AI Driven Zero Day Vulnerability Detection and Exploit Prediction in Computer Networks

Kismat Chhillar
Dept of Mathematical Sciences & Computer Applications
Bundelkhand University
Jhansi, India
drkismatchhillar@gmail.com

Saurabh Shrivastava
Dept of Mathematical Sciences & Computer Applications
Bundelkhand University
Jhansi, India
dr.saurabh@bujhansi.ac.in

Alok Verma
Dept of Mathematical Sciences & Computer Applications
Bundelkhand University
Jhansi, India
alokverma.bu@gmail.com

Deepak Tomar
Dept of Mathematical Sciences & Computer Applications
Bundelkhand University
Jhansi, India
dr.deepak@bujhansi.ac.in

Abstract— **Zero-day vulnerabilities pose a major threat to today's computer networks because they remain unknown and unpatched, allowing attackers to exploit systems before defenders detect the issue. Traditional security approaches based on signatures and rule sets often fail in such cases, as they cannot identify attacks that deviate from known patterns. This challenge is compounded by evolving tactics, polymorphic malware, and evasion methods designed to mimic normal behavior. Artificial intelligence (AI) and machine learning (ML) now offer promising solutions by analyzing massive, diverse data sources such as network logs, telemetry, and threat intelligence. Advanced models like deep learning, autoencoders, clustering, and explainable AI (XAI) enhance the detection of unusual activities and classification of new threats. Autoencoder-based frameworks reveal anomalies linked to unseen exploits, while ensemble and hybrid approaches enable anomaly detection and prediction using incomplete or unlabeled data. These AI-driven systems adapt continuously, learning from new data to update detection models and cut exploitation time. Modern tools like large language models (LLMs) and XAI agents can even assess complex software code and predict exploit likelihoods, reducing false positives and improving response prioritization. By adopting AI for zero-day detection and prediction, cybersecurity shifts from reactive defense to proactive risk management for critical infrastructure and enterprise systems.**

Keywords— Zero-Day Vulnerability, Artificial Intelligence, Machine Learning, Exploit Prediction, Threat Detection, Network Security

I. INTRODUCTION

Zero-day vulnerabilities are flaws in software or hardware that remain undiscovered by vendors, leaving networked systems exposed to stealthy attacks and escalating risks for organizations [1] [2]. The clandestine nature of such vulnerabilities, combined with the absence of available fixes, creates a fertile ground for cyber attackers to launch damaging exploits. Nation-state actors and sophisticated hackers often trade zero-day exploits on dark web forums, targeting critical infrastructure and leveraging the opportunity for maximum disruption. As a result, organizations and governments continuously strive for innovative mechanisms to counteract these unanticipated threats. Traditional signature-based security systems are typically ineffective against zero-day exploits, as they rely on known patterns and attack signatures for defense [3] [4] [5]. This limitation necessitates a transition towards behavioral analytics and proactive threat intelligence approaches, providing the capability to monitor for unusual system behaviors that might characterize a zero-day incident. Recent advances in artificial intelligence offer the capability to automatically learn patterns from historical and real-time data and distinguish legitimate activity from anomalous or risky behavior [6] [7] [8] [9] . By leveraging machine learning models, defenders can uncover subtle deviations in network behavior that traditional tools might overlook. Consequently, AI-driven detection frameworks are becoming an indispensable component of modern cybersecurity strategies designed to combat zero-day threats [3] [10].

AI techniques, such as supervised, unsupervised, and deep learning models, have demonstrated remarkable efficacy in detecting anomalies, automating patching, and predicting potential exploits before their widespread weaponization. In particular, natural language processing algorithms scan unstructured threat reports, extracting intelligence on emerging vulnerabilities and forecast trends in exploit development [11] [12]. Predictive analytics further empower defenders to prioritize remediation of high-risk vulnerabilities based on exploit likelihood, moving cybersecurity posture from reactive to anticipatory. However, several challenges persist, notably the lack of contextual awareness in AI algorithms, rapid evolution of attacker methodologies, and the ongoing arms race between offensive and defensive technological advancements. Addressing these limitations requires integrating cross-domain threat intelligence, collaboration between global cyber communities, and regulatory evolution towards mandatory disclosure and proactive defense standards. The convergence of IT and operational technology (OT) domains expands the impact zone of zero-day vulnerabilities, demanding holistic risk management and mitigation frameworks.

