Advances in Artificial Intelligence and Machine Learning; Research 5 (3) 4053-4073 Received 23-04-2025; Accepted 17-07-2025; Published 24-07-2025

The AIM-PRISM Framework: A Novel Strategic Model for Machine Learning and Artificial Intelligence Deployment in National Infrastructure Cybersecurity

Mansoor G. Al-Thani mbgalthani87@gmail.com
Technical Affairs Department, Ministry of Interior,
Qatar

Corresponding Author: Mansoor G. Al-Thani

Abstract

The increasing intricacy and prevalence of online threats, growing complexity and frequency of cyber threats, particularly those targeting energy grids, transport systems, and financial platforms, necessitate a holistic approach to integrating intelligent technologies. This research proposes the AIM-PRISM framework, a strategic and adaptable model for deploying Artificial Intelligence (AI) and Machine Learning (ML) in cybersecurity for national infrastructure protection. While significant advancements have been made in incident response, AI-driven risk detection, and data protection, a unified deployment strategy is still lacking. Building on an extensive literature review, we identify key technological developments and implementation challenges and synthesize them into a novel eight-component framework: Adaptability, Integration, Monitoring, Predictive capacity, Responsiveness, Inclusivity, Security, and Meaningful interpretation (AIM-PRISM). This framework addresses operational, ethical, and governance considerations, offering a structured guide for policymakers, engineers, and organizational leaders. The research illustrates the framework's application through real-world-inspired scenarios and presents criteria for evaluating AI/ML deployment readiness across infrastructure sectors.

Keywords: Cybersecurity, Artificial intelligence, Machine learning, Critical infrastructure, Adaptive security, Threat detection, Incident response, Resilience.

1. INTRODUCTION AND LITERATURE BACKGROUND

National infrastructure is becoming more vulnerable to cybersecurity threats as a result of our growing reliance on digital technology and interconnected systems. Cyberattacks are increasingly targeting critical industries like energy and power grids [1], transportation (e.g., smart traffic systems and airports) [2], water supply [3], communication networks [4], and financial systems [5]. Successful attacks on these infrastructures can have serious repercussions, including the potential to impair public safety, interrupt vital services, and cause large financial losses.

Citation: Mansoor G. Al-Thani. The AIM-PRISM Framework: A Novel Strategic Model for Machine Learning and Artificial Intelligence Deployment in National Infrastructure Cybersecurity. **Advances in Artificial Intelligence and Machine Learning**. 2025;5(3):228.

Cyber dangers have become more complicated and widespread. Besides, state-sponsored actors and cybercriminals are using increasingly advanced tactics. These days, attacks like ransomware, phishing, and Advanced Persistent Threats (APTs) are more specialized, covert, and able to get past traditional security measures [6]. This quickly evolving threat landscape makes it challenging for traditional rule-based cybersecurity systems to control, necessitating the development of more intelligent, adaptive, and predictive solutions. AI/ML approaches are being included in cybersecurity frameworks for infrastructure protection more and more, from anomaly detection in network traffic to predictive modeling for attack vectors [7].

A comprehensive review of the existing literature reveals that AI/ML applications in cybersecurity can be grouped into five main domains: (1) threat detection and prevention, (2) incident response and automation, (3) predictive analysis and risk assessment, (4) data protection and privacy, and (5) adaptive security mechanisms. Technologies such as anomaly detection systems [8], SOAR platforms [9], explainable AI (XAI) [10], and federated learning (FL) [11], are rapidly evolving (**TABLE 1**). Yet challenges persist, ranging from data availability and adversarial AI to integration difficulties and legal/ethical constraints. Most notably, current studies highlight fragmented implementation and the absence of a strategic framework that unifies technical capability with deployment strategy.

Table 1: Comparative Analysis of SOAR and Related Cybersecurity Platforms.

Platform	Primary Function	AI/ML Role	Strengths	Limitations
SIEM	Log collection and event correlation	Analyzes patterns in log data	Centralized visibility	High false positives, reactive
Threat Intelligence Feed	Provides threat indicators and context	Feeds predictive models with up-to-date threat data	Global threat awareness	May produce overwhelming data
SOAR	Orchestrates tools and automates incident response	Uses ML to prioritize incidents and automate decisions	Workflow automation and speed	Complex setup and integration
Automated Response	Executes predefined actions (e.g., block IP)	Can be automated through rules or AI recommendations	Fast containment	Limited to known patterns/actions
SOC Analyst	Investigates, validates, and escalates alerts	Supports AI-assisted investigations and triage	Human expertise and decision-making	Fatigue, slow for large volumes
SOAR (Detailed)	Centralizes security tools, automates incident workflows, provides playbooks, and enables response orchestration	Leverages ML to correlate alerts, assess severity, suggest actions, and reduce analyst workload through intelligent automation	Reduces mean time to detect/respond (MTTD/MTTR), enables consistent responses, integrates diverse tools, scalable automation	Requires careful rule/playbook design, high initial setup effort, may depend on quality of integrated data sources

1.1 Threat Detection and Prevention

AI and ML have transformed cybersecurity by improving real-time threat detection and response [12]. These technologies excel at analyzing large volumes of data, such as network traffic and user behavior, to identify anomalies that may signal cyber threats [13]. Their ability to detect attacks early is particularly crucial in protecting national infrastructure, where delays can have devastating consequences. Traditional cybersecurity systems often rely on static, rule-based approaches, which struggle to detect new or evolving threats. In contrast, AI and ML can identify subtle, complex patterns that indicate malicious activity [14]. For instance, they can spot spikes in data transfer or unusual login behavior, allowing organizations to mitigate potential Distributed Denial of Service (DDoS) attacks or data breaches before damage occurs [15, 16].

