ISSN: 2583-9888
DOI: doi.org/10.51219/JAIMLD/Victor-Oriakhi-Nosakhare/568

URF PUBLISHERS
connect with research world

Journal of Artificial Intelligence, Machine Learning and Data Science

https://urfpublishers.com/journal/artificial-intelligence

Vol: 3 & Iss: 2

Research Article

Machine Learning in Cybersecurity: A Multi-Industry Case Study Analysis for Enhanced Threat Detection and Response

Victor Oriakhi Nosakhare[1], Bisola Kayode[2], Samuel Akerele[3], Ibrahim Yakub[4] and Fatai Ayobami Popoola[5]

[1]Sonae, UK

[2]Independent Researcher, UK

[3]Vuhosi Limited, UK

[4]Readrly Limited, UK

[5]Middlesex University, UK

Citation: Nosakhare VO, Kayode B, Akerele S, et al. Machine Learning in Cybersecurity: A Multi-Industry Case Study Analysis for Enhanced Threat Detection and Response. *J Artif Intell Mach Learn & Data Sci 2025* 3(2), 2684-2691. DOI: doi.org/10.51219/JAIMLD/Victor-Oriakhi-Nosakhare/568

Received: 11 May, 2025; **Accepted:** 15 May, 2025; **Published:** 13 June, 2025

***Corresponding author:** Victor Oriakhi Nosakhare - Sonae, UK

ABSTRACT

Machine learning (ML) has rapidly emerged as a cornerstone technology in advancing cybersecurity across multiple industries. This review article provides a comprehensive analysis of how ML-driven techniques are being leveraged to strengthen threat detection, accelerate incident response and improve the overall security posture in diverse operational landscapes, ranging from finance and healthcare to critical infrastructure and cloud-native environments. We begin by surveying the evolving threat landscape, highlighting the limitations of traditional signature-based approaches and the need for adaptive data-driven defences. Our synthesis of recent research and industrial deployments illustrates how state-of-the-art ML methods, including supervised, unsupervised, semi-supervised and reinforcement learning, can effectively detect anomalies, identify zero-day vulnerabilities, classify malicious activities and guide automated decision making.

Drawing from real-world case studies, we examine the key factors influencing ML's performance and reliability of ML, such as data quality, model interpretability, adversarial robustness and integration into existing security architectures. We explore practical considerations for model selection, feature engineering, continuous learning and lifecycle management to ensure both scalability and resilience. Additionally, we review novel approaches that combine ML with traditional cybersecurity tools and processes as well as the emerging role of federated learning and privacy-preserving techniques in safeguarding sensitive data.

Through the integration of these findings, we offer a comprehensive strategic framework that enables researchers, practitioners and policymakers to evaluate the current status of machine learning in the field of cybersecurity. Our analysis identified research gaps and future directions, including the potential of automated ML (AutoML), transfer learning and causal inference to yield more adaptive, context-aware defences. Ultimately, this review offers a multi-industry perspective that underscores the transformative potential of ML in enhancing threat detection and response, guiding the field toward more robust and intelligent cybersecurity ecosystems.

1. Introduction

A. The growing cybersecurity threat landscape

Over the past decade, the frequency and sophistication of cyber-attacks have escalated dramatically, posing significant challenges for organisations and governments worldwide. Advanced persistent threats (APTs), zero-day vulnerabilities and large-scale distributed denial-of-service (DDoS) attacks target sensitive data and critical infrastructure[1-3]. According to recent threat intelligence reports, attackers are adopting more covert and polymorphic tactics, enabling them to evade conventional security measures and exploit inherent weaknesses in both legacy and modern network architectures[4-6].

Traditional cybersecurity defences largely rely on signature-based detection methods, which are effective against known threats, but struggle to identify previously unseen attacks. These static approaches, although computationally inexpensive and easy to maintain, falter due to their vulnerability to polymorphic attacks and adversarial techniques that exploit emerging technologies such as large language models (LLMs) [7,1]. Moreover, overreliance on manual rule creation and heuristic filtering leads to increased false positive and false negative rates, causing alert fatigue among security analysts and ultimately undermining the effectiveness of an organisation's defense posture[8]. In light of these evolving challenges, there is an urgent need for more adaptive and proactive security measures that not only detect but also anticipate emerging threats. Such measures must incorporate dynamic data-driven mechanisms capable of learning from complex, rapidly changing environments. This calls for a paradigm shift, transitioning from static, rule-based defences toward intelligent systems that continuously refine their understanding of normal and malicious behaviour patterns.

