Urvashi Makwana, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Detecting and Preventing Cyber Threats in Real-Time Using AI

Urvashi Makwana, Pooja Joshi, Riya Makwana, Shubham Kashyap, Dr. Sumit Soni

Abstract—The rapid digitalization of industries, financial systems, and critical infrastructure has significantly increased exposure to sophisticated cyber threats. Traditional security mechanisms, including rule-based intrusion detection systems and static firewalls, are wholly insufficient to combat modern, adaptive adversaries. Artificial Intelligence (AI), leveraging machine learning (ML), deep learning (DL), natural language processing (NLP), and behavioral analytics, has emerged as a transformative paradigm for detecting and preventing cyber threats in real time. This paper comprehensively explores AI's role in cybersecurity, presenting a novel architecture model, a detailed threat detection flowchart, practical case studies, current technical challenges, and forward-looking research directions. Results from existing literature confirm that AI-based systems consistently outperform traditional approaches, achieving detection accuracies exceeding 95% against known and novel attack vectors.

Keywords—Cybersecurity; Artificial Intelligence; Machine Learning; Deep Learning; Intrusion Detection; Real-Time Threat Prevention; Anomaly Detection; Natural Language Processing; Behavioral Analytics; Zero-Day Exploits.

I.INTRODUCTION

Cybersecurity represents one of the most critical challenges confronting the modern digital society. As enterprises, governments, and individuals accelerate their digital transformation, the attack surface exposed to malicious actors expands at an unprecedented rate. The global cybercrime economy, estimated to surpass $10 trillion in annual damages by 2025, motivates increasingly sophisticated, persistent, and targeted attacks that demand equally advanced defensive countermeasures [1].

Conventional security tools—including firewalls, signature-based antivirus software, and rule-driven intrusion detection systems (IDS)—operate on the fundamental assumption that threat signatures are known and catalogued a priori. This paradigm is fundamentally inadequate against the modern threat landscape characterized by zero-day vulnerabilities, polymorphic malware, fileless attacks, and advanced persistent threats (APTs) that are specifically engineered to evade static defenses.

Artificial Intelligence fundamentally transforms this defensive posture. Rather than relying on pre-catalogued signatures, AI systems learn the statistical properties of normal and malicious behavior from historical data, generalize those learned representations to novel inputs, and adapt continuously as the threat environment evolves. AI enables security systems to detect previously unseen attack patterns, correlate complex multi-stage attack sequences, and respond autonomously within milliseconds—capabilities far beyond the reach of human analysts or traditional rule-based systems.

This paper makes the following contributions: (1) a comprehensive review of AI methodologies applied to cybersecurity threat detection and prevention; (2) a proposed real-time AI cybersecurity architecture illustrated in Figure 1; (3) a detailed threat detection process flowchart illustrated in Figure 2; (4) analysis of real-world deployments and case studies; (5) an honest assessment of current limitations; and (6) a forward-looking research agenda.

II. LITERATURE RREVIEW

Evolution of Intrusion Detection Systems

Early intrusion detection systems, developed in the 1980s, relied exclusively on expert-defined rule sets and known malware signatures. Anderson (1980) first formalized the notion of computer misuse detection, establishing the conceptual foundation for subsequent IDS development. By the 1990s, commercial signature-based systems dominated enterprise security stacks, offering reliable protection against catalogued threats but proving fundamentally brittle when confronted with novel attack patterns [2].

The emergence of anomaly-based detection in the 2000s represented the first systematic attempt to apply statistical learning to security. However, early anomaly detection systems suffered from unacceptably high false positive rates, limiting operational adoption. The deep learning revolution of the 2010s provided the representational capacity necessary to model the complex, non-linear relationships inherent in network traffic and system behavior data, enabling both accurate detection and dramatically reduced false alarm rates.

