AI-Powered Threat Detection and Mitigation System: A Literature Survey

Authors:

A. MEENA (meenaananda2329@gmail.com)
ARHAM NOORAIN (arhamnoorain17@gmail.com)
SNEHA. S (sneha7114.4117@gmail.com)
SOUMYA U GANJI (soumyaug0432@gmail.com)

Guide:

Dr. Guruprasad YK Assoc. Prof. and Head
Dept. of CSE(CY) SVCE, Bengaluru

Abstract:

The rapidly evolving landscape of cyber threats, characterized by increasingly sophisticated Advanced Persistent Threats (APTs) and zero-day exploits, has exposed the critical limitations of traditional, signature-based defense mechanisms. These reactive systems struggle to identify novel attacks and adapt to the dynamic tactics of modern adversaries. This project, "AI-Powered Threat Detection and Mitigation System," aims to explore and develop a proactive defense framework that leverages Artificial Intelligence (AI) and Machine Learning (ML) to enhance threat detection accuracy, enable predictive capabilities, and facilitate automated mitigation responses.

By synthesizing findings from contemporary research, this survey investigates the application of core AI methodologies—including supervised and unsupervised learning, deep learning, and ensemble methods—within cybersecurity. The project seeks to design a system that moves beyond mere detection, integrating real-time analysis, threat intelligence extraction, and automated countermeasures into a cohesive unit. The ultimate goal is to contribute to a paradigm shift in cybersecurity, from a reactive to a proactive and adaptive posture, thereby strengthening organizational resilience against complex cyber threats.

1. Introduction

In today's digital landscape, organizations are undergoing rapid digitization, leading to unprecedented operational efficiencies and new revenue streams. However, this increased connectivity and reliance on sophisticated technology have also expanded the attack surface, attracting equally sophisticated cyber threats. Advanced Persistent Threats (APTs) represent a particularly formidable challenge, capable of operating covertly, mimicking legitimate users, and remaining undetected within networks for extended periods. Traditional security measures, such as rule-based Intrusion Detection Systems (IDS) and firewalls, which depend on static signatures and predefined rules, are increasingly inadequate against these stealthy and evolving attack methodologies.

This defensive shortfall has created an urgent need for more adaptive, intelligent, and proactive security strategies. Advances in Artificial Intelligence (AI) and Machine Learning (ML) present a transformative opportunity to bridge this gap. By leveraging big data analytics and sophisticated algorithms, AIpowered systems can process vast volumes of security telemetry in real-time, identify hidden attack patterns, predict potential threats, and automate response actions. Furthermore, the principle of threatinformed defense, which aligns security measures with known adversary Tactics, Techniques, and Procedures (TTPs) as outlined in frameworks like MITRE ATT&CK, enhances detection visibility and reduces false positives.

This literature survey, conducted under the project "AI-Powered Threat Detection and Mitigation System," synthesizes recent research to illustrate the promise of AI-driven, threat-informed security environments. It explores the core AI techniques being deployed, analyzes the capabilities and limitations of current approaches, and identifies research gaps to justify the development of a comprehensive, intelligent threat management system.

1.1 Problem Statement & Objectives

The modern cybersecurity paradigm is locked in an asymmetric battle against agile and resourceful adversaries. While organizations rely on traditional tools, attackers continuously develop new methods to bypass signature-based detection. APTs, zero-day vulnerabilities, and polymorphic malware can easily evade these legacy systems, leading to prolonged undetected breaches, significant data loss, financial damage, and reputational harm. The sheer volume and velocity of network data also overwhelm human analysts, causing alert fatigue and delayed responses.

While AI-powered solutions are emerging in the market, many are siloed, focusing either on detection *or* response, but not providing an integrated, end-to-end solution. Tools often lack explainability, creating a "black box" problem where security teams cannot understand the AI's decisions. Furthermore, many advanced AI tools are designed for experts, creating accessibility barriers for average organizations.

This project, titled “AI-Powered Threat Detection and Mitigation System,” seeks to address these limitations by proposing a system that integrates proactive threat detection, intelligent classification, and automated mitigation into a unified, user-accessible framework.

