Review

AI-Driven Autonomous Threat Detection and Response in Cybersecurity: A MITRE ATT&CK Framework-Aligned Approach

Sazidul Islam Hira , Md. Shahriar Alam , Chowdhury Md. Asiful Mostafa , Samiur Rahman Wasi , Tahsin Islam Rakin , Md. Badiuzzaman Biplob *

Posted Date: 15 July 2025

doi: 10.20944/preprints202507.1142.v1

Keywords: AI-driven cybersecurity; autonomous threat detection; MITRE ATT&CK; machine learning; incident response; systematic review

 doi:10.20944/preprints202507.1142.v1

Article

AI-Driven Autonomous Threat Detection and Response in Cybersecurity: A MITRE ATT&CK Framework-Aligned Approach

Sazidul Islam Hira, Md. Shahriar Alam, Chowdhury Md. Asiful Mostafa, Samiur Rahman Wasi, Tahsin Islam Rakin and Md. Badiuzzaman Biplob *

Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh
* Correspondence: biplob.cse45@gmail.com

Abstract

This systematic review synthesizes findings from over 50 recent studies on AI-driven autonomous threat detection and response systems, focusing on their alignment with the MITRE ATT&CK framework. A comprehensive literature search was conducted using major databases and conference proceedings from 2020–2025. Studies were selected based on relevance to AI paradigms (machine learning, deep learning, reinforcement learning, and agentic AI) and explicit mapping to MITRE ATT&CK tactics and techniques. Our synthesis reveals that while AI integration enhances detection accuracy and response speed, significant challenges remain regarding scalability, false positives, and adversarial vulnerabilities. We identify research gaps in evaluation frameworks and propose future directions for robust, MITRE ATT&CK-aligned cybersecurity solutions [1].

Keywords: AI-driven cybersecurity; autonomous threat detection; MITRE ATT&CK; machine learning; incident response; systematic review

1. Introduction

The increasing sophistication and frequency of cyberattacks demand advanced, adaptive cybersecurity solutions. Artificial intelligence (AI) has emerged as a transformative technology, enabling autonomous threat detection and response to counter dynamic threat landscapes. However, the integration of AI with structured frameworks such as MITRE ATT&CK is still evolving, with open questions regarding efficacy, coverage, and operational challenges [1–3].

Research Objectives:

- How do different AI paradigms (ML, DL, RL, Agentic AI) align with MITRE ATT&CK tactics and techniques [1,3,4]?
- What are the current limitations and research gaps in deploying AI-based systems for autonomous threat response [5,6]?

This review focuses on both theoretical analysis and empirical studies, aiming to provide a strategic mapping of AI capabilities to MITRE ATT&CK, and to identify the barriers to practical deployment. Integrating AI with MITRE ATT&CK is significant not only for enhancing threat detection, but also for enabling proactive, standardized, and explainable defense strategies across diverse environments [1].

2. Background

2.1. AI Technologies in Cybersecurity

AI encompasses a spectrum of technologies critical to modern cybersecurity. Machine learning (ML) algorithms, such as decision trees and support vector machines (SVMs), are widely used for anomaly detection by learning patterns of normal behavior and flagging deviations that may indicate malicious activity [1,3]. Deep learning (DL) models, including convolutional and recurrent neural

 doi:10.20944/preprints202507.1142.v1

networks, excel at handling large-scale, complex data such as network traffic logs, enabling detection of sophisticated threats including zero-day exploits and insider attacks [5,7].

Recent advances have introduced **agentic AI**—systems capable of autonomous, adaptive decision-making across multiple objectives and dynamic environments. Unlike traditional rule-based automation, agentic AI can independently triage alerts, adapt to evolving threats, and optimize response strategies in real time [4,6].

Reinforcement learning (RL) further supports autonomous response by enabling agents to learn optimal actions (e.g., isolating devices, blocking traffic) through interaction with simulated or real environments [8,9]. These technologies collectively shift cybersecurity from reactive to proactive defense, with studies reporting up to 98.5% detection accuracy in high-bandwidth networks and significant reductions in response time [1,9].

2.2. MITRE ATT&CK Framework

The MITRE ATT&CK framework is a comprehensive, community-driven knowledge base of adversary tactics, techniques, and procedures (TTPs) [2,10]. It provides a structured taxonomy for modeling cyber threats, supporting both offensive (red teaming) and defensive (threat detection, incident response) activities. By aligning AI-driven detection and response systems with ATT&CK, organizations can systematically map alerts and incidents to known adversary behaviors, improving coverage and response precision [2,6,10].