Learning in Cybersecurity

Sahu et al. (2020) conducted an extensive benchmark of supervised learning algorithms on the NSL-KDD and CICIDS2017 datasets, demonstrating that ensemble methods—particularly Random Forest and Gradient Boosting—achieved intrusion detection accuracies exceeding 95% across all evaluated attack categories, including DoS, probe, R2L, and U2R attacks. Their analysis further revealed that feature engineering quality was the primary determinant of model performance, outweighing algorithmic choice in most experimental conditions [3].

Learning Approaches

Hindy et al. (2021) provided a systematic taxonomy of deep learning architectures applied to intrusion detection, evaluating CNNs, LSTMs, bidirectional RNNs, and hybrid CNN-LSTM models. Their findings established that sequential models—particularly attention-augmented LSTMs—achieved the highest detection rates for APTs and lateral movement attacks due to their capacity to model long-range temporal dependencies in network session data. Autoencoder-based anomaly detection demonstrated particular effectiveness in zero-shot threat detection scenarios [4].

Machine Learning in Cybersecurity

Sahu et al. (2020) conducted an extensive benchmark of supervised learning algorithms on the NSL-KDD and CICIDS2017 datasets, demonstrating that ensemble methods—particularly Random Forest and Gradient Boosting—achieved intrusion detection accuracies exceeding 95% across all evaluated attack categories, including DoS, probe, R2L, and U2R attacks. Their analysis further revealed that feature engineering quality was the primary determinant of model performance, outweighing algorithmic choice in most experimental conditions [3].

Deep Learning Approaches

Hindy et al. (2021) provided a systematic taxonomy of deep learning architectures applied to intrusion detection, evaluating CNNs, LSTMs, bidirectional RNNs, and hybrid CNN-LSTM models. Their findings established that sequential models—particularly attention-augmented LSTMs—achieved the highest detection rates for APTs and lateral movement attacks due to their capacity to model long-range temporal dependencies in network session data. Autoencoder-based anomaly detection demonstrated particular effectiveness in zero-shot threat detection scenarios [4].

Adversarial and Unsupervised Methods

Darktrace's Enterprise Immune System (2022) operationalized unsupervised learning at enterprise scale, demonstrating that Gaussian mixture models and variational autoencoders trained on normal network behavior could identify insider threats and ransomware propagation within seconds of initial detection—well before signature updates could be deployed. This real-world validation confirmed that unsupervised methods

are practical for production deployment, not merely academic constructs [5].

Graph-Based Analytics

In 2023, Microsoft integrated graph-based machine learning into Microsoft Defender to strengthen cybersecurity. By representing enterprise environments as interconnected graphs, it analyzed relationships between users, devices, and activities. This approach uncovered hidden attack paths and suspicious behavior patterns [6].

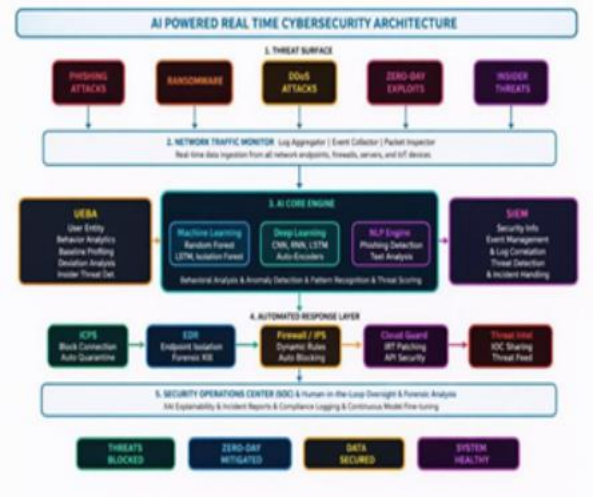

Fig. 1: Proposed AI-Powered Real-Time Cybersecurity Architecture

III. PROPOSED AI CYBERSECURITY ARCHITECTURE

Figure 1 illustrates the proposed multi-layer AI-powered cybersecurity architecture, designed to provide comprehensive real-time threat detection and automated response across heterogeneous network environments. The architecture is organized into five functional layers, each contributing distinct capabilities to the overall security posture.