Deep learning models, including convolutional and recurrent neural networks, have advanced malware detection [17]. Unlike signature-based systems that only identify known threats, these models can detect zero-day malware by analyzing behavioral patterns and file characteristics. This is especially effective against sophisticated threats such as APTs, which often deploy custom malware to evade detection [18]. Behavioral analytics is another area where AI and ML are highly effective. These tools monitor user activity and detect deviations from established norms, helping identify insider threats or compromised accounts. For example, if an employee accesses sensitive data at unusual times or downloads excessive files, the system flags this behavior as suspicious.

Platforms like Darktrace use unsupervised learning to detect such anomalies in real-time, improving responsiveness and reducing the risk of insider attacks. AI has also enhanced threat intelligence through platforms that analyze vast, unstructured data sources such as dark web forums, social media, and threat databases. Using natural language processing (NLP), these platforms can detect emerging vulnerabilities or malicious activity before they escalate [19]. For example, if a new exploit is discussed in hacker communities, AI can identify it and alert organizations to apply security patches promptly. With three distinct steps, Cascavilla et al (2022) [19], attempted to expand on earlier work done in the topic of classifying criminal acts. First, they assembled 113995 onion sites and dark marketplaces into a diverse dataset. After that, they contrasted pre-trained transferable models with more conventional text classification techniques. Lastly, they created two methods for classifying unlawful activities: one for dealing with Dark Web illicit content and another for determining the particular kinds of narcotics. Bert was able to classify the general material of the dark web and the different sorts of drugs with around 96% and 92% accuracy, respectively, demonstrating that he had the best method [19].

A real-world example of AI's effectiveness is to be deployed in financial institutions [20]. An AI-based threat detection system identifies and prevents an APT attack that had bypassed traditional defenses. The AI system detected abnormal data exfiltration patterns and reduced false positives, allowing security teams to focus on real threats. AI-driven systems are inherently adaptive, continuously learning from new data and growing threats. Unlike static security tools, they refine their models as they encounter novel attacks, making them better suited for dynamic environments like national infrastructure. However, challenges remain. High-quality data is necessary for training effective models, but many organizations face limitations due to data silos, privacy concerns, or insufficient datasets [21]. Moreover, AI systems must distinguish genuine threats from false positives, which is a complex task in dynamic networks. The rise of adversarial AI, where attackers design

inputs to fool detection systems, also poses a new risk. Researchers are responding by developing robust models and using techniques like adversarial training to improve system resilience [22].

1.2 Incident Response and Automation

The automation of incident response processes through AI has revolutionized the way organizations handle security breaches. By leveraging AI-driven systems, cybersecurity teams can rapidly identify, prioritize, and remediate threats, significantly lowering response times and minimizing the potential damage caused by cyberattacks [23]. This shift from manual to automated incident response is particularly critical in the context of national infrastructure protection, where delays in responding to threats can have severe consequences for public safety, economic stability, and national security. AI-driven incident response systems excel at analyzing vast amounts of data in real-time, enabling organizations to detect and respond to threats more efficiently than traditional methods [24].

One of the main advantages of AI in this domain is its ability to prioritize threats based on severity. For example, an AI system can analyze the characteristics of a detected threat, such as its origin, target, and potential impact, and assign it a risk score [25]. This allows cybersecurity teams to assign resources more effectively, concentrating on high-priority threats while automating the response to lower-risk incidents. By streamlining this process, AI not only accelerates response times but also ensures that critical threats are addressed before they can cause significant harm.

Recent innovations in AI and ML have further enhanced the capabilities of incident response systems. One of the most significant developments is the emergence of "Security Orchestration, Automation, and Response (SOAR)" platforms [26]. These platforms integrate AI to automate complex incident response workflows, allowing organizations to react to threats at machine speed. For instance, when a security breach is detected, a SOAR platform can automatically isolate infected machines, block malicious IP addresses, and generate detailed incident reports—all without human intervention [27]. This level of automation is particularly valuable in large organizations, where the volume of security alerts can overwhelm even the most experienced cybersecurity teams. SOAR solutions allow human analysts to concentrate on more intricate and strategic facets of incident response by automating repetitive work.

Another area where AI has made significant strides is "automated threat hunting" [28, 29].

AI has also transformed the field of "real-time forensics", enabling organizations to analyze large volumes of data in real-time to identify the root cause of a breach [30]. Traditional forensic investigations often involve manually analyzing logs and system activities, which can take days or even weeks. In contrast, AI-powered forensic tools use ML algorithms to correlate events across multiple systems, providing a comprehensive view of the attack in real-time. If a data breach occurs, an AI forensic tool can analyze network traffic, system logs, and user activities to identify how the attacker gained access, what data was compromised, and whether the attack is ongoing. This real-time analysis enables organizations to respond more effectively, containing the breach and preventing further damage.

For example, Rzwanbasha and Annamalai (2024) [30], reported that the time needed for important forensic activities can be cut by up to 93% when AI is used in these technologies, with the metadata

extraction process taking two hours instead of ten. Here, the improvements were evident: facial recognition increased from 70 to 88 percent, while video object detection increased from 65 to 90 percent. The manual complexity was greatly decreased by these automation features. It reduced the aspect of feature extraction by 80% and the output of reports by 88%. Additionally, it was noted that the legal admissibility of AI-generated evidence has been improved, and statistics indicate that the average admissibility of predictive analytics, which was previously at 70%, has increased to 85% [30].