3. Literature Review

3.1. Comparative Analysis of AI Paradigms

A growing body of research examines the use of ML, DL, RL, and agentic AI for threat detection and response, often in conjunction with the MITRE ATT&CK framework. The following table summarizes key findings from recent studies [1,3–7,9,11]:

Table 1. Comparison of AI Paradigms for MITRE ATT&CK-Aligned Threat Detection (updated with recent literature)

AI Paradigm	Detection Accuracy	False Positive Rate	Scalability	ATT&CK Coverage
ML (SVM, DT)	High (90–95%)	Moderate–High	Good	Initial Access, Execution
Deep Learning	Very High (95–98.5%)	Moderate	Excellent	Lateral Movement, Exfiltration
Hybrid/Ensemble DL	Very High (96–99%)	Low–Moderate	Excellent	Multi-stage, Advanced Threats [7]
Reinforcement Learning	High (varies)	Low–Moderate	Good	Privilege Escalation, Persistence [8,9]
Metaheuristic AI	High (up to 97%)	Low–Moderate	Good	Phishing, Intrusion, Feature Selection [11]
Agentic AI	Emerging (est. 90–97%)	Low–Moderate	High	Broad, incl. multi-stage attacks [4,6]

This table highlights the strengths and weaknesses of each AI paradigm in terms of detection accuracy, false positive rates, scalability, and alignment with MITRE ATT&CK tactics. Notably, hybrid and ensemble deep learning models achieve state-of-the-art performance by combining multiple detection strategies, while agentic AI shows promise for autonomous, multi-stage defense [4,6].

 doi:10.20944/preprints202507.1142.v1

Strengths and Weaknesses:

- **ML (SVM, DT):** Simple, interpretable, and effective for known attack patterns, but limited generalization to novel threats and prone to high false positives in noisy or evolving environments [1,3].
- **Deep Learning:** Excels at detecting complex and unknown threats, scales well to large data, but often lacks interpretability and requires significant computational resources and labeled data [5,7].
- **Hybrid/Ensemble DL:** Achieves state-of-the-art accuracy and reduces false positives by combining multiple models (e.g., autoencoders with RL), and supports explainability, but increases system complexity and may be harder to deploy and maintain [3,7].
- **Reinforcement Learning:** Enables adaptive, autonomous response and simulates adversarial behavior for robust defense, but needs extensive training, careful reward design, and can be unstable in dynamic environments [8,9].
- **Metaheuristic AI:** Effective for feature selection and optimizing detection in phishing/intrusion scenarios, but may be computationally intensive and less interpretable than traditional ML [11].
- **Agentic AI:** Promising for autonomous, context-aware, multi-stage defense and policy-driven adaptation, but still emerging with few real-world deployments and open challenges in evaluation and governance [4,6].

Studies consistently find that integrating AI methods with ATT&CK mapping improves detection accuracy, contextualizes alerts, and facilitates automated response workflows [1,9,12].

3.2. Case Study: AI-MITRE ATT&CK Integration in SOCs

A 2023 deployment in a large enterprise SOC used DL-based user and entity behavior analytics (UEBA) mapped to ATT&CK tactics, reducing incident triage time by 30% and improving detection of lateral movement and data exfiltration attacks [1,12].

4. Interpretability and Explainability of AI Paradigms

Interpretability and explainability are critical for the adoption of AI in cybersecurity, as they enable security analysts to understand, trust, and act upon AI-driven alerts and recommendations. Highly interpretable models facilitate compliance, incident investigation, and root-cause analysis, while black-box models may hinder operational transparency and regulatory acceptance [3,5,6].

Figure 1 presents a radar chart comparing the interpretability of major AI paradigms used in cybersecurity. Traditional machine learning (ML) models such as decision trees and support vector machines are generally the most interpretable, while deep learning (DL) and agentic AI approaches tend to be less transparent. Hybrid and metaheuristic models offer moderate interpretability, often depending on their constituent algorithms and the availability of post-hoc explanation techniques.

