the area of knowledge engineering will help create systems that can solve problems better than humans.

V. Results And Discussion

The implementation of Artificial Intelligence (AI) techniques in Cyber Security Threat Intelligence has shown promising and impactful results. During the study, various AI models, such as Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP), were analyzed to understand their effectiveness in detecting and predicting cyber threats.

The results indicate that AI-based systems significantly improve the speed and accuracy of threat detection compared with traditional security methods. Machine Learning algorithms can identify unusual patterns in network traffic, which assists in the early detection of potential cyber-attacks, such as phishing, malware, and intrusion attempts.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 04, April 2026)

Deep Learning models, particularly neural networks, have demonstrated higher accuracy in recognizing complex attack patterns owing to their ability to process large volumes of data.

Another important observation from the results is the role of Natural Language Processing in extracting threat intelligence from unstructured data sources, such as security reports, blogs, and dark web content. This capability allows organizations to remain updated on emerging threats in real time.

However, this discussion also highlights certain challenges. One of the major issues is the dependency on high-quality labeled datasets. Poor data quality can lead to incorrect predictions and an increased number of false positives. Additionally, AI models, especially deep learning systems, require significant computational resources, which may not be feasible for all organizations to obtain.

Furthermore, although AI enhances automation in cybersecurity, it also introduces new risks. Attackers can potentially use AI techniques to create more sophisticated and harder-to-detect cyber threats. This creates a continuous cycle in which both defenders and attackers leverage advanced technologies.

Overall, the findings suggest that AI-driven Cyber Threat Intelligence systems are highly effective in improving cybersecurity defenses. However, for optimal performance, challenges related to data quality, system scalability, and ethical concerns must be addressed. Future improvements should focus on developing more efficient models with lower computational requirements and better adaptability to evolving cyber threats.

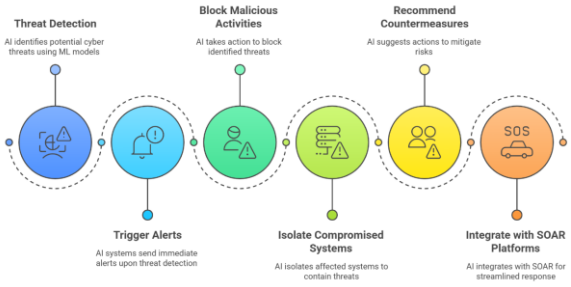

Figure 5: Graphical Representation of Various Parameters Cyber Security Threat Intelligence using AI.

VI. Conclusion

This research clearly shows that Artificial Intelligence plays an important role in improving cyber threat intelligence.

Compared to traditional security methods, AI-based systems are much faster and more accurate in detecting cyber-attacks. By using techniques like machine learning and deep learning, these systems can analyse large amounts of data, identify unusual patterns, and respond to threats in real time. These assists organizations protect their data and systems more effectively. Overall, AI makes cybersecurity smarter, more efficient, and better prepared to handle modern cyber threats.

REFERENCES

[1] Abrahams, T., Ewuga, S., Dawodu, S., Adegbite, A., & Hassan, A. (2024). A REVIEW OF CYBERSECURITY STRATEGIES IN MODERN ORGANIZATIONS: EXAMINING THE EVOLUTION AND EFFECTIVENESS OF CYBERSECURITY MEASURES FOR DATA PROTECTION. Computer Science & IT Research Journal. https://doi.org/10.51594/csitrj.v5i1.699

[2] Ande, R., Adebisi, B., Hammoudeh, M., & Saleem, J. (2020). Internet of Things: Evolution and technologies from a security perspective. *Sustainable Cities and Society*. https://doi.org/10.1016/J.SCS.2019.101728.

[3] Ozkan-Okay, M., Akin, E., Aslan, Ö., Kosunalp, S., Iliev, T., Stoyanov, I., & Beloev, I. (2024). A Comprehensive Survey: Evaluating the Efficiency of Artificial Intelligence and Machine Learning Techniques on Cyber Security Solutions. IEEE Access, 12, 12229-12256. https://doi.org/10.1109/ACCESS.2024.3355547.

[4] Sarker, I., Sarker, I., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big Data, 7. https://doi.org/10.1186/s40537-020-00318-5

[5] Rhode, M., Burnap, P., & Jones, K. (2017). Early Stage Malware Prediction Using Recurrent Neural Networks. Compute. Secure. 77, 578-594. https://doi.org/10.1016/j.cose.2018.05.010.

[6] Benzaid, C., & Taleb, T. (2020). AI for Beyond 5G Networks: A Cyber-Security Defense or Offense Enabler?. IEEE Network, 34, 140-147. https://doi.org/10.1109/MNET.011.2000088.

[7] Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., & Wang, C. (2018). Machine Learning and Deep Learning Methods for Cybersecurity. IEEE Access, 6, 35365-35381. https://doi.org/10.1109/ACCESS.2018.2836950.

[8] K. K. R. Yanamala, "Integration of AI with traditional recruitment methods," Journal of Advanced Computing Systems, vol. 1, no. 1, pp. 1–7, Jan. 2021.

[9] Sarker, I., Furhad, M., & Nowrozy, R. (2021). AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. SN Computer Science, 2. https://doi.org/10.1007/s42979-021-00557-0.

[10] Timmers, P. (2019). Ethics of AI and Cybersecurity When Sovereignty is at Stake. Minds and Machines, 29, 635 - 645. https://doi.org/10.1007/s11023-019-09508-4.