1.3 Predictive Analysis and Risk Assessment

AI has also been a game-changer in the fields of risk assessment and predictive analysis, allowing businesses to foresee possible cyber threats before they become true [31]. Organizations can take proactive steps to improve their security posture by using AI systems to analyze past data and spot trends that might provide them with important insight into new threats. This capability is particularly critical for protecting national infrastructure, where the consequences of a successful cyberattack can be catastrophic, ranging from widespread service disruptions to significant economic and public safety impacts.

AI's capacity to process and analyze enormous volumes of data at previously unheard-of speeds is the foundation of its predictive powers [32]. Static risk models and manual study of past accidents are common components of traditional risk assessment techniques, which can be laborious and have a limited capacity to adjust to emerging risks. AI-driven systems, on the other hand, can continuously examine data from several sources, including user activity, threat intelligence feeds, and network logs, to spot patterns and trends that might point to an imminent attack. For instance, an AI system can notify cybersecurity teams of the potential for an impending assault and allow them to take preventive action if it notices a sharp rise in reconnaissance activity directed at a certain industry.

One of the most substantial advancements in this area is "predictive threat modeling", where AI models analyze historical attack data to predict future attack vectors [33]. By identifying trends and correlations in past incidents, AI can predict the probability of specific types of attacks, such as ransomware or Distributed Denial of Service (DDoS) attacks. For instance, if a particular vulnerability has been exploited in multiple recent attacks, an AI system can predict that similar vulnerabilities in other systems may also be targeted. This foresight allows organizations to prioritize patching and other mitigation efforts, reducing the risk of a successful attack. Predictive threat modeling is particularly valuable in the context of national infrastructure, where attackers often target known vulnerabilities in critical systems. In order to close this gap, Hoseini et al. (2024) [33], suggested a quick, easy, and time-saving method for assessing possible attacks and their security concerns using the attack tree and a risk analysis technique in the field of adversarial machine learning (AML). Assessing the danger associated with each node in an attack tree is one of the most crucial phases in figuring out the attack's overall risk. This study also outlines a methodical methodology that includes characterizing the system architecture and identifying its assets under different operating environment situations. Security professionals may also benefit greatly from this method, which can help them comprehend and reduce possible risks and analyze risk in AML systems [33].

Another key advancement is the development of "AI-driven risk scoring systems", which assess the likelihood of a cyberattack based on a combination of factors, including system vulnerabilities, threat intelligence, and user behavior [34]. These systems provide a quantitative measure of risk, enabling organizations to prioritize their security investments more effectively. Practically, a risk scoring system might assign a high-risk score to a system that has multiple unpatched vulnerabilities, a history of being targeted by attackers, and a high volume of suspicious user activity. By focusing on high-risk systems, organizations can distribute their resources more effectively, ensuring that the most significant assets are protected.

AI has also revolutionized the field of "scenario analysis", enabling organizations to simulate various attack scenarios and assess their potential impact on critical infrastructure [35]. Traditional scenario analysis often involves manual simulations, which can be time-consuming and limited in their scope. AI-powered technologies, on the other hand, can quickly model thousands of attack scenarios while accounting for variables like attacker behavior, network architecture, and system configurations.

These simulations provide valuable insights into potential vulnerabilities and the likely impact of different types of attacks, helping organizations develop robust contingency plans. For example, an AI-powered scenario analysis tool might simulate a ransomware attack on a power grid, identifying critical systems that could be affected and recommending measures to mitigate the impact [36]. A compelling case study that highlights the effectiveness of AI in predictive analysis and risk assessment involves a national power grid operator that used AI to predict and prevent a potential cyberattack on its control systems. The operator's AI system analyzed historical attack data and identified a pattern of reconnaissance activities targeting the grid. These activities included repeated scans of the grid's control systems and attempts to exploit known vulnerabilities. Based on this analysis, the AI system predicted that an attack was imminent and alerted the operator's cybersecurity team. The team was able to implement additional security measures, such as patching vulnerabilities and enhancing monitoring, before the attack could be launched. As a result, the operator successfully prevented a potential blackout, safeguarding the stability of the national power grid.

Despite these advancements, there are still challenges that need to be addressed to fully realize the potential of AI in predictive analysis and risk assessment. One of the key challenges is the quality and availability of data [37]. AI systems rely on large volumes of high-quality data to train their models, but in many cases, organizations may not have access to sufficient data or may face challenges related to data silos and privacy concerns [38]. Additionally, the effectiveness of AI systems depends on their ability to distinguish between genuine threats and false positives, which can be difficult in complex and dynamic environments [39]. Another challenge is the increasing use of adversarial AI by cybercriminals. Adversarial AI refers to the use of AI techniques to deceive or manipulate AI systems [40]. More specifically, attackers may use adversarial machine learning to create malware that can evade detection by AI systems or to launch attacks that exploit vulnerabilities in AI models.

1.4 Data Protection and Privacy

AI enhances data protection by improving encryption methods and access controls, which are critical components of any cybersecurity strategy [41]. Traditional encryption methods, while effective,

often struggle to keep pace with the increasing sophistication of cyberattacks. AI-driven encryption techniques, such as homomorphic encryption [42]. Homomorphic encryption permits data to be processed without being decoded, meaning that sensitive information can remain secure even while it is being used for computations. This technology is particularly valuable in cloud environments, where data is often processed and stored across multiple servers. By enabling secure data processing, AI-driven encryption helps organizations protect sensitive information from unauthorized access, even in complex and distributed environments.