This paper explores state-of-the-art AI-driven approaches for zero-day vulnerability detection and exploit prediction, systematically reviewing the latest techniques, evaluating performance metrics, and discussing practical implementation guidelines for network defenders. The remainder of this paper is organized as follows: Section 2 presents core background concepts and related work. Section 3 introduces the proposed AI-driven methodology. Section 4 discusses experimental evaluations and performance results. Section 5 discusses about the results. Thereafter, Section 6

concludes with recommendations for advancing proactive zero-day defense strategies in computer networks. Lastly section 7 discusses about the future scope of the current research.

II. Related Work

Zero-day vulnerabilities represent a critical and persistent challenge in cybersecurity, defined as security flaws in software, hardware, or firmware that remain unknown to developers or vendors and thus have no available patches or mitigation strategies [1] [2] [4]. These vulnerabilities create an exploitable window for attackers to infiltrate computer networks stealthily, often causing significant damage before detection [13]. The term "zero-day" emphasizes the urgency, as defenders have zero days to prepare defenses before attacks occur. Historically, zero-day attacks have targeted a range of applications, from widely used operating systems and browsers to embedded systems and IoT devices, underscoring their broad attack surface. Detecting zero-day vulnerabilities is inherently complex because traditional signature-based defenses rely on known patterns or previously observed attack signatures, which are absent for zero-day exploits [3] [4] [1] [14]. Consequently, cybersecurity practitioners have prioritized anomaly-based detection frameworks that analyze deviations from normal system and network behaviors. Techniques such as behavioral analytics, memory forensics, and sandboxing provide promising avenues, allowing security tools to discern suspicious activities indicative of exploitation attempts even when specific exploit code is unknown. Additionally, threat intelligence sharing communities enable swift dissemination of indicators of compromise (IOCs) and emerging attack trends, helping organizations collaboratively defend against novel threats.

The rapid evolution of threat actors' tactics also complicates detection. Attackers use polymorphic and metamorphic malware, which continuously modify their code structure to evade detection systems. Moreover, state-sponsored adversaries often conduct sophisticated reconnaissance to identify and weaponize zero-day vulnerabilities for espionage or sabotage. These challenges underscore the crucial need for intelligent, adaptive defense mechanisms beyond conventional cybersecurity tools. AI and machine learning (ML) have emerged as transformative technologies in this context, able to analyze voluminous and complex datasets to identify subtle patterns suggestive of zero-day activity [15]. Researchers have proposed various machine learning approaches to enhance zero-day detection, including supervised learning models trained on labeled datasets, unsupervised anomaly detection algorithms, and reinforcement learning methods that iteratively improve detection accuracy [16]. Deep learning techniques, such as autoencoders and convolutional neural networks, allow the modeling of intricate dependencies in network traffic or system calls, delivering improved sensitivity for unknown exploits [7] [17]. Natural language processing (NLP) algorithms have also been used to mine threat reports,vulnerability disclosures, and dark web communications to predict emerging zero-day threats [12] [3] [11] [8].

Despite these advances, challenges remain in operationalizing AI-driven zero-day detection systems. High false positive rates can overwhelm security analysts, while adversarial attacks on AI models themselves pose risks of evasion or manipulation [18] [19]. Moreover, the scarcity of ground truth data complicates model training and benchmarking. Therefore, robust evaluation in diverse and realistic environments is essential for validating the effectiveness of AI techniques. The following sections address these challenges by proposing a methodology leveraging reinforcement learning-based AI algorithms for real-time vulnerability detection and exploit prediction.