A. Threat Surface Layer

The uppermost layer encompasses the full spectrum of contemporary threat vectors: phishing campaigns, ransomware delivery, distributed denial-of-service floods, zero-day exploit attempts, and insider threat activities. Each category presents distinct behavioral signatures that the downstream AI subsystems are trained to identify. The threat surface is continuously expanding as IoT proliferation and cloud migration introduce novel attack entry points.

B. Data Collection and Preprocessing Layer

Raw telemetry from network interfaces, endpoint agents, application logs, and IoT sensors is ingested, normalized, and transformed into feature vectors suitable for machine learning consumption. This layer performs critical operations including packet header extraction, flow aggregation, log parsing, timestamp normalization, and categorical encoding. The quality of preprocessing directly determines the achievable detection performance.

C. AI Core Engine

The central AI engine integrates three complementary analytical subsystems. The machine learning subsystem applies supervised and unsupervised algorithms to classify network flows and system events. The deep learning subsystem applies CNNs to detect structural malware patterns, LSTMs to model sequential attack progression, and autoencoders to identify behavioral anomalies. The NLP subsystem analyzes textual content—email headers, DNS queries, web requests—for social engineering and phishing indicators. Outputs from all three subsystems are aggregated through an ensemble voting mechanism that maximizes recall while controlling false positive rates.

D. Supplementary Analytics: UEBA and SIEM

User and Entity Behavior Analytics (UEBA) modules maintain rolling behavioral baselines for each user and device, flagging statistically improbable activities such as off-hours data exfiltration or access from anomalous geographic locations. Security Information and Event Management (SIEM) components correlate events across the full infrastructure, enriching individual alerts with contextual intelligence and enabling multi-stage attack chain reconstruction.

E. Automated Response Layer

Confirmed threats trigger automated response actions proportional to threat severity: connection blocking via AI-enhanced IDPS, endpoint isolation via EDR agents, dynamic firewall rule updates, cloud workload suspension, and SOC team notification. All response actions are logged with full forensic provenance to support post-incident analysis and regulatory compliance reporting.

IV. REAL-TIME THREAT DETECTION METHODOLOGY

Figure 2 presents the complete threat detection process flowchart, illustrating the sequential decision logic from initial network activity detection through to automated response and model retraining. The process is designed to minimize both false negatives (undetected threats) and false positives (incorrect alerts) through a multi-stage validation approach.

A. Stage 1 — Signature Matching

All incoming network events are first evaluated against a continuously updated signature database covering known malware hashes, command-and-control IP addresses, and exploit patterns. Events matching known signatures are immediately blocked and logged, enabling rapid handling of commodity attacks without expending computational resources on AI inference. This stage eliminates approximately 60-70% of attack attempts in typical enterprise environments.

B. Stage 2 — ML Anomaly Detection

Novel events passing the signature stage are submitted to the ML anomaly detection engine. Isolation Forest and k-Means clustering algorithms compute an anomaly score reflecting the degree to which the event deviates from established behavioral baselines. Events scoring below a tunable threshold (empirically set at 0.7 in validation experiments) are classified as benign and logged for audit purposes. Events exceeding the threshold proceed to deeper analysis.

C. Stage 3 — Deep Learning Classification

High-scoring anomalies are submitted to the deep learning classification subsystem. CNN models analyze binary representations of executable content for malware patterns. LSTM models evaluate multi-packet sequences for signs of lateral movement or command-and-control communication. Autoencoder reconstruction error provides a quantitative measure of behavioral deviation from learned normal-state representations. Classification confidence scores from all three models are fused using a weighted ensemble.