Another area where AI has made significant contributions is in the development of privacy-preserving machine learning techniques [43]. Traditional ML models often require large amounts of centralized data to train effectively, which can pose significant privacy risks. Techniques like federated learning and differential privacy address this challenge by enabling ML models to be trained on decentralized data without compromising user privacy [44]. Multiple parties can work together to train a model using federated learning without disclosing their raw data, and differential privacy makes sure that specific data points cannot be deduced from the model's output.

These techniques are increasingly being used in sectors such as finance and healthcare, where data privacy is of critical sensitivity and importance. For example, a hospital network might use federated learning to train a diagnostic AI model on patient data from multiple hospitals without ever transferring the data to a central server, thereby preserving patient privacy.

Chen et al. (2020) [43], introduce a novel framework aimed at mitigating information leakage in federated learning (FL) by leveraging the principles of differential privacy (DP). Their approach, referred to as *Noising before Aggregation Federated Learning (NbAFL)*, involves injecting artificial noise into client-side parameters prior to model aggregation. The authors formally demonstrate that NbAFL can achieve differential privacy guarantees by appropriately adjusting the variance of the added noise to correspond with different privacy levels. Furthermore, they established a theoretical bound on the convergence of the FL model's loss function under the NbAFL scheme. The analysis highlights three key insights: (1) a fundamental trade-off exists between privacy and convergence—enhanced convergence generally comes at the cost of reduced privacy protection; (2) for a fixed privacy level, increasing the number of participating clients NNN can lead to improved convergence performance; and (3) for a given privacy budget, there exists an optimal number of aggregation rounds that maximizes convergence efficiency [43].

AI has also revolutionized the field of Data Loss Prevention (DLP), which focuses on monitoring and controlling the flow of sensitive information within an organization [45]. Traditional DLP systems often rely on predefined rules and signatures to detect unauthorized data transfers, which can be limited in their ability to adapt to new threats. In contrast, AI-powered DLP systems use machine learning to analyze data flows and identify patterns indicative of data exfiltration [46]. For example, if an employee suddenly attempts to download a large volume of sensitive files or transfer data to an external device, the AI system can flag this activity as suspicious and automatically block the transfer. By continuously learning from new data, AI-driven DLP systems can adapt to evolving threats and provide more robust protection against data breaches. Sabir et al. (2021) [47], conducted a review to highlight existing research gaps in ML-based countermeasures against data exfiltration. Through thier analysis, the authors were able to categorize ML techniques used in these defenses as either data-driven or behavior-driven; group features into six main categories—behavioral, content-based, statistical, syntactical, spatial, and temporal; classify the types of evaluation datasets employed as simulated, synthesized, or real-world; and identify 11 commonly used

performance metrics across the studies. Based on their findings, the research recommended a greater focus on integrating data-driven and behavior-driven methods; the need for developing large, high-quality evaluation datasets; the adoption of incremental training techniques for ML models in defensive mechanisms; taking proactive consideration of adversarial robustness to mitigate the risk of poisoning attacks; and the promotion of automated feature engineering to enhance the efficiency of detecting data exfiltration threats [47].

A compelling case study that highlights the effectiveness of AI in data protection is the use and implementation in healthcare systems [48]. In order to automatically detect and eliminate identifying information, the authors developed a novel GPT4-enabled de-identification framework. It showed high accuracy and exceptional dependability in concealing private information from the unstructured medical text while maintaining the text's original structure and meaning when compared to other widely used medical text data de-identification techniques [48].

The integration of AI into data protection and privacy has also led to the development of more adaptive and proactive cybersecurity systems. Unlike traditional systems, which often rely on reactive measures, AI-driven systems can continuously monitor for threats and respond to them in real-time. For example, if a new type of data exfiltration technique is detected, an AI system can update its models to recognize similar threats in the future and automatically implement countermeasures. This adaptability is particularly important in the context of national infrastructure, where the threat landscape is constantly evolving, and attackers are increasingly using AI to develop more sophisticated attacks. A study used a simulation-based approach, combining reinforcement learning frameworks like Deep Q-Learning with synthetic data for zero-day threat simulations and datasets like CICIDS2017. The findings demonstrate that adaptive algorithms outperformed static models by achieving 94.8% detection accuracy, 54.5% fewer false positives, and 53.1% faster response times. Furthermore, in attack scenarios that were simulated, the adaptive systems showed an exceptional ability to recognize new threats [49].

Another challenge is the increasing use of adversarial AI by cybercriminals. Adversarial AI refers to the use of AI techniques to deceive or manipulate AI systems. For example, attackers may use adversarial machine learning to create malware that can evade detection by AI systems or to launch attacks that exploit vulnerabilities in AI models. To counter this threat, researchers are developing more robust AI models that are resistant to adversarial attacks. Techniques such as adversarial training, where AI models are trained on data that includes adversarial examples, are being explored to improve the resilience of AI systems.

1.5 Adaptive Security Mechanisms

Adaptive security mechanisms represent a paradigm shift in cybersecurity, moving beyond static rule-based systems toward dynamic models capable of evolving in response to evolving threats. Contrary to traditional security protocols that rely on pre-configured responses, adaptive systems leverage AI and ML to continuously assess risk, detect anomalies, and reconfigure defenses in real time. These mechanisms are particularly vital in safeguarding critical national infrastructure, where the threat landscape is characterized by complexity, velocity, and adversarial innovation.