B. The emergence of machine learning (ML) in cybersecurity

Machine Learning, a branch of artificial intelligence focused on data-driven pattern recognition and predictive modelling, has emerged as a pivotal technology in modernising cybersecurity strategies. By leveraging vast amounts of heterogeneous data, ML techniques can automatically identify anomalous traffic patterns, classify malicious binaries, detect intrusions and predict potential vulnerabilities more accurately and efficiently than traditional methods[2,9]. ML models capture subtle relationships and correlations that often elude human experts, thereby enhancing threat detection and response capabilities[4]. In practical terms, ML-powered cybersecurity solutions can adapt to evolving attacker behaviours, improving detection rates for zero-day exploits and reducing response times during incident handling[10].

For example, anomaly detection algorithms can uncover previously unknown attack patterns hidden within network telemetry, whereas supervised classification models can provide early warnings of impending breaches based on historical attack signatures and known attacker profiles[1]. Additionally, ML models facilitate automated data analysis and alert prioritisation, reducing false positives and mitigating analyst fatigue[11]. Collectively, these capabilities streamline the remediation process, helping security teams prioritise alerts, allocate resources efficiently and implement more informed and proactive strategies to mitigate future threats.

C. Scope and objectives of the review

This review focuses on the implementation and impact of ML-driven cybersecurity solutions across multiple industries, including finance, healthcare, critical infrastructure and cloud-based environments. By examining real-world case studies, we aim to present a cross-sector perspective that highlights the versatility and efficacy of ML techniques in vastly different operational contexts[4].

Specifically, this study provides a comparative analysis of various ML approaches, detailing lessons learned, best practices and key factors influencing model selection, performance and scalability. It also identifies critical research gaps, such as the need for more explainable models, improved adversarial robustness and standardised evaluation frameworks, while outlining potential directions for future investigation, including federated learning, causal inference and automated machine learning (Auto ML). By synthesising these insights, this review seeks to guide researchers, practitioners and policymakers in leveraging ML to create more resilient, adaptive and intelligent cybersecurity ecosystems.

2. Background: Foundations of Machine Learning in Cybersecurity

A. Overview of machine learning techniques applicable to cybersecurity

The application of machine learning (ML) to cybersecurity relies on a variety of algorithmic paradigms, each suited to different aspects of threat detection and defence. Supervised learning, one of the most widely adopted approaches, involves training models using labelled datasets, enabling them to classify malicious traffic or predict the probability of a known exploit recurring[2,9]. Supervised techniques such as decision trees, random forests and deep neural networks are used to distinguish benign from hostile activities when sufficient ground-truth data are available.

However, real-world scenarios often involve emerging previously unseen threats that defy traditional labelling and pattern recognition. In these instances, unsupervised learning methods, including clustering and anomaly detection algorithms, can detect novel or rare attack behaviours without relying on labelled examples[1]. By modelling "normal" system behaviour, unsupervised techniques can flag deviations symptomatic of intrusions or data exfiltration attempts, enabling early detection of zero-day vulnerabilities.

Beyond strictly supervised or unsupervised paradigms, semi-supervised and reinforcement learning (RL) approaches have emerged to address complex adaptive security scenarios[12]. Semi-supervised methods efficiently use partially labelled datasets, bridging the gap between expert knowledge and the vast amounts of unlabelled data available in network telemetry, enabling more accurate and resource-efficient threat detection and response. RL agents, on the other hand, continuously improve defence strategies by interacting with a dynamic environment-evaluating the outcomes of detection and response actions and refining policies to maximise long-term system security[12].

B. Data sources and feature engineering

Effective ML-driven cybersecurity solutions are predicated on the availability and quality of underlying data. Network traffic logs, derived from routers, switches and firewalls, provide low-level data that can be mined for patterns indicative of malicious activity[4]. Endpoint telemetry, including host-based events, process behaviour and file system modifications, enriches these analyses by offering granular insight into the attacker's presence and lateral movement. In addition, security event logs generated by intrusion detection systems (IDS), antivirus tools and authentication servers serve as valuable labelled signals for supervised training and model validation[13].

Feature engineering plays a critical role in extracting informative and discriminative attributes from raw data. This may include protocol analysis (for example, HTTP headers, DNS queries and SSL handshake sequences) as well as deriving behavioural metrics such as traffic volume anomalies, unusual login times or abnormal command execution patterns[2]. However, building robust models often requires balancing data quality, volume and diversity. Large-scale datasets can improve generalisation but may introduce noise and complexity, whereas overly curated data can fail to represent realistic operational environments. Striking the right balance ensures that ML models are both accurate and resilient to adversarial manipulation.

C. Integrating ML models into security infrastructures

Implementing ML-driven defences requires seamless integration into existing security frameworks and tool chains. Placing ML models within Security Information and Event Management (SIEM) systems enables automated correlation, prioritisation and escalation of alerts, thereby reducing the manual overhead for security analysts[14].