1.2 Objectives of the Study

The primary objectives of this literature survey and the subsequent project are as follows:

1. **To study and evaluate** existing AI/ML techniques, tools, and research related to cybersecurity threat detection, prediction, and mitigation.
2. **To explore the potential** of integrating various AI methodologies—such as anomaly detection,

supervised classification, and deep learning—into a cohesive defense system.

3. **To analyze existing AI-driven security solutions** and threat intelligence platforms in terms of their capabilities, limitations, and real-world applicability.

4. **To identify research gaps** in current solutions, particularly regarding explainability, automated response, and proactive threat hunting.

5. **To justify the need** for a holistic system that consolidates threat detection, risk assessment, and automated mitigation.

6. **To lay the foundation** for designing an effective, privacy-aware, and user-friendly solution that empowers organizations to take a proactive stance against cyber threats.

1.3 Methodology

The methodology for this literature survey was structured to identify, evaluate, and compare contemporary research, tools, and techniques relevant to AI in cybersecurity.

1.3.1 Research Approach

A qualitative research approach was followed, focusing on the critical analysis of peer-reviewed articles, conference papers, industry whitepapers, and technical documentation published primarily between 2020 and 2025. The selection was guided by relevance to the core themes of this project.

1.3.2 Source of Data

The following sources were used to gather literature:

- **Academic Databases:** IEEE Xplore, SpringerLink, ACM Digital Library, Google Scholar.
- **Security Forums & Whitepapers:** SANS Institute, OWASP, MITRE.
- **Industry Reports:** Publications from leading cybersecurity firms (e.g., Kaspersky Lab).
- **Open-source Tools Documentation:** Official repositories and guides for AI/ML frameworks.

1.3.3 Keyword Strategy

Keywords included: "AI in cybersecurity," "machine learning threat detection," "Advanced Persistent Threat (APT) detection," "anomaly detection," "predictive cyber analytics," "automated incident response," "Explainable AI (XAI) cybersecurity," and "threat-informed defense."

1.3.4 Selection Criteria Selection

was based on:

- **Relevance:** Direct connection to AI/ML in threat detection and mitigation.

- **Recency:** Preference for works published after 2020 to reflect the current threat landscape.
- **Credibility:** Peer-reviewed publications, reputable conference proceedings, and authoritative industry sources.

1.4 Literature Survey

The integration of AI into cybersecurity is a rapidly advancing field. This section presents a thematic review of existing research, tools, and studies that form the foundation for this project.

Source	Focus / Objective	AI Techniques Discussed	Key Contributions / Findings	Context / Application	Challenges / Limitations Mentioned
Mori (2023)	AI for cyber resilience in critical infrastructure (prediction, detection, recovery)	Predictive analytics (ML) , Real-time anomaly detection (Unsupervised ML - Autoencoders, Clustering) , Automated response, Self-healing systems, Reinforcement Learning (Risk Assessment)	AI enables proactive defense through prediction , addresses limitations of traditional methods , automates response/recovery , enhances resilience. Real-time anomaly detection crucial for zero-day threats.	Critical Infrastructure (Energy grids, Transport, Water treatment, Healthcare).	Data quality/availability , complexity (false positives, unintended consequences) , adversarial adaptation , explainability ("black box").
Pulyala (2024)	AI-powered SIEM for proactive threat hunting and risk mitigation	ML, DL , Pattern Recognition , Predictive Analytics , Behavioral Analysis , Automation , User & Event Behavioral Analytics (UEBA)	AI transforms SIEMs from data aggregation to intelligent analysis. AI enables prediction, faster detection/response, operational efficiency, reduced manpower, continuous improvement . AI also powers sophisticated attacks. Use cases: malware analysis, data analysis, ransomware detection, early detection, rapid response, vulnerability management .	SIEM systems, Threat Hunting, Risk Mitigation, Malware Analysis, Vulnerability Management .	AI can be weaponized by attackers. Addresses traditional SIEM issues (data volume, false positives).
Labu & Ahammed (2024)	Next-gen cyber threat detection/mitigation using AI/ML, focusing on fraud detection & algorithm comparison	ML, DL. Supervised (Random Forest) , Unsupervised , Reinforcement Learning . Feedzai system . Comparison: Naïve Bayes, KNN .	AI/ML address inadequacy of traditional methods. Provide proactive detection, real-time analysis, automation, anomaly detection, reduced false positives . Random Forest showed highest accuracy (83.94%) vs Naïve Bayes (79.23%) and KNN (78.74%) in experiment .	General Cybersecurity, Fraud Detection (Financial), Algorithm Comparison .	Dynamic nature of threats may make methods outdated.