Adaptive security can be identified as the capability of a system to autonomously adjust its protection levels based on contextual risk assessment and behavioral analysis [50]. This is made possible through the integration of self-learning algorithms, behavioral modeling, and predictive analytics. Such systems monitor network activity continuously, establish baselines for normal behavior, and flag abnormalities that may indicate the presence of malicious actors or zero-day exploits. As noted by Khayat et al. (2025) [51], adaptive mechanisms can also support decision-making in security operations centers (SOCs) by prioritizing incidents based on dynamic threat scoring models [51].

Rapid AI integration into vital industries has exposed a complicated web of cybersecurity issues specific to these cutting-edge technologies. Due to their complicated algorithms and wide-ranging data dependencies, AI systems are vulnerable to a wide range of cyberthreats that could compromise their integrity and impair their functionality [52]. The authors carefully considered these risks, which include adverse attacks, data poisoning, and systemic weaknesses brought on by the AI's infrastructure and operating frameworks. This research explores a comprehensive framework for creating and implementing trustworthy AI systems to address the limitations of current approaches. In order to increase the resilience of AI systems, this approach places a strong emphasis on creating dynamic, adaptive security mechanisms that may change in reaction to fresh and emerging cyberthreats. The study also discusses the ethical aspects of AI cybersecurity, emphasizing the necessity of tactics that safeguard systems while maintaining user privacy and promoting equity in all business dealings. This study examines future directions in AI cybersecurity in addition to existing tactics and ethical issues [52].

One of the most widely adopted approaches in adaptive cybersecurity is anomaly detection using unsupervised learning [53]. Algorithms such as Isolation Forests, autoencoders, and clustering models enable systems to identify patterns that diverge from expected norms without requiring labeled attack data [54]. In critical infrastructure environments—such as energy grids or transportation networks—this capability is crucial for early warning against stealthy, persistent threats. Recent advancements in reinforcement learning (RL) have further enriched the adaptive security domain [55]. The authors developed an attack-defense model for networks using finite random games that reflects the dynamic adversarial aspect of the attack-defense process. The autonomous defensive agent is made to create and modify defense methods on its own by utilizing reinforcement learning techniques. A network attack agent is created in order to improve the defense capabilities. In order to tackle the issue of environment instability brought on by the separation of observation space and action space among several offensive and defensive agents that coexist in the same environment, a synchronized interactive training mechanism is also proposed. This mechanism is inspired by MINIMAX Q-learning. The usefulness of the suggested approach in automated attack-defense scenarios is confirmed by experimental simulations, which also examine the generalization capacity in networks with varying sizes [55].

However, the deployment of adaptive security in real-world infrastructure also raises challenges. Issues such as false positives, model drift, adversarial inputs, and the lack of explainability in AI decisions must be addressed to maintain system trustworthiness. Moreover, integration with legacy systems and regulatory frameworks often limits the pace of adoption. Despite these limitations, adaptive security mechanisms remain at the forefront of AI-driven cyber defense. Their ability to adjust to fluid threat conditions, prioritize high-risk anomalies, and automate response workflows positions them as essential components of next-generation infrastructure protection frameworks.

Despite these advances, the implementation of AI/ML technologies often lacks a cohesive strategy that balances innovation with operational compatibility, ethical considerations, and system resilience. This research addresses that gap by proposing the AIM-PRISM framework, a structured, strategic model for guiding AI/ML adoption in cybersecurity, particularly within national critical infrastructure domains.

2. THE AIM-PRISM FRAMEWORK

The AIM-PRISM framework (**FIGURE 1**) is proposed to offer a unified strategy for the integration of AI/ML technologies into national infrastructure cybersecurity. It is composed of the following interconnected components:

1. **Adaptabilit**y: Ensures continuous learning and system evolution in response to new threats.
2. **Integration:** Focuses on compatibility with legacy systems and minimizing operational disruption.
3. **Monitoring:** Supports real-time anomaly detection and behavioral surveillance.
4. **Predictive Capacity:** Enables threat modeling and scenario forecasting based on historical patterns.
5. **Responsiveness:** Facilitates automated incident response via SOAR and EDR tools.
6. **Inclusivity:** Emphasizes privacy-preserving ML, federated learning, and equitable system design.
7. **Security:** Prioritizes robustness against adversarial AI and data poisoning.
8. **Meaningful Interpretation:** Incorporates explainable AI to enhance transparency and trust.

3. APPLICATION SCENARIOS OF THE AIM-PRISM FRAMEWORK

To demonstrate the practical applicability of the AIM-PRISM framework, we present three illustrative scenarios across critical national infrastructure domains: energy, transportation, and finance. These use cases are designed to showcase how various components of the framework can guide implementation strategies, even in environments constrained by legacy systems and data privacy regulations.