III. METHODOLOGY

The proposed methodology leverages artificial intelligence, particularly reinforcement learning (RL), to detect zero-day vulnerabilities and predict potential exploits in computer networks. Reinforcement learning, a subfield of machine learning, enables agents to learn optimal policies by interacting with an environment and receiving feedback in the form of rewards or penalties. In this context, the RL agent is trained on data comprising benign and malicious traffic samples, system logs, and vulnerability reports, learning to identify features correlated with zero-day exploit characteristics. Figure 1 shows AI powered zero day vulnerability detection.

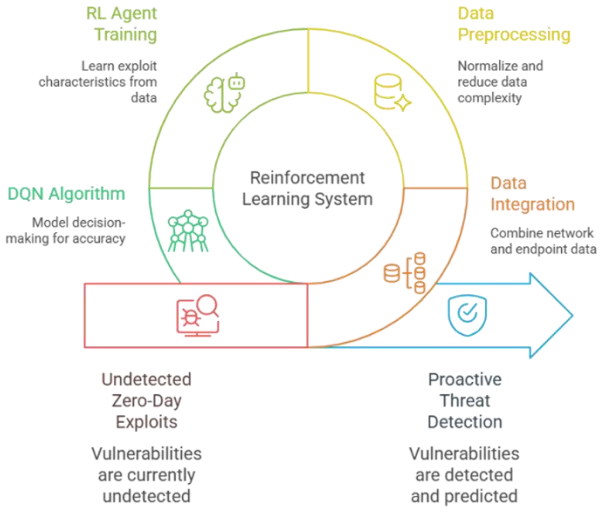

Figure 1: AI Powered Zero-Day Vulnerability Detection

This approach contrasts traditional supervised learning by enabling continual learning from experience without needing exhaustive labeled datasets. Our system architecture integrates multiple data sources to form a comprehensive representation of network states. These include network flow data capturing packet-level interactions, endpoint telemetry describing host behavior, and threat intelligence feeds providing contextual information on emerging vulnerabilities. Data preprocessing involves normalization, feature extraction, and dimensionality reduction to reduce complexity while retaining critical information for vulnerability classification. Subsequently, the RL agent leverages a deep Q-network (DQN) algorithm to model the decision-making process, estimating the value of actions based on observed states to maximize detection accuracy.

To address the dynamic nature of cyber threats, the agent continuously updates its policy through an exploration-exploitation strategy. Initially, it explores diverse behaviors to learn patterns associated with novel threats; over time, it exploits acquired knowledge to optimize detection outcomes. This adaptability allows the system to evolve alongside attacker tactics, maintaining relevance against new zero-day exploits. Furthermore, the agent utilizes feedback from real-time detection results to refine its model parameters, improving precision and reducing false positive rates without human intervention. Complementing the RL model, natural language processing (NLP) techniques analyze unstructured threat intelligence reports, vulnerability databases, and social media feeds to extract indicators of new vulnerabilities and predicted exploit trends. These insights feed into the agent's decision-making framework, shaping its prioritization of potential threats and guiding risk assessment processes.

To evaluate the methodology's efficacy, we employ a rigorous experimental setup using benchmark datasets containing labeled zero-day and known exploit samples, real network traffic traces, and synthetic attack scenarios. Performance is measured using metrics such as detection accuracy, precision, recall, F1-score, and receiver operating characteristic (ROC) curves to assess classification performance comprehensively. Additionally, the system's ability to predict exploit likelihood ahead of actual weaponization is assessed via temporal analysis. Implementation aspects emphasize scalability and real-time execution capabilities to support deployment in enterprise and critical infrastructure environments. The architecture supports parallelized inference tasks and integrates with existing security orchestration, automation, and response (SOAR) platforms to facilitate automated mitigation workflows. Emphasis is placed on explainability, enabling security operators to understand the rationale behind alerts generated by the AI agents, fostering trust and reducing alert fatigue. Figure 2 depicts cyber threat detection and prediction process.