Wang (2024)	General application of AI in cybersecurity threat detection; overview of core methods	ML (Supervised, Unsupervised, Reinforcement, Transfer) , DL (CNN, RNN/LSTM, Autoencoders, GANs, Transfer,	AI overcomes limitations of traditional methods against modern threats (APTs, zerodays) . Provides data analysis, pattern recognition, automation.	General Cybersecurity, Threat Detection, Malware Detection, Spam Filtering,	Data privacy/security , Adversarial attacks , Interpretability ("black box") & Compliance .
		GNN) , Integrated/Ensemble Learning , Multimodal approaches , Heterogeneous integration .	Improves accuracy, detects anomalies, adapts in real-time. Discusses specific uses of ML/DL types. Integrated/multimodal approachesimprove robustness.	Intrusion Detection Systems (IDS), Anomaly Detection .	
Aldawsari & Kouchay (2024)	Integrating AI/ML into cloud security frameworks for proactive threat detection/mitigation	AI/ML generally. Supervised, Unsupervised, Semi-supervised. DL (Autoencoders, RNNs). Predictive Analytics , Automated Incident Response , Selfhealing Systems. RL. NLP. Hybrid/Ensemble.	AI/ML needed for complex cloud threats. Provides accuracy, scalability, effectiveness in detecting anomalies. Improves detection, reduces response times, enhances security posture. Case studies show benefits. Benefits: accuracy, speed, scalability, adaptation, efficiency .	Cloud Security, Anomaly Detection, Intrusion Detection, Predictive Maintenance, Access Control, Performance Monitoring, Resource Management.	Need for robust/scalable models, model bias, ethics/legal (data protection). Data quality, Interpretability (XAI), Adversarial attacks, Integration complexity, Compliance, False positives .
Balantrapu (2024)	AI for predictive cyber threat intelligence (CTI); anticipating attacks	ML , NLP (unstructured data) , DL (pattern recognition) , RL (adaptive security) . Supervised (DT, RF, SVM), Unsupervised (clustering), Semisupervised. DL: CNN, RNN, Autoencoders.	Addresses limitations of traditional CTI (reactive, volume, false positives, context, effort) . Enables proactive prediction. Reduces false positives, enhances threat hunting. Use cases: real-time detection, predictive hunting, anomaly detection, vulnerability management/forecasting . Case studies discussed .	Cyber Threat Intelligence (CTI), Threat Prediction, Threat Hunting, Anomaly Detection, Vulnerability Management .	Data privacy/ethics , Algorithmic bias/transparency ("black box") , Technical/operational challenges (cost, integration, updates) , Skill gaps .
Venkoba et al. (2024)	General review of AI for threat detection/prevention; trends, challenges, future directions	ML (Supervised , Unsupervised , RL), DL (CNN, RNN/ConvLSTM) , GANs , Graphbased analysis. Bayesian Networks.	Addresses insufficiency of traditional methods against modern threats (zero-days, APTs) . Provides pattern recognition, automated response. Details ML types for known/novel/ adaptive threats. Details DL in malware detection. Covers predictive analysis & threat hunting.	General Cybersecurity, Threat Detection, Malware Identification, Predictive Analysis, Intrusion Detection, Threat Hunting.	Data quality/availability , False positives , Adversarial attacks , Generalization/Adaptation. Need for XAI. Computational cost (DL).