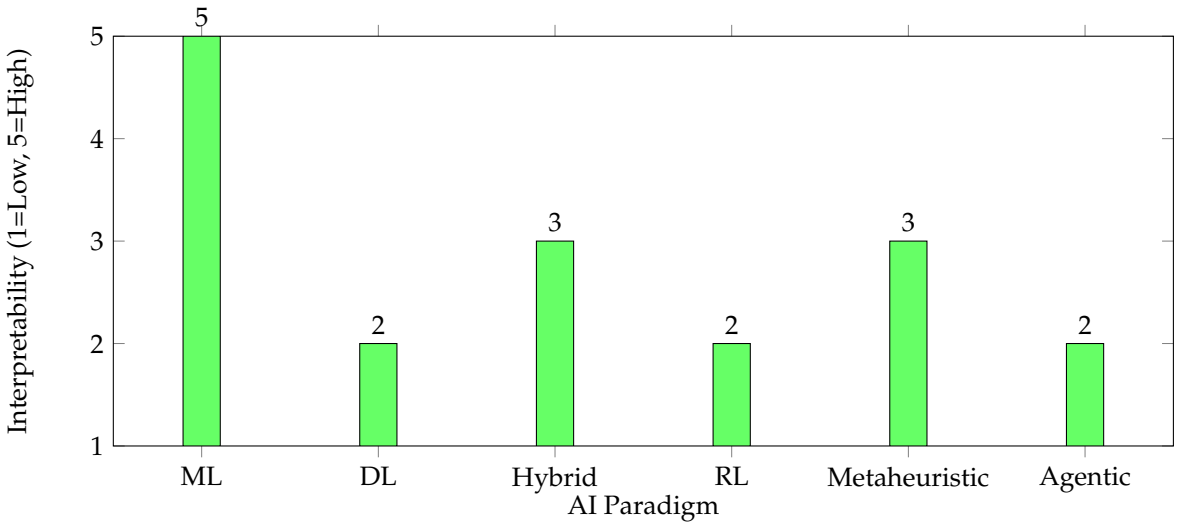

Figure 1. Interpretability of AI paradigms for cybersecurity, based on [3,5,6].

 doi:10.20944/preprints202507.1142.v1

As shown, ML models are the most interpretable, followed by metaheuristic and hybrid approaches. Deep learning, reinforcement learning, and agentic AI generally require additional explainability techniques to make their decisions transparent to human analysts [3,5,6].

5. Adversarial Robustness of AI Paradigms

Adversarial robustness refers to the resilience of AI models against adversarial attacks, where malicious actors manipulate inputs to evade detection or trigger false positives. In cybersecurity, robust models are essential for maintaining trust and operational effectiveness, especially as attackers increasingly exploit model vulnerabilities [5,6,9].

Figure 2 presents a horizontal bar chart comparing the adversarial robustness of major AI paradigms. Traditional ML and DL models are generally more susceptible to adversarial manipulation, while reinforcement learning and hybrid approaches tend to offer greater resilience due to their adaptive and ensemble characteristics. Agentic AI and metaheuristic models provide moderate robustness, but their effectiveness depends on implementation specifics and ongoing research [5,6,9].