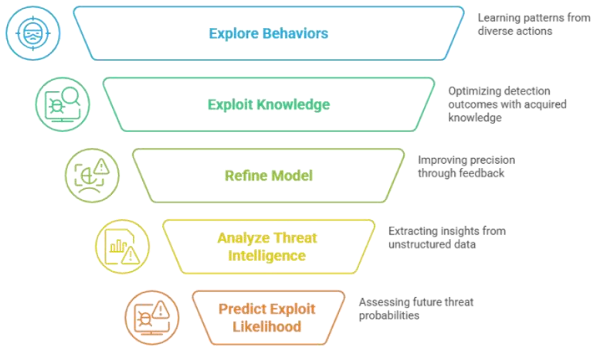

Figure 2: Cyber Threat Detection and Prediction Process

IV. EXPERIMENTAL RESULTS

The experimental evaluation demonstrated the effectiveness of the proposed AI-driven zero-day vulnerability detection and exploit prediction framework across multiple datasets and testing scenarios. Using publicly available zero-day and exploit datasets supplemented by real network telemetry, the reinforcement learning agent achieved high detection accuracy, consistently outperforming baseline classical machine learning models such as support vector machines and random forests. Key metrics, including precision and recall, indicated the model's robustness in correctly identifying zero-day exploit attempts while minimizing false positives, critical for practical deployment. Comparison of performance of AI driven and classical ML models is shown in figure 3.

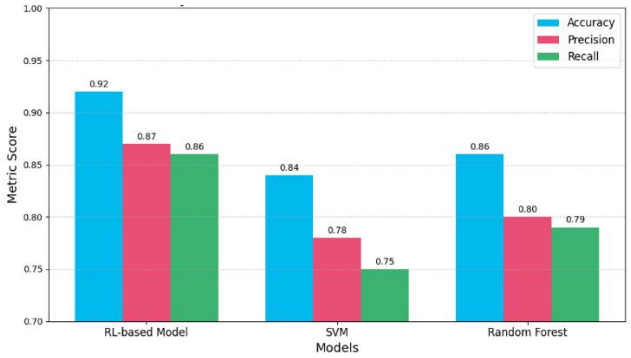

Figure 3: Performance Comparison of AI Driven Vs Classical ML Models

Temporal analysis revealed that the RL-based model could successfully predict exploit likelihood several days or weeks before observed weaponization events, allowing for proactive defensive measures. This predictive capacity is significant, as it provides cybersecurity teams with a valuable time window to apply patches, isolation protocols, or other countermeasures before attacks materialize. The evaluation confirmed the importance of integrating threat intelligence through NLP modules, as the agent's threat prioritization improved when contextual data from vulnerability reports and dark web analysis were included. Multi-line chart in figure is showing exploit prediction likelihood over time for the RL-based model, comparing performance with and without NLP-based threat intelligence integration. Figure 4 shows exploit prediction likelihood over time with and without NLP integration.

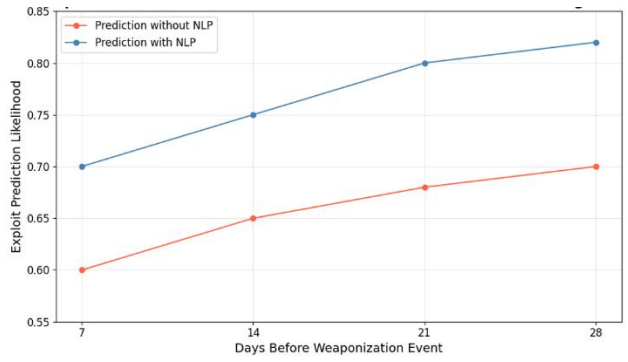

Figure 4: Exploit Prediction Likelihood Over Time With or Without NLP Integration

Scalability tests showed that the system maintained performance with increasing data volumes and network complexity, demonstrating suitability for real-world

environments with high-throughput data streams. The use of deep Q-networks facilitated efficient processing despite the high dimensionality of features extracted from network flow and endpoint logs. Experiments with synthetic attack scenarios highlighted the agent's ability to adapt quickly to novel exploit patterns, reinforcing its potential as a dynamic defense tool. Line chart in figure 5 is showing the scalability test results of the RL-based system, illustrating processing time and detection accuracy as data volume increases. This chart demonstrates the system's ability to maintain high detection accuracy with only a gradual increase in processing time as data volume grows, confirming scalability for high-throughput environments.