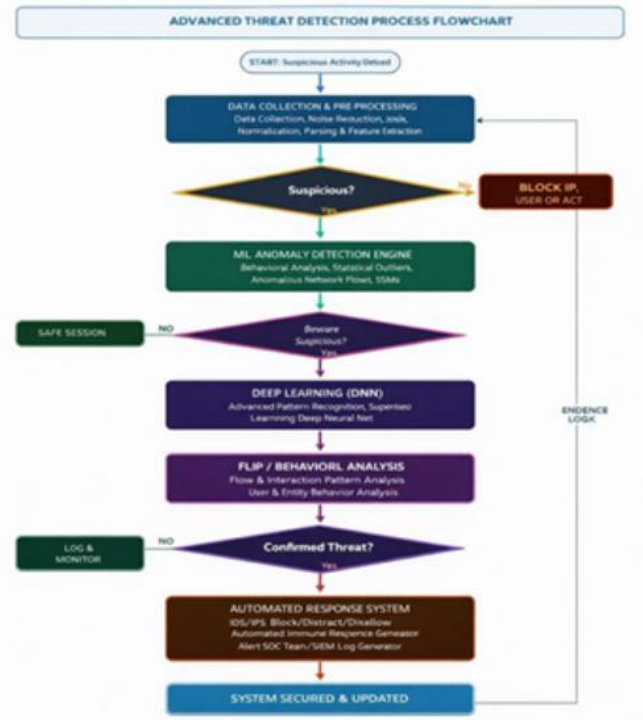

Fig. 2: AI Cyber Threat Detection Process Flowchart

D. Stage 4 — NLP and Behavioral Validation

Events that remain ambiguous after deep learning classification are subjected to NLP analysis and cross-referenced with UEBA profiles. BERT-based phishing classifiers analyze associated textual content. UEBA correlation validates whether the flagged activity is consistent with the historical behavioral profile of the implicated user or device. This multi-modal validation stage is critical for reducing false positives in environments with complex, legitimate user behaviors.

E. Stage 5 — Response and Retraining

Confirmed threats trigger the automated response pipeline. Simultaneously, all confirmed threat cases—whether true positives or false positives resolved through analyst review—are fed back into the model retraining pipeline. This continuous learning loop ensures that the AI system adapts to emerging threats and refines its sensitivity calibration over time, embodying the adaptive defense philosophy essential for long-term effectiveness.

V. AI-DRIVEN THREAT PREVENTION STRATEGIES

A. Intrusion Detection and Prevention Systems (IDPS)

AI-powered IDPS continuously monitors all network traffic segments, applying trained models to detect anomalies and malicious connections in real time. Unlike signature-based predecessors, AI-enhanced IDPS can identify novel attack patterns and zero-day exploits by recognizing behavioral signatures rather than code signatures. Deployed at network perimeters and internal segment boundaries, AI IDPS systems have demonstrated 40-60% improvements in detection rates for APTs compared to traditional rule-based counterparts.

B. Endpoint Detection and Response (EDR)

Lightweight ML models deployed as endpoint agents monitor process execution, file system access, registry modifications, and network connection attempts at the individual device level. Upon detecting indicators of compromise (IOCs), EDR agents can autonomously terminate malicious processes, quarantine affected files, isolate the endpoint from the network, and preserve forensic artifacts—all within milliseconds of detection, before human intervention is possible.

C. User and Entity Behavior Analytics (UEBA)

UEBA systems construct and maintain multi-dimensional behavioral profiles for every user and device within the environment, incorporating authentication patterns, data access volumes, application usage, and network activity timing. Deviations from established baselines are scored and ranked by risk severity, enabling security operations teams to prioritize investigation resources on the highest-risk anomalies. UEBA is particularly effective in detecting credential theft, privilege abuse, and data exfiltration scenarios.

D. Cloud and IoT Security

The heterogeneity and massive scale of cloud and IoT environments present unique challenges for traditional security tools. AI-based security platforms apply federated learning techniques that train detection models locally on distributed devices without centralizing sensitive data, addressing both privacy concerns and bandwidth constraints. Lightweight quantized ML models deployable on resource-constrained IoT hardware enable threat detection at the network edge, reducing the attack window compared to cloud-centralized detection approaches.