3.1 Energy Sector: Securing National Power Grids

Modern power grids are becoming increasingly digitized and interconnected, which makes them more vulnerable to sophisticated cyber threats such as APTs, zero-day vulnerabilities, and ransomware. The AIM-PRISM framework offers a comprehensive approach to securing such critical infrastructure. Adaptability is ensured through AI-enabled anomaly detection systems that

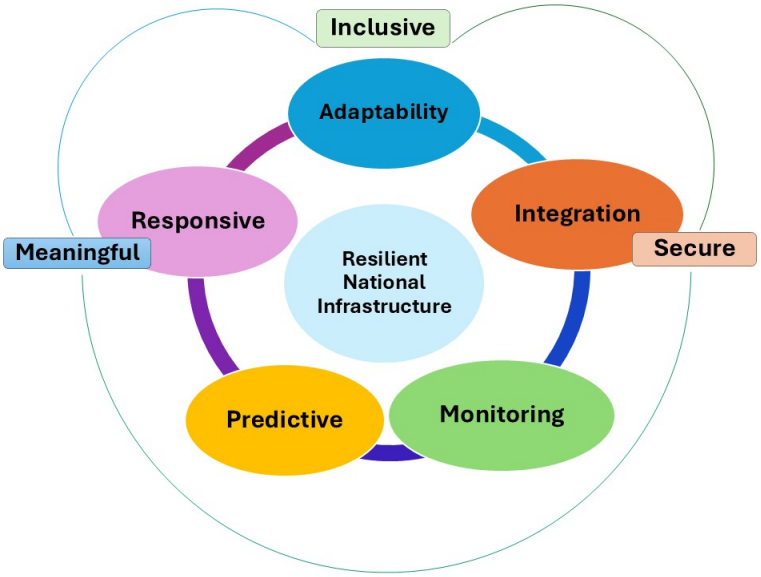

Figure 1: Scheme of the proposed AIM-PRISM Framework.

continuously evolve to recognize new attack patterns. Integration is achieved by embedding AI-based intrusion detection within SCADA systems without disrupting their core control protocols. Continuous monitoring is facilitated through federated learning, allowing real-time oversight of grid operations while preserving data privacy at the local level. Predictive analytics are employed to forecast cascading failures by analyzing historical outage and cyber incident data. Responsive mechanisms, such as SOAR platforms, enable the automatic isolation of compromised nodes to prevent system-wide disruption.

The framework also emphasizes inclusivity by training AI models on diverse grid configurations to prevent regional biases. To enhance security, adversarial training is used to fortify models against spoofing attacks. Finally, meaningful oversight is maintained through the use of explainable AI tools that assist grid operators in interpreting alerts and understanding system behavior. This integrated approach underscores the urgent need for resilient, AI-driven security systems in critical infrastructure, especially where public safety is at stake and legacy technologies impose strict operational constraints..

3.2 Transportation Sector: Smart Traffic Management and Aviation

The transportation sector increasingly depends on artificial intelligence for functions such as traffic optimization, surveillance, and autonomous control. Given its cyber-physical nature, this sector requires intelligent safeguards against threats like spoofing, signal manipulation, and denial-of-service attacks. The AIM-PRISM framework offers a structured approach to enhancing cybersecurity in such environments. Adaptability enables the detection of emerging attack vectors targeting autonomous transport systems, while integration ensures that AI-enhanced surveillance technologies, such as facial recognition, are seamlessly connected with airport security infrastructures. Monitor-

ing leverages machine learning to identify real-time anomalies in traffic flow, allowing for swift intervention.

Predictive models anticipate congestion and system failures that may be initiated through cyber-attacks, enhancing preparedness. Responsive protocols empower traffic systems to autonomously reroute or initiate lockdowns when malicious signals are detected. An inclusive design philosophy ensures that AI models are trained on diverse geographic and demographic mobility data, preventing bias and increasing reliability. Secure development practices are embedded by designing AI models that are resistant to GPS spoofing and data injection threats. Finally, meaningful transparency is achieved through explainable AI, allowing both control center operators and the public to understand how decisions are made. This application of the AIM-PRISM framework illustrates its value in guiding the creation of adaptive, fail-safe transportation systems that secure both digital information and the physical movement of people and goods.

3.3 Financial Sector: AI-Augmented Fraud and Risk Management

Financial institutions are increasingly targeted by sophisticated cyber threats such as phishing attacks, data breaches, and AI-generated synthetic identities. The AIM-PRISM framework provides a structured approach to strengthening cybersecurity in this high-stakes sector. Adaptability is achieved through continuous learning models trained on patterns of user behavior and transaction history, enabling dynamic detection of anomalies. Integration ensures these models are embedded directly into mobile banking applications and fraud monitoring systems, facilitating seamless threat detection.

Monitoring tools operate in real time to flag unusual login attempts or suspicious fund transfers, while predictive components identify emerging fraud rings or concentrations of credit risk before they escalate. Responsive capabilities are built into the system to automatically freeze accounts or escalate high-risk activities for immediate review. Inclusivity is addressed by training AI models on a broad range of transaction data across diverse customer segments, helping to mitigate systemic bias. Secure operations are maintained through the use of federated learning techniques and encrypted transaction logs, ensuring that sensitive data remains protected. Finally, meaningful outputs are provided through explainable risk dashboards that offer transparent scoring and support compliance with regulatory standards. By adopting the AIM-PRISM framework, the financial sector can align AI-driven innovation with the rigorous demands of security, fairness, and regulatory transparency.

4. EVALUATION CRITERIA FOR AIM-PRISM IMPLEMENTATION

To support practical adoption, we propose an evaluation rubric that allows organizations to assess their preparedness and maturity across the eight dimensions of the AIM-PRISM framework. This model can serve as both a diagnostic tool and a roadmap for continuous improvement in AI-integrated cybersecurity systems.

4.1 Maturity Levels

Each dimension of AIM-PRISM can be evaluated on a four-level scale:

1. **Level 0:** Absent – No implementation or strategic planning exists.
2. **Level 1:** Initial – Ad hoc, experimental, or pilot applications are present.
3. **Level 2:** Developing – Defined strategies and partial implementations exist; integration is underway.
4. **Level 3:** Advanced – Full, operationalized, and continuously monitored deployment is in place.