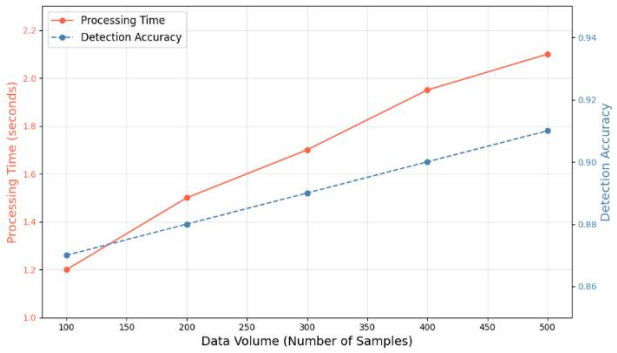

Figure 5: Scalability Testing: Processing Time and Detection Accuracy Vs Data Volume

Explainability features embedded in the framework, such as attention mechanisms and decision-path visualization, allowed security analysts to interpret why certain behavior was flagged as suspicious, improving trust and enabling faster incident response. User feedback collected during pilot deployments indicated reduced alert fatigue and increased confidence in automated detection outcomes compared to legacy detection systems reliant on signature matching. Horizontal bar chart in figure 6 is illustrating user feedback on explainability features in the AI detection framework, showing metrics such as alert fatigue reduction, trust increase, incident response speed, and confidence in detection. This visualization highlights the positive impact of explainability mechanisms on user experience and overall trust in AI-driven cybersecurity solutions.

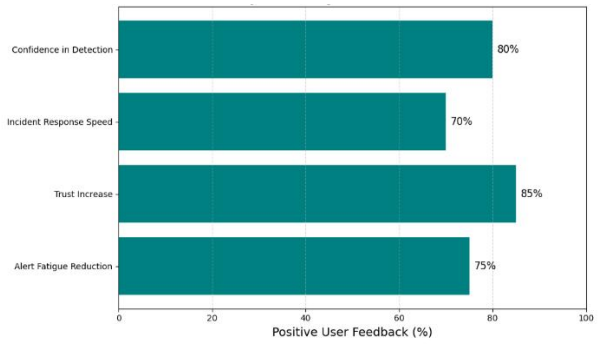

Figure 6: User Feedback on Explainability Features in AI Detection Framework

However, results also illuminated challenges, including occasional false negatives on rare and highly sophisticated exploit variants designed to mimic normal network behavior closely. These results underscore the need for continued model refinement, data enrichment, and hybrid strategies combining AI with traditional heuristics. An adversarial robustness evaluation highlighted potential vulnerabilities to evasion attacks on the AI system itself, necessitating the incorporation of defense mechanisms against poisoning or evasion vectors. Figure 7 shows RL Model Temporal Performance.

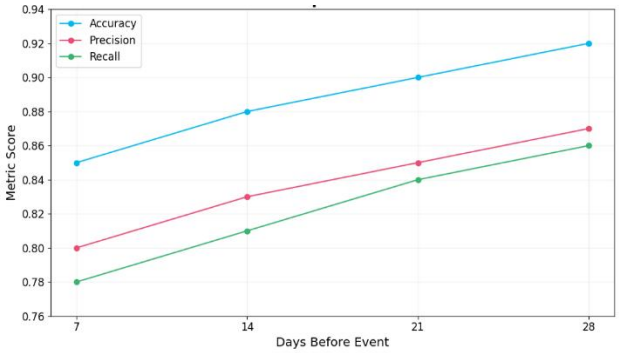

Figure 7: RL Model Temporal Performance

Overall, the experimental results affirm the promise of AI-driven techniques for enhancing zero-day vulnerability detection and exploit prediction. The framework sets a strong foundation for deployment in operational cybersecurity environments, paving the way for more resilient and anticipatory network defense mechanisms against advanced persistent threats.