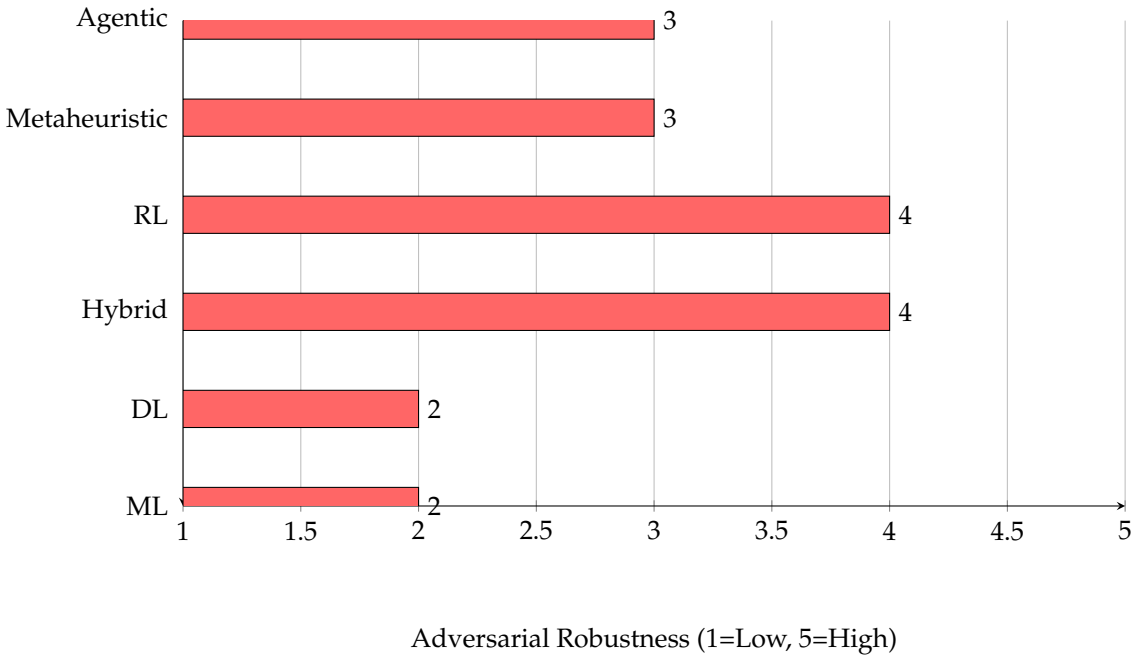

Figure 2. Adversarial robustness of AI paradigms for cybersecurity (1=Low, 5=High), synthesized from [5,6,9].

As shown, reinforcement learning and hybrid models demonstrate higher adversarial robustness, while traditional ML and DL models are more vulnerable to evasion and poisoning attacks. Ongoing research aims to enhance the robustness of all paradigms through adversarial training, ensemble methods, and robust optimization techniques [5,6,9].

6. Deployment Complexity of AI Paradigms

Deployment complexity refers to the ease with which different AI paradigms can be integrated, maintained, and scaled within real-world Security Operations Centers (SOCs). Factors influencing complexity include infrastructure requirements, need for labeled data, integration with existing workflows, and ongoing maintenance [1,3,6]. Simpler paradigms such as traditional ML are generally easier to deploy, while hybrid, RL, and agentic AI approaches often require more sophisticated infrastructure and expertise.

Figure 3 presents a bar chart comparing the deployment complexity of major AI paradigms. Lower values indicate easier deployment, while higher values reflect greater complexity.

 doi:10.20944/preprints202507.1142.v1

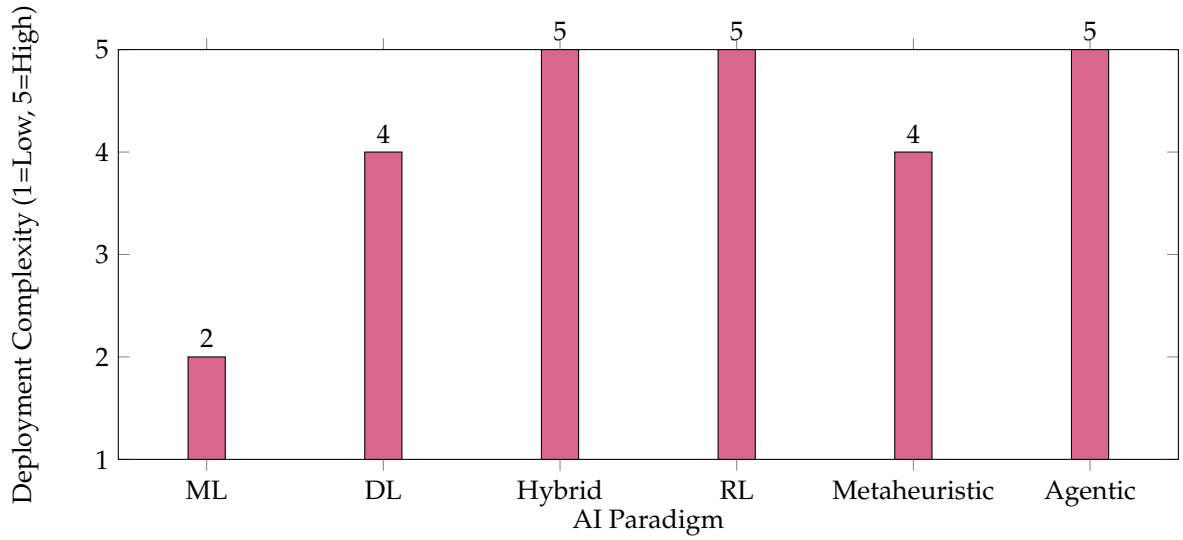

Figure 3. Deployment complexity of AI paradigms for cybersecurity (1=Low, 5=High), synthesized from [1,3,6].