This integration allows defenders to merge ML-based insights with rule-based detection and threat intelligence feeds for a more holistic security posture.

When deploying ML solutions organisations must also consider the real-time and batch-processing trade-offs. Real-time analysis is essential for rapid threat detection and response; however, it may require more computational resources and robust feature selection strategies[1]. Batch processing, in contrast, can support periodic retraining and retrospective analysis, which helps models adapt to evolving threat landscapes over time.

Finally, continuous monitoring and lifecycle management are fundamental to long-term success. As attacker tactics evolve, models must be periodically retrained using fresh data, validated against emerging threats and updated to maintain their effectiveness. This iterative improvement cycle ensures that ML systems remain aligned with operational needs and resilient against adversarial attempts to subvert or evade detection[4].

3. ML Techniques for Threat Detection and Response

A. Intrusion detection and intrusion prevention systems (IDS/IPS)

Intrusion detection and prevention systems are central to modern cybersecurity frameworks. Traditional signature-based IDS approaches rely on predefined patterns and known attack signatures, making them vulnerable to adversaries employing novel or polymorphic threats[1,15]. In contrast, machine-learning-driven IDS/IPS harness data-driven models learn to recognise suspicious patterns of behaviour. These behaviour-based approaches enable the detection of both known and previously unseen threats with greater agility.

A key class of ML models deployed for intrusion detection involves anomaly detection techniques, which model "normal" network conditions and identify deviations that may indicate an intrusion[4]. For instance, algorithms such as Isolation Forest efficiently isolate anomalous instances by recursively partitioning the dataset, whereas autoencoders, a type of neural network trained to reconstruct input data, highlight anomalies by measuring reconstruction errors[16,17]. These approaches enhance the accuracy, reduce false positives and adapt to evolving threats, making them indispensable in dynamic environments[18].

B. Malware classification and zero-day detection

Malware poses a pervasive threat in diverse computing environments. Traditional static analysis methods leverage supervised learning approaches, such as decision trees, support vector machines and gradient-boosted ensembles, trained on labelled datasets of known malware samples, enabling the high-fidelity classification of known malware families[2]. Beyond these known threats, zero-day vulnerabilities and new malware strains often evade signature-based methods. In such cases, unsupervised methods help identify anomalous file or network behaviours that do not match any known malicious pattern, detecting zero-day exploits before the corresponding signatures are available[19].

Additionally, deep learning has emerged as a powerful tool for malware analysis, enabling models to learn complex hierarchical feature representations. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been applied to analyse both static binary features and dynamic system call sequences, identifying subtle indicators of compromise that simpler models might miss[20]. By combining classical supervised learning with advanced representation learning, these deep models can improve zero-day detection rates and reduce the time-to-response.

C. Phishing, social engineering and fraud detection

Phishing campaigns, social engineering attempts and fraudulent activities exploit human vulnerabilities and trust relationships. To counter these threats, ML models increasingly integrate natural language processing (NLP) techniques to parse email content, webpage text and other communication artefacts. For example, NLP-based classifiers can detect suspicious linguistic patterns, unusual grammatical structures or deceptive cues in phishing emails, thereby achieving higher precision and recall than manual keyword heuristics[21,18].

In the context of fraudulent financial transactions, graph-based ML models leverage the interconnected nature of user accounts and transaction history. By modelling relationships as nodes and edges, these approaches can identify fraudulent rings or collusion patterns that are not evident when examining isolated events[22]. Detecting these hidden structures enables security teams to proactively prevent cascading financial damage and to safeguard both institutions and customers.

D. Automated incident response

Beyond detection, machine learning can help orchestrate an efficient data-driven incident response. Reinforcement learning (RL) agents, for example, can autonomously evaluate defence actions such as isolating compromised hosts, blocking suspicious IPs, deploying patches and refining their strategies over time to minimise damage[12]. By simulating a range of attack scenarios, RL agents learn from the continuous feedback to generate automated playbooks that respond swiftly and effectively.

Furthermore, predictive analytics can guide resource allocation during an ongoing cyberattack, helping security operations centres prioritise which systems or alerts demand immediate attention. By leveraging historical data, ML models can estimate the potential severity and impact of an intrusion, enabling informed decisions that reduce downtime, data loss and the overall operational risk. Together, RL-driven automation and predictive analytics not only enhance the speed and quality of incident response but also free human analysts to focus on complex, high-level strategic tasks.