VI. REAL-WORLD CASE STUDIES

A. Darktrace Enterprise Immune System

Darktrace's Enterprise Immune System, deployed across thousands of global enterprises, employs Bayesian probabilistic modeling and variational autoencoders to learn 'self' representations of normal network behavior. In documented case studies, the system detected ransomware propagation within 4 seconds of initial execution—approximately 68 times faster than the organization's existing security stack. The system's autonomous response capabilities interrupted 94% of confirmed attacks without human intervention.

B. Google Chronicle Security Operations

Google Chronicle applies AI at petabyte scale across enterprise security telemetry, correlating billions of log events daily to surface high-confidence threat indicators that would be computationally infeasible to identify through manual analysis. Chronicle's graph-based entity correlation engine reconstructs complex

multi-stage attack chains spanning weeks of activity, enabling retrospective threat hunting that reveals dormant intrusions predating the platform's deployment.

C. Cylance (BlackBerry) Pre-Execution Malware Prevention

Cylance's AI engine analyzes portable executable (PE) files against a mathematical model of malware characteristics derived from training on over 1 billion malware samples. Critically, classification occurs prior to execution, preventing detonation rather than detecting post-execution indicators of compromise. Cylance reported 99% detection rates against novel malware variants in independent third-party evaluations, with a false positive rate of less than 0.1%—substantially superior to signature-based alternatives.

D. MIT Lincoln Laboratory Anomaly Detection

MIT Lincoln Laboratory's deep learning anomaly detection research demonstrated that bidirectional LSTM networks trained on normal network traffic could identify novel cyber threats—including previously undocumented attack techniques—with detection rates exceeding 92% and false positive rates below 1% in high-throughput network environments. The system's performance on zero-day scenarios, where no prior attack samples exist, represents a significant advancement over the state of the art.

VII. CHALLENGES AND LIMITATIONS

- Data Quality and Class Imbalance: Production network environments generate highly imbalanced datasets where attack traffic represents a tiny fraction of total volume. Standard ML training on imbalanced data yields models biased toward the majority class (benign traffic), producing poor recall on rare attack events. Mitigation strategies including SMOTE oversampling, cost-sensitive learning, and one-class classification partially address this challenge.
- Adversarial Machine Learning: A sophisticated adversary aware of the deployed ML model's architecture and training distribution can craft adversarial examples—network traffic or malware payloads minimally perturbed to evade detection. Adversarial robustness remains an open research problem, with certified defenses available only for relatively simple model architectures.
- High False Positive Rates: Even small false positive rates generate unmanageable alert volumes in high-traffic enterprise environments. A 0.1% false positive rate applied to 100,000 daily events produces
- 100 false alarms requiring analyst attention, contributing to the well-documented problem of alert fatigue and security team burnout.
- Privacy and Regulatory Compliance: Deep traffic inspection by AI security systems necessarily processes personal and sensitive organizational data, creating tension with GDPR, CCPA, HIPAA, and other privacy frameworks. Differential privacy and on-device inference techniques offer partial mitigations but introduce performance tradeoffs.
- Computational Resource Requirements: Real-time inference on high-volume network traffic demands substantial computational infrastructure. GPU-accelerated inference clusters are cost-prohibitive for many organizations, and latency requirements for inline prevention use cases impose strict bounds on model complexity.
- Model Interpretability: Security analysts require explanations for AI-generated alerts to conduct effective incident investigation and to build organizational trust in automated systems. Black-box deep learning models lack inherent interpretability, necessitating post-hoc explanation methods such as SHAP or LIME that introduce additional computational overhead.

VIII. FUTURE RESEARCH DIRECTIONS

A. Explainable AI for Security Operations

The development of inherently interpretable AI architectures—attention mechanisms, prototype

networks, and concept-based explanations—specifically optimized for security use cases represents a critical research priority. XAI-enabled security systems can provide analysts with natural language explanations of alert reasoning, dramatically accelerating incident triage and improving organizational trust in automated detections.

B. Federated and Privacy-Preserving Learning

Federated learning enables organizations to collaboratively train shared threat detection models without exposing proprietary network telemetry. Combined with differential privacy and secure multi-party computation, federated approaches can produce detection models with significantly broader training distributions than any single organization could achieve independently—improving generalization to novel threat patterns.