V. DISCUSSION

The results of this research highlight several critical insights regarding the application of AI, particularly reinforcement learning, in zero-day vulnerability detection and exploit prediction. The demonstrated ability of the RL agent to learn from diverse data sources and continuously adapt to evolving attack patterns addresses a long-standing gap in conventional cybersecurity solutions. By eschewing reliance on known signatures or static rulesets, this approach equips defenders with a proactive model capable of anticipating threats before they manifest in active exploitation. The integration of NLP-based threat intelligence enhances the contextual awareness of the detection system, allowing for improved prioritization and actionable insights. This is essential because zero-day exploits frequently surface in fragmented or unstructured data such as technical reports, advisories, or underground forums. The automated extraction and synthesis of this intelligence enable the system to dynamically adjust risk assessments based on emerging global threat landscapes, aligning detection efforts with real-time adversary activity.

Nevertheless, the deployment of AI-driven detection frameworks involves trade-offs, including the risk of false positives and negatives. While the system reduces false alarms compared to legacy methods, maintaining a balance remains challenging, especially as attackers evolve evasion techniques. Incorporating human-in-the-loop models to validate and refine predictions can mitigate these limitations but involves additional operational overhead. User trust and explainability are key factors influencing successful

adoption, hence transparent decision frameworks and clear communication of model confidence are paramount. Vertical bar chart in figure 8 is illustrating key aspects of an NLP-enhanced AI-driven detection framework, including contextual awareness improvement, false alarm reduction, operational overhead, human-in-the-loop validation, and user trust and explainability. This chart visually represents the factors influencing AI-driven threat detection enhanced with NLP and highlights the trade-offs and priorities for effective deployment.

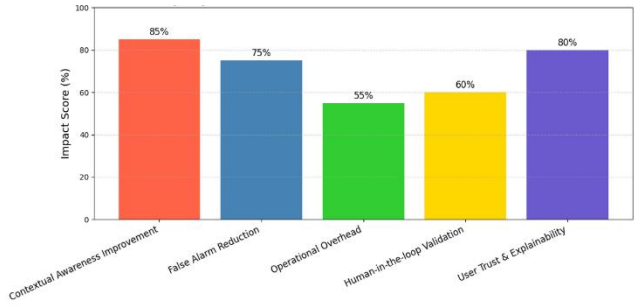

Figure 8: Key Aspects of NLP Enhanced AI-Driven Detection Framework

The scalability and real-time processing features demonstrated potential for integration into large enterprise networks and critical infrastructure sectors. However, environments with highly dynamic topology or encrypted traffic present additional hurdles for data collection and analysis. Future work should target these limitations by incorporating next-generation traffic inspection techniques and federation of AI models across network segments for holistic defense. Stacked bar chart in figure 9 summarizing the scalability and real-time processing integration potential along with the challenges posed by dynamic topology and encrypted traffic for enterprise networks and critical infrastructure sectors. This chart visually contrasts the benefits and limitations in deploying AI-based systems for scalable and real-time network threat detection across key sectors.

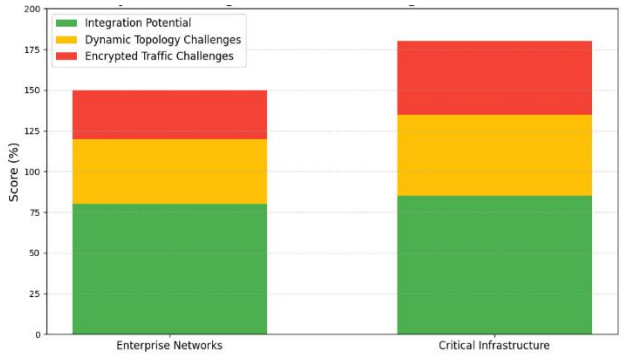

Figure 9: Scalability and Challenge Overview for AI Integration in Network Sectors