4. Multi-Industry Case Studies

A. Finance and banking

The finance sector, entrusted with safeguarding highly sensitive customer information and substantial monetary assets, exemplifies the strategic application of machine learning to combat fraud and maintain trust. Transaction anomaly detection, powered by machine learning algorithms such as gradient-boosted ensembles and deep neural networks, identifies suspicious financial activities by modelling normal spending patterns and pinpointing deviations in real time[22,23]. These systems help flag fraudulent credit card transactions or money laundering attempts swiftly, thereby minimising financial losses and preserving brand reputation.

Beyond transactional analytics, behavioural biometrics such as keystroke dynamics, mouse movement patterns and mobile sensor data provide additional layers of authentication. These ML-driven behavioural models continuously adapt to users' unique interaction styles, delivering ongoing, unobtrusive security checks that transcend static credentials like passwords or traditional challenge-response prompts[24]. Moreover, financial institutions increasingly embed ML-derived insights into broader compliance and risk management frameworks, integrating anomaly detection outputs into regulatory reporting structures and capital risk assessments. In this way, machine learning not only mitigates cyber threats but also supports strategic, long-term operational planning[2].

Real-world implementations of these technologies further validate their impact. Visa's cloud-native ARIC Risk Hub profiles over 500 million cardholders and blocks anomalous transactions in under 300 milliseconds, resulting in a 90% reduction in phishing-related financial losses for the Norwegian Eika bank consortium within its first operational year[25]. Similarly, Mastercard's "Decision Intelligence Pro" analyses up to 160 billion transactions annually, generating real-time risk scores within 50 milliseconds and helping prevent an estimated $40 billion in fraud during 2024 alone[26].

These production-scale deployments corroborate the earlier academic consensus on the superiority of adaptive models over static rule-based systems. However, they also expose two critical implementation challenges. First is the issue of bias drift. Models initially trained on predominantly North American datasets misclassified Eastern European e-commerce flows, prompting urgent domain-adaptation iterations. Second is the growing challenge of explainability. Regulators under the EU PSD2 directive required granular, interpretable outputs before authorising model deployment, compelling institutions to integrate SHAP-based local explainability dashboards to meet audit and compliance standards.

B. Healthcare and medical devices

Healthcare organisations face escalating cyber threats, including ransomware attacks on hospital networks and the potential compromise of Internet of Things (IoT) medical devices. ML-based solutions for network anomaly detection can distinguish legitimate medical-device communications from unauthorised intrusions or data-exfiltration attempts[4]. For example, unsupervised clustering models can identify unusual traffic patterns originating from connected pacemakers or infusion pumps and alert security teams before patient safety is jeopardised.

In the era of telemedicine and remote patient monitoring, maintaining the confidentiality and integrity of patient data is of paramount importance. ML-driven identity verification and anomaly detection tools can continuously authenticate healthcare providers and validate patient requests, thereby reducing the likelihood of data breaches or unauthorised access to electronic health records[12]. In addition, the use of natural language processing and encryption-aware ML algorithms can ensure that sensitive diagnostic information and treatment plans remain both accessible and protected.

C. Critical infrastructure (energy, transportation, utilities)

Critical infrastructure systems, including energy grids, water treatment plants and transportation networks, present unique cybersecurity challenges owing to their massive scale and potentially catastrophic consequences of compromise. ML models integrated into Supervisory Control and Data Acquisition (SCADA) systems enable real-time anomaly detection and early threat identification. These models can distinguish normal operational states, such as predictable fluctuations in power demand, from malicious manipulations that could degrade services or cause physical damage[27]. To counter highly sophisticated nation-state-level attacks, ML-based early warning systems aggregate threat intelligence across multiple data streams, correlating sensor readings, network telemetry and industrial process parameters. By pre-emptively identifying subtle patterns of infiltration or sabotage, these systems empower infrastructure operators to mitigate disruptions and preserve the stability of essential public services[28].

D. Cloud and virtualised environments

The shift towards cloud computing and virtualisation has introduced new attack surfaces and complexities. In multi-tenant architectures, attackers can attempt lateral movement by navigating from one compromised virtual machine to others within the same physical host. ML-based anomaly detection agents monitor east-west traffic flows to detect unexpected communication patterns and isolate compromised instances[29]. Such a layered defense ensures that even if an attacker breaches one segment of the cloud infrastructure, they cannot easily expand their reach.

Moreover, threat intelligence in containerised and serverless environments relies on ML's ability of ML to adapt to short-lived and rapidly scaling services. By modelling normal container startup times, resource utilisation and inter-service calls, ML techniques can highlight suspicious deviations and help security teams contain breaches before attackers exploit ephemeral workloads[30]. The fluidity and dynamism of cloud infrastructure make ML's speed, adaptability and automation essential components of an effective defense strategy.