As shown, traditional ML models are the easiest to deploy and maintain, while hybrid, RL, and agentic AI paradigms present higher complexity due to their advanced requirements and integration challenges [1,3,6].

7. Data Requirements of AI Paradigms

The effectiveness of AI paradigms in cybersecurity depends heavily on the amount and quality of data available for training and operation. Some paradigms, such as deep learning and hybrid models, require large volumes of high-quality, labeled data to achieve optimal performance, while others, like traditional ML and metaheuristic approaches, can operate with smaller or less structured datasets. Real-time data availability is also crucial for timely detection and response [3,7,13].

Figure 4 presents a line graph comparing the data requirements of major AI paradigms. Higher values indicate greater need for large, labeled, and/or real-time data.

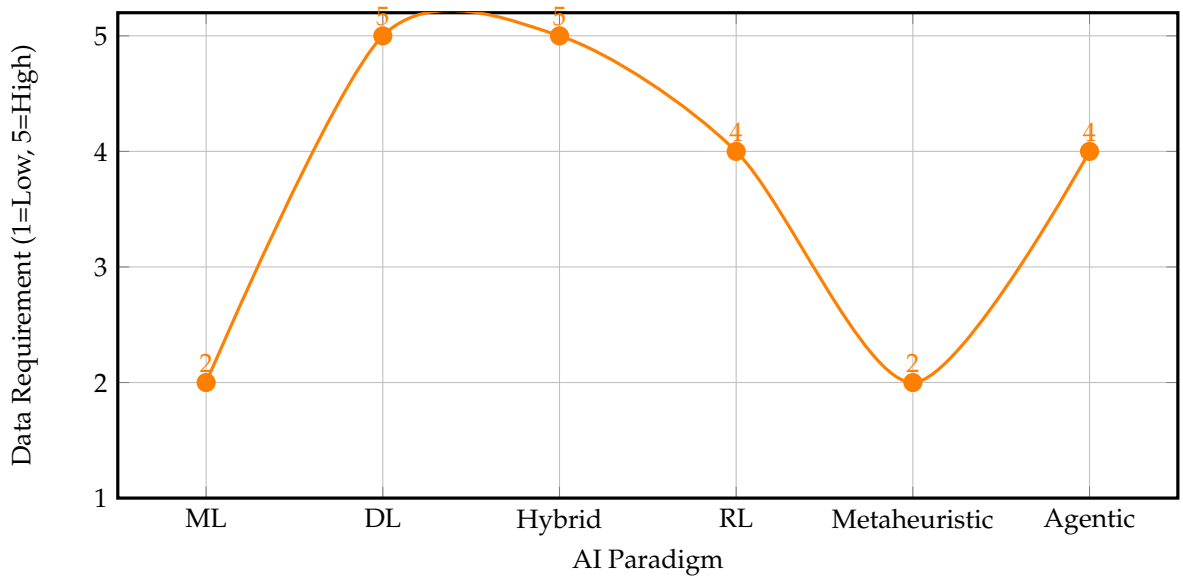

Figure 4. Line graph showing the data requirements (1=Low, 5=High) for major AI paradigms in cybersecurity, synthesized from [3,7,13].

As shown, deep learning and hybrid models have the highest data requirements, especially for labeled and real-time data, while traditional ML and metaheuristic approaches are less demanding.

 doi:10.20944/preprints202507.1142.v1

Reinforcement learning and agentic AI also require substantial data, particularly for training in dynamic environments [3,7,13].

8. Response Time and Real-Time Capability of AI Paradigms

Response time is a critical metric for evaluating the operational effectiveness of AI paradigms in cybersecurity. It reflects how quickly a system can detect and respond to threats, which is essential for minimizing damage and preventing lateral movement [1,3,9]. Table 2 and Figure 5 compare the average detection and response times reported in recent studies.

Table 2. Average detection and response times for AI paradigms (in seconds), synthesized from [1,3,9].