Niveditha et al. (2024)	Predictive analytics using AI for cyber threat intelligence	Predictive Analytics. ML (RF, SVM, Gradient Boosting). DL (CNN, RNN, LSTM). Supervised, Unsupervised, RL mentioned . Ensemble methods.	Addresses shortcomings of traditional measures. Enables preemptive identification/mitigation. High accuracy (92% precision). DL improves accuracy (15%). Reduced false positives (20%), increased timeliness (30%). Case studies showed significant	Cyber Threat Intelligence, Predictive Analytics, Threat Detection, APT Detection, Anomaly Detection, Phishing Mitigation .	Data quality/availability, interpretability, evolving threats. Data privacy/ethics.
			improvements . High model performance metrics reported .		
Kumar et al. (2023)	General overview of AI (specifically ML) in cybersecurity for threat detection and response	AI/ML generally. ML algorithms (Supervised, Unsupervised - KMeans, Reinforcement Learning) . Behavioral analysis. Behavioral Biometrics. UEBA.	AI/ML enhance detection/response efficiency. Analyze vast data, spot anomalies/patterns. Needed because traditional methods are inadequate. Enables realtime detection, faster response. Bolsters authentication. AI can also facilitate attacks. Benefits listed .	General Cybersecurity, Threat Detection, Threat Response, Intrusion Detection, Authentication, Access Control.	AI usable by attackers. False positives/negatives. Ethics (privacy, bias). Security risks of AI systems. Data privacy. Lack of skilled workforce.
LeBlanc (2025)	Detecting APTs using threat modeling and metadetections; comparison with RBA/UEBA	Discusses UEBA. Focuses on meta-detections (correlating TTPbased alerts), often enabled by AI/ML platforms. RiskBased Alerting (RBA). SIEM mentioned.	Proposes threat-informed detection using meta-detections for APTs. Meta-detections correlate alerts based on TTPs/kill-chain. Findings: Meta-detections did not improve speed/accuracy vs. RBA in short-term test, but provided better attribution . May be better for long dwell times. Implementation advice provided .	Advanced Persistent Threat (APT) Detection, Threat Modeling, Detection Engineering, Meta-Detection, RBA, SIEM, MITRE ATT&CK.	Meta-detection effectiveness may depend on longer dwell times. Requires mature security environment . Traditional/RBA can cause alert fatigue.
Our Solution (2025)	AI-driven threat detection and mitigation	Hybrid AI (Supervised, Unsupervised, RL), DL (CNN, RNN), NLP, Ensemble	Integrated PyQt6–Flask system enabling real-time detection, automated mitigation, and XAI-based transparency. Uses predictive analytics and CTI for adaptive defense.	Cybersecurity, Threat Detection, CTI, SIEM, Cloud Security	Reduces false positives, improves interpretability, ensures scalability, and enables proactive automated response.

1.4.1 Core AI Techniques in Threat Detection

Research consistently highlights a suite of AI techniques that are proving effective in modern threat detection.

- **Machine Learning (ML):** Wang (2024) and Venkoba et al. (2024) provide comprehensive overviews, noting that **supervised learning** (e.g., Random Forests, SVM) is widely used for classifying known threats like malware, while **unsupervised learning** (e.g., Isolation Forests, K-Means) is crucial for detecting novel attacks and anomalies without prior labeling. **Reinforcement Learning (RL)** is explored for adaptive security policies that evolve with the threat environment Mori (2023).

- **Deep Learning (DL):** As discussed by Wang (2024) and Niveditha et al. (2024), **CNNs** are effective for analyzing malicious code as images, and **RNNs/LSTMs** excel at detecting patterns in time-series data like network traffic logs. **Autoencoders** are prominently used for unsupervised anomaly detection by learning a model of "normal" behavior and flagging deviations.

- **Integrated and Multimodal Approaches:** Wang (2024) emphasizes that **ensemble methods** (e.g., combining Random Forest with Gradient Boosting) improve robustness and accuracy. **Multimodal deep learning**, which fuses diverse data sources (network traffic, logs, user behavior), provides a more holistic view for detecting complex threats like APTs.

1.4.2 AI for Proactive and Predictive Defense

A key theme in recent literature is the shift from reactive to proactive defense.

- **Predictive Analytics:** Balantrapu (2024) and Niveditha et al. (2024) focus on AI for predictive Cyber Threat Intelligence (CTI). By applying NLP to unstructured data and ML to historical attack patterns, AI can forecast potential vulnerabilities and attack campaigns, allowing for preemptive hardening of defenses.

- **Threat-Informed Defense:** LeBlanc (2025) discusses the importance of **meta-detections**—correlating alerts based on adversary TTPs from the MITRE ATT&CK framework. This approach moves beyond isolated alerts to uncover the broader narrative of an attack campaign.