E. Key lessons from cross-industry comparisons

The application of ML in finance, healthcare, critical infrastructure and cloud environments underscores the need to tailor solutions to domain-specific challenges. Data availability, labelling and privacy concerns differ markedly across industries. While financial institutions have ample transaction logs, healthcare organisations must manage sensitive patient data under strict regulatory constraints[2]. Consequently, domain-specific data challenges influence the choice of algorithms, feature engineering techniques and performance evaluation metrics.

Likewise, customisation of ML models to unique operational contexts is essential. Models trained on financial datasets may not seamlessly transfer to detecting medical device anomalies or SCADA system intrusions. Instead organisations must carefully adapt architectures, hyperparameters and training strategies to align with the nuances of each environment[1]. Nevertheless, there remain opportunities for transferability of best practices across sectors. Techniques that prove effective in detecting lateral movement in cloud infrastructure, for instance, may inform strategies for identifying similar patterns in industrial control networks, provided that careful domain adaptation occurs.

These cross-industry insights collectively highlight the potential of ML for robust, adaptive cybersecurity measures while simultaneously illuminating the complexities inherent in scaling, generalising and operationalising these solutions in diverse real-world settings.

F. Large-language-model (LLM)-enabled threats and defences

As large language models (LLMs) continue to advance, their application within cyber threat vectors is transforming traditional social engineering tactics into highly adaptive and automated attack mechanisms. LLMs now enable the automated production of grammatically flawless, contextually tailored phishing emails and, increasingly, real-time synthetic voice impersonations that mimic the speech patterns of corporate executives. These capabilities significantly reduce the time, cost and expertise required to mount targeted cyberattacks[31]. Recent intelligence from investigative reports has uncovered a proliferation of "LLM as a service" offerings on dark web forums, where even individuals with minimal technical background can deploy multilingual, context-sensitive spear-phishing campaigns at scale[32]. The commodification of generative AI in cybercrime circles marks a paradigm shift, transforming social engineering from a manual craft into a scalable automated operation.

In response, cybersecurity operations centres (SOCs) and enterprise defenders are actively deploying countermeasures that leverage the same class of models to detect and neutralise these emerging threats. One such approach is semantic anomaly detection, wherein SOCs retrain natural language classifiers on transformer-based embeddings to identify subtle inconsistencies, unexpected tone shifts or unnatural language patterns within message threads, which are key signals that often betray generative origin[33]. In parallel, graph neural networks (GNNs) are being utilised to uncover coordination patterns across attack surfaces. These models integrate heterogeneous data including email content, metadata, sender infrastructure and behavioural user-click pathways to detect and isolate distributed LLM-generated phishing campaigns in near real time[34].

Another promising defence involves the integration of user-facing LLMs into productivity and messaging platforms. These embedded assistants serve as "phish-check" agents, offering real-time coaching by paraphrasing potentially suspicious messages and drawing attention to anomalous requests. Preliminary deployments have shown measurable success. In pilot studies, users equipped with these tools demonstrated a 14 percent reduction in phishing click-through rates, indicating a positive impact on human-in-the-loop security outcomes[26]. Together, these developments illustrate the dual use nature of LLMs in cybersecurity, functioning both as a vector for novel attacks and as a foundation for next-generation defence architectures. The challenge for researchers and practitioners lies in staying ahead of offensive innovation, deploying proactive and adaptive models that can scale in step with adversarial capabilities.

5. Challenges and Limitations

A. Data quality, bias and label scarcity

A key challenge in building robust machine learning (ML) models for cybersecurity lies in ensuring access to high-quality and representative data. Many real-world datasets suffer from insufficient labelled samples, limiting the effective use of supervised learning and hindering the performance of anomaly detection methods[2]. Further compounding this issue, datasets commonly exhibit class imbalance, in which benign events vastly outnumber malicious instances. Such skewed distributions may lead to models that excel at detecting normal activities but underperform in identifying subtle or rare attacks[4].

Additional complications arise from noise, incomplete data and the prevalence of false positives and negatives in network logs and security event streams. Noise can mask attack signals, whereas incomplete data may omit the critical contextual information necessary for accurate classification. Misclassifications, such as false alarms that overload security personnel or missed intrusions that compromise systems, can erode trust in ML-driven solutions. Overcoming these data challenges requires rigorous data engineering, active learning strategies to refine labelling and continual dataset maintenance to preserve the model performance over time.

B. Interpretability and explainability of ML models

As ML models become more complex, the difficulty of interpreting their decisions increases. Regulatory and compliance requirements, particularly in sectors such as finance and healthcare, demand that organisations justify security decisions for auditing and legal purposes[1]. Highly complex deep learning models, although powerful, often function as "black boxes", whose internal logic remains opaque. Lack of transparency can raise concerns about bias, fairness and accountability, posing hurdles to widespread adoption.