4.2 Evaluation Matrix

AIM-PRISM Dimension	Evaluation Focus	Maturity Indicators
Adaptability	Use of self-learning AI models to respond to emerging threats	Dynamic threat model updates, automated retraining, adversarial resilience
Integration	Compatibility with legacy systems and cybersecurity workflows	Seamless data flow, Application Programming Interface (API) connectivity, minimal system disruption
Monitoring	Real-time analytics and behavioral tracking	Continuous anomaly detection, endpoint telemetry, system logs ingestion
Predictive	Forecasting of future risks using AI/ML	Predictive scoring, threat simulation, attack vector modeling
Responsive	Automation of incident handling	Use of SOAR platforms, Endpoint Detection and Response (EDR) integration, automated containment
Inclusive	Bias mitigation and diversity in model training	Representative datasets, fairness checks, federated learning approaches
Secure	Defense against adversarial attacks and model manipulation	Use of adversarial training, penetration testing, model explainability under stress
Meaningful	Explainability and compliance with ethics/legal frameworks	Integration of XAI tools, General Data Protection Regulation (GDPR) alignment, accountability audits

Organizations can complete a self-assessment across these dimensions using internal audits, expert interviews, or checklists. Each rating provides insight into areas of strength and those needing strategic investment.

4.3 Composite Readiness Score

A composite score can be derived by averaging the maturity level across all eight dimensions. This score provides a high-level benchmark that can:

1. Guide resource allocation
2. Support regulatory compliance reporting
3. Prioritize capability development

For example, a readiness score of 2.1 might indicate strong foundational systems but a need to mature in predictive modeling and ethical explainability before full-scale deployment.

5. DISCUSSION

The development of the AIM-PRISM framework addresses a significant gap in current AI-integrated cybersecurity strategies for national infrastructure: the absence of a holistic, evaluative, and implementation-ready model that aligns technical innovation with operational and ethical considerations. While existing literature and industry standards focus on discrete elements, such as anomaly detection, SOAR platforms, or adversarial robustness, few attempts have been made to synthesize these components into a unified framework that supports institutional decision-making and cross-sector adoption.

5.1 Bridging Fragmentation in AI-Cybersecurity Integration

Current AI adoption practices in cybersecurity are often fragmented, siloed, and reactive. Organizations may implement machine learning models for intrusion detection or deploy explainable AI for regulatory compliance without a coherent strategy linking these tools across the cybersecurity lifecycle. AIM-PRISM bridges this fragmentation by offering a multidimensional model that includes not only technical robustness (e.g., adaptability, security, and responsiveness) but also operational feasibility (e.g., integration and monitoring) and governance mechanisms (e.g., inclusiveness and explainability).

Unlike traditional maturity models (e.g., NIST's Cybersecurity Framework), which emphasize risk management and policy compliance, AIM-PRISM embeds AI-specific capabilities into the core of infrastructure defense strategies. This integration of AI characteristics with cyber defense requirements makes AIM-PRISM distinctively suited to 21st-century threats where attackers themselves may use AI to automate, evade, and scale their attacks.

5.2 Strategic Adaptability and Cross-Sector Relevance

The flexibility of AIM-PRISM allows it to be customized to various critical infrastructure sectors—energy, finance, transport—each with distinct threat profiles, regulatory obligations, and system architectures. For instance, while the energy sector may emphasize predictive grid stability and incident containment, the financial sector requires federated learning and zero-trust endpoint control. AIM-PRISM's layered structure supports this sectoral adaptability, enabling policy-makers and technical teams to co-develop AI strategies tailored to mission-critical environments.

This positions AIM-PRISM as more than a theoretical contribution; it becomes a strategic instrument for national cyber resilience planning. Its readiness assessment matrix, in particular, can guide ministries, regulators, and infrastructure providers in benchmarking capabilities, identifying risks, and allocating funding toward high-impact upgrades.

5.3 From Technical Focus to Governance-Aware Innovation

Another key distinction is AIM-PRISM's explicit attention to ethical, inclusive, and legal dimensions of AI deployment. While many technical reviews mention data privacy and adversarial attacks in passing, this framework places explainability, data equity, and regulatory alignment on equal footing with detection accuracy and automation. In doing so, it supports the growing policy imperative for trustworthy and human-centered AI systems, as outlined by bodies such as UNESCO, OECD, and the EU AI Act.

By structuring its components around measurable criteria, AIM-PRISM can also serve as a pre-assessment framework for AI certification or accreditation in cybersecurity tools—something that is increasingly needed as AI applications become mainstream in sensitive domains.

6. CONCLUSION AND FUTURE WORK

This research has introduced the AIM-PRISM framework, a comprehensive and adaptable strategic model for the integration of Artificial Intelligence and Machine Learning into cybersecurity systems protecting national infrastructure. Unlike conventional technical reviews or narrowly scoped deployment strategies, AIM-PRISM offers a multi-dimensional, actionable structure that combines algorithmic robustness with system integration, real-time monitoring, predictive intelligence, automated response, and ethical governance.

The framework bridges the existing gap between fragmented AI deployments and the need for institutional readiness, sector-specific adaptability, and policy-aligned implementation. Through its eight core dimensions, Adaptability, Integration, Monitoring, Predictive, Responsive, Inclusive, Secure, and Meaningful, it serves as both a diagnostic tool and a roadmap for organizations navigating the evolving threat landscape.