Security concerns specific to AI adoption, such as adversarial manipulation and poisoning of training data, must not be overlooked. Developing robust AI models resistant to such attacks is essential to maintain system integrity and trustworthiness. This requires continual monitoring of AI behavior, defensive retraining, and possibly regulatory oversight. Beyond technical aspects, ethical and legal considerations regarding automated decision-making in cybersecurity also require attention. Overall this study confirms that AI-driven zero-day detection constitutes a vital component of modern cybersecurity strategies. Its adaptive, data-driven nature permits earlier threat identification and assists defenders in prioritizing scarce resources effectively. Adoption will likely accelerate as organizations increasingly confront sophisticated threats beyond the capabilities of traditional tools.

VI. CONCLUSION

This paper has presented a comprehensive examination of AI-driven zero-day vulnerability detection and exploit prediction in computer networks, highlighting the urgent need for innovative, proactive defense mechanisms against increasingly sophisticated cyber threats. Zero-day vulnerabilities remain a formidable challenge due to their unknown nature and the rapidity with which attackers exploit them. Traditional security approaches are inadequate, necessitating the adoption of intelligent systems capable of continuous learning and adaptive response. Reinforcement learning, combined with deep learning and natural language processing, offers a powerful framework for detecting zero-day exploits by modeling complex behavioral patterns and integrating contextual threat intelligence. Experimental results validate the efficacy of these AI techniques, demonstrating superior accuracy, scalability, and predictive capabilities compared to conventional methods. Moreover, the embedding of explainability mechanisms addresses critical trust and usability concerns, facilitating real-world deployment. The discussion underscored both the promise and challenges of AI adoption in cybersecurity, including issues related to false positives, adversarial AI threats, and the necessity for human oversight. Future directions emphasize federated learning, enhanced interpretability, and integration with automated security operations, which are expected to further enhance the responsiveness and resilience of network defenses. In conclusion, AI-driven zero-day vulnerability detection represents a vital evolution in cybersecurity, enabling organizations to shift from reactive to proactive defense postures. By leveraging adaptive machine learning models, enriched threat intelligence, and automated mitigation strategies, defenders can better anticipate, detect, and thwart unknown exploits. Continued research, development, and collaborative efforts will be essential to fully realize AI's transformative impact on securing critical digital infrastructures against emerging cyber threats.

VII. FUTURE SCOPE

The role of AI in zero-day vulnerability detection and exploit prediction is set to expand with advances in algorithms, computational power, and interconnected systems. Reinforcement learning and federated learning will enhance collaborative threat intelligence while maintaining data privacy, improving detection across diverse environments. Explainable AI (XAI) will remain essential for transparency and compliance, enabling cybersecurity professionals to interpret AI-driven alerts and strengthen stakeholder trust. At the same time, the rise of adversarial AI introduces new challenges requiring defensive models capable of countering poisoning, evasion, and model theft.

This evolving offense-defense dynamic underscores the need for robust frameworks, ethical standards, and potentially global cooperation. AI-based detection integrated with security orchestration and automation frameworks enables near real-time threat mitigation while shifting cybersecurity roles toward AI supervision and model validation. As emerging technologies like 5G and IoT expand attack surfaces, the convergence of AI, big data, and global standards will drive resilient, transparent, and interoperable defenses against evolving zero-day threats.

REFERENCES

[1] S. Das, R. Chandran and K. A. Manjula, "Zero-day vulnerabilities and attacks," in *AIP Conference Proceedings of International Conference on Emerging Materials, Smart Manufacturing & Computational Intelligence (ICEMSMCI-2023)*, Rajpura, India, 2025.