From an operational perspective, trust and usability for security analysts hinge on interpretable outputs. To make informed response decisions, analysts must understand why a model flagged a particular event as malicious. Post-hoc explanation tools, such as feature importance rankings, rule extraction methods and local approximation techniques, aid in bridging this gap[35]. However, achieving a balance between predictive accuracy and interpretability remains an ongoing area of research and practice, influencing how effectively these models are integrated into day-to-day cybersecurity workflows.

C. Adversarial attacks against ML models

Attackers are becoming increasingly aware that ML models now guard critical infrastructure and sensitive data. Consequently, adversarial attacks, including poisoning (tampering with training data), evasion (crafting inputs that fool detection) and model inversion (deriving information about the model or training data) have emerged as significant threats[36]. Such attacks can degrade model performance or reveal sensitive patterns, reducing the overall effectiveness of ML-based defences.

Researchers and practitioners have developed techniques to improve the model robustness and resilience, such as adversarial training, input sanitisation and the use of robust feature representations that are less susceptible to manipulation[17]. Continual adaptation and rigorous testing against known adversarial scenarios are vital for preserving the integrity and reliability of ML-driven cybersecurity measures.

D. Scalability and integration costs

Although ML solutions promise enhanced detection and response capabilities, practical implementation at scale often involves computational overhead and resource constraints. Processing massive datasets, training large models and conducting real-time inference can strain organisational infrastructure, both in terms of hardware capacity and latency requirements[2]. Ensuring that these systems remain efficient, cost-effective and responsive as data volumes increase is a persistent engineering challenge.

In addition organisational resistance, skill gaps and the cost of adoption may impede the seamless integration of ML solutions into existing cybersecurity programs. Specialised data science and ML engineering expertise are required to select the right models, tune hyperparameters and maintain model performance over time. Without adequate training and buy-in, operational teams may hesitate to trust ML-driven alerts or fail to harness the full potential of these technologies. Overcoming these barriers demands not only technical innovation but also effective communication, training and alignment with organisational priorities to ensure that ML-driven cybersecurity solutions deliver both technological and operational value.

6. Emerging Trends and Future Research Directions

A. Federated learning and privacy-preserving approaches

As organisations and industries increasingly collaborate to address evolving cyber threats, safeguarding sensitive data remains a critical concern. Federated learning, a paradigm in which models are trained across decentralised data sources without transferring raw information, offers a promising approach to protecting privacy while gaining collective intelligence[35]. Paired with secure multiparty computation and differential privacy techniques, federated learning enables the sharing of threat intelligence across financial institutions, healthcare providers and critical infrastructure operators without disclosing proprietary datasets or vulnerable system details[38]. Such cooperative models can capture broader threat patterns and improve detection rates against emerging attacks, while maintaining strict data governance and regulatory compliance. Additionally, adherence to privacy regulations, such as GDPR and CCPA, becomes more manageable, as federated learning inherently minimises data transfer risks[39].

B. Automated machine learning (AutoML)

AutoML frameworks aim to streamline the selection, tuning and deployment of ML models, thereby reducing the need for deep domain or data science expertise. By automating tasks, such as hyperparameter optimisation, feature selection and ensemble construction, AutoML can rapidly iterate through myriad configurations, converging on robust, well-calibrated solutions[40]. This efficiency not only accelerates the pace at which organisations deploy and update their cybersecurity models, but also enables smaller or resource-constrained security teams to benefit from sophisticated ML techniques. In effect, AutoML democratises access to ML-driven cybersecurity solutions, enhancing overall resilience and enabling continuous improvements in threat detection and response.

C. Transfer learning and domain adaptation

However, attacks are rarely confined to a single domain. Patterns identified in one industry, such as the lateral movement techniques observed in cloud infrastructure, may offer valuable insights for detecting similar tactics in industrial control systems or healthcare networks[2]. Transfer learning and domain adaptation techniques allow models trained on one dataset to be fine-tuned or adapted to another, conserving computational resources and reducing the need for extensive labelled data in new domains[41]. These approaches not only accelerate the deployment of ML solutions to emerging sectors but also help ensure that lessons learned from one field inform and strengthen defences elsewhere. As threat landscapes evolve, ML models capable of adapting their knowledge and recontextualising learned patterns are crucial for maintaining robust security.

D. Causal inference and complex system modelling

While many ML methods excel at pattern recognition, the next frontier involves understanding the cause-effect relationships underlying cyber-attacks. Causal inference methods allow practitioners to move beyond correlation-based modelling, providing insights into what triggers or influences certain attack patterns and how different defense strategies produce tangible outcomes[42]. By constructing complex system models, ML-driven cybersecurity solutions can simulate the interplay between attackers, defenders and evolving infrastructure, leading to more strategic planning and a better allocation of defensive resources.