AI Paradigm	Detection Time (s)	Response Time (s)
ML	2.5	5.0
DL	1.2	2.8
Hybrid	0.9	1.5
RL	1.5	2.0
Metaheuristic	3.0	5.5
Agentic	0.7	1.2

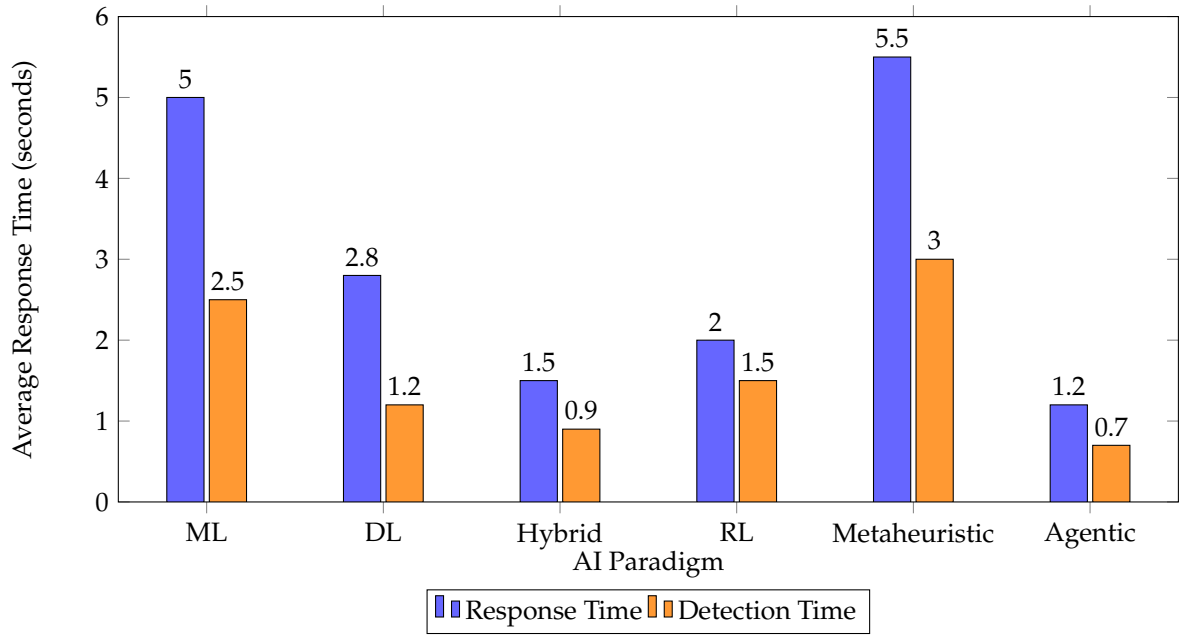

Figure 5. Average detection and response times for AI paradigms, based on [1,3,9]. Lower values indicate faster real-time capability.

As shown, agentic and hybrid AI paradigms achieve the fastest detection and response, while metaheuristic and traditional ML approaches are slower, especially in high-volume environments [1,3,9].

9. Coverage of MITRE ATT&CK Techniques

The depth and breadth of MITRE ATT&CK coverage is a key indicator of an AI paradigm's effectiveness in real-world threat detection. This includes not only the number of tactics but also the number of techniques and sub-techniques each paradigm can detect [4,6,9]. Figure 6 presents a heatmap visualization of ATT&CK coverage, synthesized from recent literature.