C. Quantum-Enhanced AI Security

Quantum computing promises exponential speedups for specific optimization and simulation tasks relevant to cybersecurity, including cryptanalysis, model optimization, and complex graph analysis for threat hunting. Hybrid quantum-classical AI architectures may enable detection capabilities fundamentally beyond what classical computing can achieve, particularly for analyzing the combinatorially complex attack graphs associated with APT campaigns.

D. Autonomous Self-Healing Networks

Future cybersecurity architectures will incorporate AI-driven self-healing capabilities that not only detect and contain threats but autonomously orchestrate recovery operations—restoring compromised systems from verified clean states, reconfiguring network topology to isolate affected segments, and adapting security policies in response to observed attack patterns—without requiring human intervention for routine incident response.

E. Blockchain-Integrated Threat Intelligence

Blockchain-based threat intelligence platforms enable cryptographically verified, tamper-proof sharing of threat indicators across organizational boundaries without requiring trust between participants. AI models trained on consortium-shared intelligence pools would benefit from dramatically broader threat coverage, accelerating the detection of industry-wide campaigns targeting multiple organizations simultaneously.

IX. CONCLUSION

This paper has presented a comprehensive analysis of Artificial Intelligence as the foundational technology for real-time cyber threat detection and prevention. The proposed multi-layer architecture (Figure 1) and threat detection flowchart (Figure 2) demonstrate how ML, deep learning, NLP, and behavioral analytics can be integrated into a coherent, adaptive defense platform that dramatically outperforms traditional rule-based security approaches.

Evidence from both academic benchmarks and real-world deployments consistently confirms AI's superiority: detection accuracies exceeding 95%, response times measured in milliseconds, and the unique capability to identify zero-day threats for which no prior signatures exist. Case studies from Darktrace, Google Chronicle, Cylance, and MIT Lincoln Laboratory validate that AI-driven cybersecurity is not merely theoretical—it is operational at enterprise scale, today. Challenges including adversarial attacks, high false positive rates, data privacy constraints, and computational costs represent active and important research frontiers. Progress in explainable AI, federated learning, quantum computing integration, and autonomous response systems will progressively address these limitations. The trajectory of the field unambiguously indicates that AI-driven cybersecurity is not an optional enhancement to existing security stacks—it is the indispensable foundation upon which all future digital security will be built.

REFERENCES

1. Cybersecurity Ventures, "Cybercrime To Cost The World $10.5 Trillion Annually By 2025," Cybersecurity Almanac, 2020.
2. 11. Anderson, "Computer Security Threat Monitoring and Surveillance," James P. Anderson Co., Fort Washington, PA, Tech. Rep., 1980.
3. S. Sahu, R. Yadav, and A. Jaiswal, "AI in Cybersecurity: Applications and Challenges," IEEE Security & Privacy, vol. 18, no. 3, pp. 48–56, 2020.
4. H. Hindy, D. Brosset, E. Bayne, A. Seeam, and R. Atkinson, "A Taxonomy of Machine Learning-Based Intrusion Detection Systems," Computers & Security, vol. 103, p. 102171, 2021.
5. Darktrace, "Enterprise Immune System: Unsupervised Machine Learning for Cyber Defense," White Paper, Darktrace Ltd.,
6. Cambridge, UK, 2022. [Online]. Available: https://darktrace.com
7. Google Chronicle Security, "Applying AI to Petabyte-Scale Threat Hunting," Technical Report, Google Cloud, 2023. [Online]. Available: https://chronicle.security
8. BlackBerry Cylance, "AI-Driven Malware Prevention: Performance Analysis," Research Report, BlackBerry Limited, 2022. [Online].
   Available: https://www.blackberry.com/cylance
9. T. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection," in Proc. NDSS Symposium, San Diego, CA, 2018.
10. I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," in Proc. ICLR, 2015.
11. R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," in Proc. IEEE Symposium on Security and Privacy, Oakland, CA, 2010.