These advancements have the potential to enhance predictive accuracy and improve strategic defense planning. For example, by understanding which infrastructure components or user behaviours are causally linked to increased vulnerability organisations can prioritise hardening those elements, resulting in more proactive and cost-effective defence measures. Ultimately, integrating causal inference techniques into ML-driven cybersecurity can yield more trustworthy, explainable and targeted defences that can adapt to changing adversary tactics.

7. Conclusion

This study has demonstrated the pivotal role that machine learning (ML) plays in strengthening cybersecurity strategies across multiple sectors. By enabling the detection of complex attack patterns and subtle anomalies that frequently elude traditional signature-based defences, ML techniques significantly enhance early threat detection and response capabilities[2,1]. Through case studies drawn from finance, healthcare, critical infrastructure and cloud environments, this review has illustrated the flexibility and domain adaptability of ML-based solutions, offering insights that transcend industry boundaries and inform broader digital defence strategies. Strategic integration of ML models into Security Information and Event Management (SIEM) systems, alongside careful alignment with regulatory frameworks and the deliberate curation of training data, has emerged as a best practice for deploying robust and scalable defences. These systems must also account for adversarial evolution, making continuous validation and adaptation essential for sustained efficacy.

From a research perspective, this synthesis reveals several urgent areas for advancement. Chief among them is the development of interpretable models that provide transparent decision-making pathways for both practitioners and regulators. Additionally, there is a pressing need for more resilient architectures capable of withstanding adversarial perturbations, as well as domain adaptation techniques that allow effective model transfer across operational contexts[43]. These challenges underscore the necessity of advancing explainable AI frameworks and adversarially robust methodologies that can evolve in parallel with the threat landscape. For practitioners, the findings offer concrete guidelines for implementation. Key recommendations include the deployment of continuously retrained models, the prioritisation of diverse and representative datasets and the adoption of privacy-preserving techniques that uphold user trust and regulatory compliance. Ensuring transparency through explainability not only enhances operational clarity but also contributes to stakeholder confidence in AI-driven security workflows.

Policymakers also have a vital role in shaping the future of ML in cybersecurity. Regulatory foresight, standard-setting and incentives for secure data sharing can help build a trustworthy foundation for AI adoption. The European Union's AI Act (Regulation (EU) 2024/1689), which came into force on 1 August 2024, exemplifies this regulatory momentum by imposing specific cyber resilience obligations on high-risk AI systems used in security monitoring. These include mandatory accuracy thresholds, robustness testing protocols and post-incident reporting requirements within 72 hours[44]. Notably, the Act mandates demonstrable human oversight and the integration of failsafe mechanisms, prompting system architects to implement dual control features that permit manual intervention in autonomous response scenarios[45]. Machine learning offers a transformative opportunity to construct intelligent, adaptive and resilient cybersecurity ecosystems. Through ongoing innovation, interdisciplinary collaboration and supportive policy environments, ML-driven security strategies can extend beyond reactive defences to enable proactive and pre-emptive resilience. As cyber threats continue to grow in scale and complexity, these intelligent systems are increasingly positioned to serve as the foundation of next-generation digital security, capable not only of mitigating today's risks but also of anticipating and addressing the threats of tomorrow.