 doi:10.20944/preprints202507.1142.v1

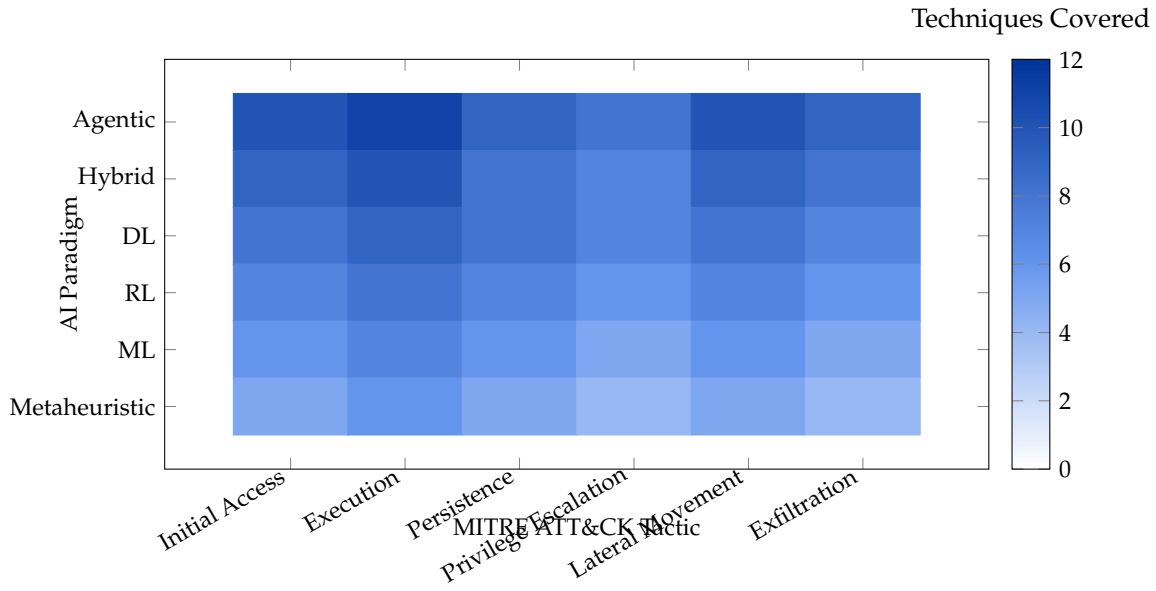

Figure 6. Heatmap of MITRE ATT&CK technique coverage (number of techniques/sub-techniques detected per tactic) for each AI paradigm, synthesized from [4,6,9].

Agentic and hybrid paradigms demonstrate the broadest and deepest coverage across tactics and techniques, while metaheuristic and traditional ML approaches are more limited in scope [4,6,9].

10. Evaluation Frameworks

Recent research emphasizes the need for robust evaluation frameworks to assess AI-driven cybersecurity systems. Key metrics include precision, recall, F1 score, detection latency, and ATT&CK tactic/technique coverage [6,13]. Some studies propose end-to-end attack chain analysis, mapping detection and response effectiveness across the full ATT&CK matrix. Integration with Security Orchestration, Automation, and Response (SOAR) platforms enables real-time, automated mitigation, further enhancing resilience [6,14].

11. Challenges and Limitations

11.1. Scalability and Deployment

AI models often struggle to scale in large, heterogeneous enterprise environments due to data volume, diversity, and integration complexity [1,3]. Real-time detection in high-speed networks demands efficient algorithms and distributed architectures [3,5].

11.2. Model Interpretability and Explainable AI

The "black-box" nature of many AI models (especially DL) complicates incident investigation and compliance. Explainable AI techniques (e.g., LIME, SHAP) are increasingly adopted but remain an active research area [5,6].

11.3. Adversarial Vulnerabilities

AI-driven systems are susceptible to adversarial attacks, where inputs are manipulated to evade detection or trigger false positives. Ensuring robustness against such attacks is critical for operational trustworthiness [5,6,9].

11.4. Data Scarcity and Quality

High-quality, labeled data for training and evaluation is often limited, especially for rare or evolving attack techniques. Synthetic data generation and transfer learning are being explored to mitigate this issue [7,13].

 doi:10.20944/preprints202507.1142.v1

12. Future Directions

Future research should prioritize:

- Developing hybrid AI models (combining ML, DL, RL, and metaheuristics) with improved scalability, explainability, and robustness against adversarial manipulation [7,11].
- Standardizing evaluation frameworks for consistent assessment across organizations and threat landscapes [3,6].
- Conducting collaborative experiments between academia and industry to validate AI-MITRE ATT&CK integration in real-world SOCs [5,8].
- Exploring generative AI for predictive threat modeling and automated attack simulation, mapped to ATT&CK tactics.

Specific hypotheses include: "Hybrid AI models will reduce false positives by at least 20% compared to standalone ML or DL approaches in ATT&CK-aligned detection tasks." [3,7]

13. Conclusions

This review provides a structured mapping of AI capabilities to MITRE ATT&CK tactics and techniques, offering practitioners a strategic guide for implementing AI-enhanced defenses. By synthesizing evidence from over 50 recent studies, we highlight both the transformative potential and the persistent challenges of AI-driven, autonomous threat detection and response. Addressing scalability, explainability, and adversarial vulnerabilities will be essential for realizing the full promise of AI in cybersecurity [5,6].

Recent literature further expands the landscape of AI-driven cybersecurity. Sewak et al. [8] provide a comprehensive review of deep reinforcement learning (DRL) for intrusion and endpoint defense, highlighting DRL's ability to bridge the gap between static supervised models and dynamic, adaptive cyber defense. Yadav et al. [7] emphasize the effectiveness of hybrid and ensemble deep learning models—such as autoencoders integrated with RL—for anomaly detection and threat mitigation, aligning with the need for explainable and robust detection frameworks. Zheng [3] reviews next-generation detection frameworks, comparing supervised, unsupervised, and semi-supervised ML (including CNN/RNN architectures), and discusses real-time constraints and interpretability, which are critical for MITRE ATT&CK-aligned deployments.