1.4.3 Automated Mitigation and Response

The literature points to automation as a critical component for reducing the time between detection and response.

Mori (2023) and Aldawsari & Kouchay (2024) discuss "self-healing systems" and **Automated Incident Response** in cloud environments. These systems can automatically execute predefined actions such as blocking malicious IPs, isolating compromised endpoints, or terminating malicious processes, thereby containing threats before they can spread.

1.5 Identified Research Gaps

Despite the wealth of research and development in AI for cybersecurity, several critical gaps remain unaddressed by existing solutions.

1.5.1 Fragmented Application of AI

Most research and tools focus on specific aspects of the security lifecycle, such as standalone anomaly detection or breach notification. There is a **lack of a unified framework** that seamlessly integrates proactive prediction, accurate detection, explainable analysis, and automated mitigation into a single, cohesive system.

1.5.2 The "Black Box" Problem and Lack of Explainability

As noted by Wang (2024) and Balantrapu (2024), the complexity of many AI models, particularly deep learning, makes them opaque "black boxes." Security operators are often unable to understand *why* a model flagged a particular event, hindering trust, validation, and effective response. There is a significant gap in integrating **Explainable AI (XAI)** principles into operational threat detection systems.

1.5.3 Reactive vs. Proactive Posture

While tools for analyzing past breaches exist, and predictive analytics is a growing field, there are very few systems that actively use this intelligence to drive **automated, proactive mitigation** *before* a fullblown incident occurs. The link between prediction and automated action is often weak or non-existent. **1.5.4 Technical and Accessibility Barriers**

Many of the most powerful AI/ML tools for security (e.g., custom-built models in Python, complex SIEM queries) require significant expertise to deploy and manage. This creates a high barrier to entry for small and medium-sized enterprises and contributes to the cybersecurity skills gap. There is a distinct **lack of user-friendly, integrated platforms** that democratize AI-powered security.

1.5 Identified Research Gaps

Despite the wealth of research and development in AI for cybersecurity, several critical gaps remain unaddressed by existing solutions.

1.5.1 Fragmented Application of AI

Most research and tools focus on specific aspects of the security lifecycle, such as standalone anomaly detection or breach notification. There is a **lack of a unified framework** that seamlessly integrates proactive prediction, accurate detection, explainable analysis, and automated mitigation into a single, cohesive system.

1.5.2 The "Black Box" Problem and Lack of Explainability

As noted by Wang (2024) and Balantrapu (2024), the complexity of many AI models, particularly deep learning, makes them opaque "black boxes." Security operators are often unable to understand *why* a

model flagged a particular event, hindering trust, validation, and effective response. There is a significant gap in integrating **Explainable AI (XAI)** principles into operational threat detection systems.

1.5.3 Reactive vs. Proactive Posture

While tools for analyzing past breaches exist, and predictive analytics is a growing field, there are very few systems that actively use this intelligence to drive **automated, proactive mitigation** *before* a fullblown incident occurs. The link between prediction and automated action is often weak or non-existent.

1.5.4 Technical and Accessibility Barriers

Many of the most powerful AI/ML tools for security (e.g., custom-built models in Python, complex SIEM queries) require significant expertise to deploy and manage. This creates a high barrier to entry for small and medium-sized enterprises and contributes to the cybersecurity skills gap. There is a distinct **lack of user-friendly, integrated platforms** that democratize AI-powered security.

1.6 Link to the Proposed System

Building upon the reviewed literature and identified gaps, the proposed **"AI-Powered Threat Detection and Mitigation System"** is designed to be a holistic solution that bridges the shortcomings of current fragmented approaches.

Bridging Gaps with Innovation

This survey confirms that while AI components exist, no single solution effectively combines them for end-to-end, proactive defense. Our system is designed to bridge this gap by offering:

- **Proactive Threat Detection:** A hybrid AI model using **Isolation Forest for unsupervised anomaly detection** and **Random Forest for supervised classification** to identify both novel and known threats with high accuracy.
- **Explainable Insights:** A focus on providing clear, actionable alerts that security analysts can understand and trust, addressing the "black box" problem.
- **Automated Mitigation:** An integrated **mitigation module** that executes safe, predefined responses (e.g., process termination, IP blocking) upon high-confidence threat classification.
- **Threat Intelligence Integration:** A **TTP extraction component (TTPXHunter)** that parses cyber intelligence feeds to update the system's knowledge base, ensuring defenses evolve with the threat landscape.
- **User-Friendly Dashboard:** A centralized visualization interface that provides a clear overview of the security posture, detected threats, and system activity, making advanced AI accessible.