8. References

1. Sommer R, Paxson V. Outside the closed world: On using machine learning for network intrusion detection. 2010 IEEE Symposium on Security and Privacy, 2010: 305-316.
2. Buczak AL, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 2016;18: 1153-1176.
3. Hesham M, Essam M, Bahaa M, et al. Evaluating predictive models in cybersecurity: A comparative analysis of machine and deep learning techniques for threat detection. MSA University 2024.
4. Moustafa N, Slay J. The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 dataset and the comparison with the KDD99 dataset. Information Security Journal: A Global Perspective, 2016;25: 18-31.
5. ENISA Threat Landscape 2024: July 2023 to June 2024. European Union Agency for Cybersecurity, 2024.
6. Shakya S, Abbas R. A comparative analysis of machine learning models for DDoS detection in IoT networks. Victoria University, 2024.
7. Afane K, Wei W, Mao Y, et al. Next-Generation Phishing: How LLM Agents Empower Cyber Attackers. In 2024 IEEE International Conference on Big Data (BigData), 2024: 2558-2567.
8. Tariq S, Chhetri BM. Alert fatigue in security operations centres: Research challenges and opportunities. ACM Computing Surveys, 2025;57: 1-38.
9. Lashkari AH, Draper-Gil G, Mamun M, et al. Characterization of tor traffic using time-based features. ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy, 2017: 253-262.
10. Edmund E, Enemosah A. AI and machine learning in cybersecurity: Leveraging AI to predict, detect and respond to threats more efficiently, 2024.
11. Tariq S, Baruwal Chhetri M, Nepal S, et al Alert fatigue in security operations centres: Research challenges and opportunities. ACM Computing Surveys, 2025;57: 1-38.
12. Nguyen TM, Reddi S. Deep reinforcement learning for cyber security. arXiv preprint, 2019.
13. Shiravi A, Shiravi H, Ghorbani AA. A survey of visualization systems for network security. IEEE Transactions on Visualization and Computer Graphics, 2011;18: 1313-1329.
14. Thorat S, Dari SS, Ahuja K, et al. Machine Learning-Driven Security Information and Event Management (SIEM). In International Conference on Frontiers of Intelligent Computing: Theory and Applications, 2024: 525-542.
15. Fogla P, Sharif M, Perdisci R, et al. Polymorphic blending attacks. In Proceedings of the 15th USENIX Security Symposium (USENIX Security 06) 2006: 241-256.
16. Liu FT, Ting KM, Zhou Z-H. Isolation Forest. 2008 Eighth IEEE International Conference on Data Mining, 2008: 413-422.
17. Ajala OA, Okoye CC, Ofodile OCA, et al. Review of AI and machine learning applications to predict and thwart cyber-attacks in real-time. Open Access Research Journal of Science and Technology, 2024.
18. Shad R, Olukemi A, Egon A. Zero-Day Attack Detection with Unsupervised Anomaly Detection, 2024.
19. Chukwunweike JN, Adewale AA, Osamuyi O. Advanced modelling and recurrent analysis in network security: Scrutiny of data and fault resolution. World Journal of Advanced Research and Reviews, 2024;23: 2373-2390.
20. Chandrasekaran M, Narayanan K, Upadhyaya S. Phishing email detection based on structural properties. NYS Cyber Security Conference, 2006: 1-7.
21. Akoglu L, Tong H, Koutra D. Graph-based anomaly detection and description: A survey. Data Mining and Knowledge Discovery, 2015;29: 626-688.
22. Phua C, Lee V, Smith K, et al. A comprehensive survey of data mining-based fraud detection research, 2010.

23. Monrose F, Rubin AD. Keystroke dynamics as a biometric for authentication. Future Generation Computer Systems, 2000;16; 351-359.

24. Visa. Visa boosts AI capabilities to further reduce fraud, Press release, 2025.

25. Villano M. At Mastercard, AI is helping to power fraud-detection systems. Business Insider, 2025.

26. Carlini N, Jagielski M, Choquette-Choo CA, et al. Poisoning web-scale training datasets is practical. In 2024 IEEE Symposium on Security and Privacy (SP), 2024: 407-425.

27. Chang H, Kodialam M, Lakshman TV, et al. MAGNet: machine learning guided application-aware networking for data centers. IEEE Transactions on Cloud Computing, 2021;11: 291-307.

28. Moustafa N, Creech G, Sitnikova E, et al. Big data analytics for intrusion detection system: Statistical decision-making using finite mixture models. Data Analytics and Decision Support for Cybersecurity, 2017: 45-66.

29. Traynor O. AI-powered phishing is on the rise: What to do? CybelAngel, 2025.

30. Sabin S. Future of Cybersecurity. Axios, 2025.

31. Das I. The impact of LLMs on cybersecurity: New threats and solutions. Qualys Blog, 2025.

32. Darktrace. Darktrace enhances Cyber AI Analyst with advanced machine learning for improved threat investigations, 2025.

33. Lakkaraju H, Bach SH, Leskovec J. Interpretable decision sets: A joint framework for description and prediction. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 1675-1684.

34. Lopez Perez R, Adamsky F, Soua R, et al. Machine Learning for Reliable Network Attack Detection in SCADA Systems, 2018.

35. Yang Q, Liu Y, Chen T, et al. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 2019;10: 1-19.

36. Hard A, Rao K, Mathews R, et al. Federated learning for mobile keyboard prediction, 2018.

37. Odume BW, Akintola AS, Nzenwa C. Regulating AI in cybersecurity: Challenges and opportunities, 2024.

38. He X, Zhao K, Chu X. AutoML: A survey of the state-of-the-art. Knowledge-Based Systems, 2021;212: 106622.

39. Pan SJ, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010;22: 1345-1359.

40. Pearl J. Causality: Models, reasoning and inference. Cambridge University Press, 2009.

41. Wang Y, Hazimeh H, Ponomareva N, et al. DART: A Principled Approach to Adversarially Robust Unsupervised Domain Adaptation. In 2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML), 2025: 773-796.

42. European Commission. Regulatory framework: AI Act, 2024.

43. White & Case. AI-Watch global regulatory tracker - European Union, 2024.