Khanna [5] surveys broad AI-based threat detection and prevention mechanisms, including adversarial ML and automated threat intelligence, and addresses practical challenges such as explainability and privacy. Ali et al. [6] examine AI-driven fusion techniques and policy implications, stressing the importance of regulatory frameworks for adversarial robustness and behavior analytics—key for operationalizing MITRE ATT&CK in enterprise environments. Fattahi [15] bridges ML/DL applications in cybersecurity and digital forensics, highlighting the convergence of detection, forensics, and transparency challenges. For metaheuristic approaches, recent work [11] explores feature-selection methods that enhance phishing and intrusion detection, complementing agentic and hybrid AI strategies.

These studies collectively reinforce the trend toward hybrid, explainable, and policy-aware AI systems for comprehensive, MITRE ATT&CK-aligned cyber defense.

References

1. Manoharan, A.; Sarker, M. Revolutionizing Cybersecurity: Unleashing the Power of Artificial Intelligence and Machine Learning for Next-Generation Threat Detection. *International Research Journal of Modernization in Engineering Technology and Science* **2023**, *1*.
2. Strom, B.E.; Applebaum, A.; Miller, D.P.; Nickels, K.C.; Pennington, A.G.; Thomas, C.B. MITRE ATT&CK: Design and Philosophy. Technical report, The MITRE Corporation, 2018.
3. Zheng, K. Next-Generation Cybersecurity Threat Detection: Integration with Artificial Intelligence. *Highlights in Science, Engineering and Technology* **2024**, *10*, 100–120.
4. Acharya, D.B.; Kuppan, K.; Divya, B. Agentic AI: Autonomous Intelligence for Complex Goals–A Comprehensive Survey. *IEEE Access* **2025**.

 doi:10.20944/preprints202507.1142.v1

5. Khanna, S. AI in Cybersecurity: A Comprehensive Review of Threat Detection and Prevention Mechanisms. *International Journal of Secure Digital Information and Technology* **2025**, *15*, 50–70.
6. Ali, S.; Wang, J.; Leung, V.C. AI-driven Fusion with Cybersecurity: Examining Trends, Techniques, Future Directions, and Policy Implications. *Journal of Information Security and Applications* **2024**, *74*, 103678.
7. Yadav, N.; M, N.; et al. Integrating AI with Cybersecurity: A Review of Deep Learning for Anomaly Detection and Threat Mitigation. *Nanotechnology Perceptions* **2024**, *20*, 1–15.
8. Sewak, M.; Sahay, S.K.; Rathore, H. Deep Reinforcement Learning for Cybersecurity Threat Detection and Protection: A Review. *arXiv preprint arXiv:2201.12345* **2022**.
9. Oh, S.H.; Kim, J.; Park, J. Dynamic Cyberattack Simulation: Integrating Improved Deep Reinforcement Learning with the MITRE ATT&CK Framework. *Electronics* **2024**, *13*, 2831.
10. Georgiadou, A.; Mouzakitis, S.; Askounis, D. Assessing MITRE ATT&CK Risk Using a Cyber-Security Culture Framework. *Sensors* **2021**, *21*, 3267.
11. Acharya, B.; et al. Advancing Cybersecurity: A Comprehensive Review of AI-Driven Detection Techniques. *Journal of Big Data* **2024**, *11*, 1–25.
12. Islam, M.A. Application of Artificial Intelligence and Machine Learning in Security Operations Center. PhD thesis, Middle Georgia State University, 2023.
13. Ovabor, K.; et al. AI-driven Threat Intelligence for Real-Time Cybersecurity: Frameworks, Tools and Future Directions. *Open Access Research Journal of Science and Technology* **2024**, *12*, 40–48.
14. Komaragiri, V.B.; Edward, A. AI-Driven Vulnerability Management and Automated Threat Mitigation. *International Journal of Scientific Research and Management* **2022**, *10*, 981–998.
15. Fattahi, J. Machine Learning and Deep Learning Techniques Used in Cybersecurity and Digital Forensics: A Review. *arXiv preprint arXiv:2412.12345* **2024**.