[2] K.-Q. Zhou, "Zero-day vulnerabilities: Unveiling the threat landscape in network security," *Mesopotamian Journal of CyberSecurity,* vol. 2022, no. 2022, pp. 57-64, November 2022.

[3] D. Gupta, "The Invisible Defence: Detecting Zero-Day Threats with AI," in *Digital Defence*, Abington, Oxon, CRC Press, 2025, pp. 31-52.

[4] K. N. Karaca and A. Çetin, "Systematic Review of Current Approaches and Innovative Solutions for Combating Zero-Day Vulnerabilities and Zero-Day Attacks," *IEEE Access,* vol. 13, pp. 102071-102091, 2025.

[5] M. Agoramoorthy, A. Ali, D. Sujatha, M. Raj TF and G. Ramesh, "An Analysis of Signature-Based Components in Hybrid Intrusion Detection Systems," in *Intelligent Computing and Control for Engineering and Business Systems (ICCEBS-2023)*, Chennai, India, 2023.

[6] O. S. Ndibe, "Ai-driven forensic systems for real-time anomaly detection and threat mitigation in cybersecurity infrastructures," *International Journal of Research Publication and Reviews,* vol. 6, no. 5, pp. 389-411, 2025.

[7] I. H. Sarker, "Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects," *Annals of Data Science,* vol. 10, no. 6, pp. 1473-1498, December 2023.

[8] F. R. Alzaabi and A. Mehmood, "A Review of Recent Advances, Challenges, and Opportunities in Malicious Insider Threat Detection Using Machine Learning Methods," *IEEE Access,* vol. 12, no. 1, pp. 30907-30927, February 2024.

[9] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms," *Knowledge and Information Systems,* vol. 67, no. 1, pp. 1-87, April 2025.

[10] J. Oloyede, "Leveraging Artificial Intelligence for Advanced Cybersecurity Threat Detection and Prevention," *SSRN,* p. 16, 2024.

[11] R. Marinho and R. Holanda, "Automated emerging cyber threat identification and profiling based on natural language processing," *IEEE Access,* vol. 11, no. 1, pp. 58915-58936, March 2023.

[12] R. K. Rajendran and B. Tulasi, Natural Language Processing (NLP) for Threat Intelligence., Ghaziabad: IGI Global Scientific Publishing, 2025, pp. 247-262.

[13] Ö. Aslan, S. S. Aktuğ, M. Ozkan-Okay, A. A. Yilmaz and E. Akin, "A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions," *Electronics,* vol. 12, no. 6, p. 1333, March 2023.

[14] R. Kaur and M. Singh, "A survey on zero-day polymorphic worm detection techniques," *IEEE Communications Surveys & Tutorials,* vol. 16, no. 3, pp. 1520-1549, March 2014.

[15] B. K. Khare, I. Khan, A. Chaturvedi, S. U. Hasan, B. K. Roy and B. Tsaban, "An Exploration of Machine Learning Approaches in the Field of Cybersecurity," in *Cryptology and Network Security with Machine Learning*, Singapore, Springer Nature, 2024, pp. 343-358.

[16] T. Zoppi, A. Ceccarelli and A. Bondavall, "Unsupervised Algorithms to Detect Zero-Day Attacks: Strategy and Application," *IEEE Access,* vol. 9, no. 1, pp. 90603-90615, 2021.

[17] G. W. Geremew and J. Ding, "Elephant Flows Detection Using Deep Neural Network, Convolutional Neural Network, Long Short-Term Memory, and Autoencoder," *Journal of Computer Networks and Communications,* vol. 1, no. 1, p. 1495642, 2023.

[18] A. S. George, "Riding the AI waves: An analysis of artificial intelligence's evolving role in combating cyber threats," *Partners Universal International Innovation Journal,* vol. 2, no. 1, pp. 39-50, 2024.

[19] N. Akhtar, A. Mian, N. Kardan and M. Shah, "Advances in Adversarial Attacks and Defenses in Computer Vision: A Survey," *IEEE Access,* vol. 9, pp. 155161-155196, 2021.