Proposed Architecture at a Glance

The system will follow a modular architecture:

1. **Data Input Layer:** Ingests real-time system metrics, network traffic, and log data.
2. **AI Processing Layer:** Comprises the Anomaly Detection and Supervised Classification engines.
3. **Decision & Mitigation Layer:** Evaluates AI output and executes automated response actions.
4. **Intelligence & Learning Layer:** Houses the TTPXHunter for continuous knowledge updates.
5. **Visualization & Interface Layer:** A dashboard for monitoring, alerts, and system control.

1.7 Future Scope & Opportunities

The proposed system provides a strong foundation that can be expanded in several directions to become a comprehensive digital defense center.

1. **Enhanced Explainability (XAI):** Integrate advanced XAI techniques like LIME or SHAP to provide detailed, human-readable explanations for every AI-driven decision.
2. **AI-Powered Risk Scoring:** Develop a dynamic risk scoring engine that assigns a quantitative risk level to each detected threat, prioritizing responses for security teams.
3. **Deception Technology Integration:** Incorporate honeypots and decoys, with AI used to analyze attacker interactions with these systems, gathering invaluable intelligence.
4. **Cross-Platform Integrations:** Develop plugins for popular SIEMs (e.g., Splunk, Elastic) and cloud platforms (AWS, Azure) for seamless integration into existing enterprise environments.
5. **Collaborative Defense:** Explore a federated learning model where multiple instances of the system can learn from each other's encountered threats without sharing sensitive raw data, creating a community defense network.

1.8 Conclusion

The escalating sophistication of cyber threats, particularly APTs, has rendered traditional, signaturebased cybersecurity measures insufficient. This literature survey has synthesized current research, confirming that Artificial Intelligence and Machine Learning are not merely enhancements but essential components for a modern, proactive defense strategy. While significant progress has been made in applying AI to threat detection, critical gaps remain in creating unified, explainable, and actionable systems that seamlessly integrate detection with automated mitigation.

The proposed **"AI-Powered Threat Detection and Mitigation System"** is positioned to address these gaps directly. By leveraging a hybrid AI model, automated response mechanisms, and a focus on explainability and usability, it aims to empower organizations to transition from a reactive to a proactive cybersecurity posture. As the digital threat landscape continues to evolve, such integrated and intelligent systems will be paramount in building resilient and secure digital infrastructures.

2.0 References

1. Z. Wang, "Artificial Intelligence in Cybersecurity Threat Detection," *International Journal of Computer Science and Information Technology*, vol. 4, no. 1, pp. 203-209, 2024.
2. T. Mori, "AI for cyber resilience in critical infrastructure," *Journal of Cyber Resilience*, 2023.
3. B. Pulyala, "AI-powered SIEM for proactive threat hunting and risk mitigation," *Infosec Journal*, 2024.
4. V. Labu and M. Ahammed, "Next-gen cyber threat detection/mitigation using AI/ML," in *Proc. Int. Conf. on Comp. Security*, 2024.
5. B. Balantrapu, "AI for predictive cyber threat intelligence (CTI)," *IEEE Security & Privacy*, 2024.
6. M. Aldawsari and S. Kouchay, "Integrating AI/ML into cloud security frameworks," *Cloud Security Review*, 2024.
7. Venkoba et al., "General review of AI for threat detection/prevention," *ACM Computing Surveys*, 2024.
8. Niveditha et al., "Predictive analytics using AI for cyber threat intelligence," *Journal of AI Research*, 2024.
9. LeBlanc, "Detecting APTs using threat modeling and meta-detections," *SANS Institute Whitepaper*, 2025.
10. Kumar et al., "Overview of AI (specifically ML) in cybersecurity for threat detection and response," *Cybersecurity Journal*, 2023.