By grounding each component in operational realities and providing a maturity evaluation rubric, AIM-PRISM empowers organizations to benchmark their current capabilities, prioritize investment,

and align technical innovation with regulatory and ethical mandates. Its real-world application potential has been illustrated across the energy, transportation, and financial sectors, emphasizing its flexibility and strategic utility.

To further validate and evolve AIM-PRISM, several research and implementation avenues are proposed:

1. Expert Validation: Conduct structured interviews and Delphi studies with cybersecurity professionals, AI developers, and policy-makers to refine the framework's criteria and maturity levels.
2. Sector-Specific Case Studies: Apply AIM-PRISM retrospectively to analyze past cyber incidents in infrastructure sectors, demonstrating how the framework might have improved resilience or response.
3. Tool Development: Translate the framework into a self-assessment toolkit or dashboard for use by national security agencies, critical infrastructure providers, or certifying bodies.
4. Policy Integration: Collaborate with regional and international institutions (e.g., ITU, ENISA, Q-CERT) to align AIM-PRISM with emerging AI and cybersecurity governance frameworks.

In a world where AI is both a weapon and a shield, the AIM-PRISM framework offers a timely and necessary contribution to structuring the responsible and resilient adoption of AI for cybersecurity. By transforming a rich body of existing knowledge into a forward-looking model, this work advances the field from review to implementation.

6.1 Nomenclature

AI	Artificial Intelligence
AML	Adversarial Machine Learning
API	Application Programming Interface
DDoS	Distributed Denial of Service
DP	Differential Privacy
DPL	Deep Predictive Learning
EDR	Endpoint Detection and Response
EU	European Union
FL	Federated Learning
GDPR	General Data Protection Regulation
GPS	Global Navigation Satellite Systems
IP	Internet Protocol
ML	Machine Learning
MTTD	Mean Time to Detect
MTTR	Mean Time to Repair
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
OECD	Organization for Economic Co-operation and Development
RL	Reinforcement Learning
SOAR	Security Orchestration, Automation, and Response
SOC	Security Operations Center
UNESCO	United Nations Educational, Scientific, and Cultural Organization
XAI	Explainable Artificial Intelligence

6.2 Funding:

This research received no external funding.

6.3 Acknowledgments:

NA

6.4 Conflicts of Interest:

The authors declare no conflicts of interest.

References

[1] Soltan S, Yannakakis M, Zussman G. React to Cyber Attacks on Power Grids. IEEE Trans Netw Sci Eng. 2019;6:459-473.

[2] Koroniotis N, Moustafa N, Schiliro F, Gauravaram P, Janicke H. A Holistic Review of Cybersecurity and Reliability Perspectives in Smart Airports. IEEE Access. 2020;8:209802-209834.

[3] Adepu S, Mathur A. An Investigation Into the Response of a Water Treatment System to Cyber Attacks. In: 17th International Symposium on High Assurance Systems Engineering (HASE) IEEE 17th International Symposium on High Assurance Systems Engineering (HASE). IEEE. 2016:141-148.

[4] Buinevich M, Vladyko A. Forecasting Issues of Wireless Communication Networks. cyber resilience for an intelligent transportation system: an overview of cyber attacks. Information. 2019;10:27.

[5] Gulyás O, Kiss G. Impact of Cyber-Attacks on the Financial Institutions. Procedia Comput Sci. 2023;219:84-90.

[6] Alshamrani A, Myneni S, Chowdhary A, Huang D. A Survey on Advanced Persistent Threats: Techniques Solutions Challenges and Research Opportunities. IEEE Commun SurvTutorials. 2019;21:1851-1877.

[7] Ma Q, Sun C, Cui B, Jin X. A Novel Model for Anomaly Detection in Network Traffic Based on Kernel Support Vector Machine. Comput Sec. 2021;104:102215.

[8] M. H. Bhuyan, D. K. Bhattacharyya and J. K. Kalita, Network Anomaly Detection: Methods, Systems and Tools. IEEE Commun Surv Tutor. 2014;16:303–336.

[9] https://dione.lib.unipi.gr/xmlui/handle/unipi/14560

[10] Holzinger A, Saranti A, Molnar C, Biecek P, Samek W. Explainable AI Methods – A Brief Overview. In: AI international workshop Held in Conjunction with ICML Vienna Austria. Revised and extended papers. Cham: Springer International Publishing. 2022:13-38.

[11] Li L, Fan Y, Tse M, Lin KY. A Review of Applications in Federated Learning. Comput Ind Eng. 2020;149:106854.

[12] Danish M. Enhancing Cyber Security Through Predictive Analytics: Real-Time Threat Detection and Response. 2024. ArXiv Preprint: https://arxiv.org/pdf/2407.10864.

[13] Anh Huynh N, Keong Ng W, Ulmer A, Kohlhammer J. Uncovering Periodic Network Signals of Cyber Attacks. In: IEEE Symposium on Visualization for Cyber Security (VizSec). IEEE. 2016:1-8.

[14] Blauth TF, Gstrein OJ, Zwitter A. Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI. IEEE Access. 2022;10:77110-77122.

[15] de Neira AB, Kantarci B, Nogueira M. Distributed Denial of Service Attack Prediction: Challenges Open Issues and Opportunities. Comput Netw. 2023;222:109553.

[16] Zargar ST, Joshi J, Tipper D. A Survey of Defense Mechanisms Against Distributed Denial of Service (DDos) Flooding Attacks. IEEE Commun Surv Tutorials. 2013;15:2046-2069.

[17] Jha S, Prashar D, Long HV, Taniar D. Recurrent Neural Network for Detecting Malware. Comput Sec. 2020;99:102037.

[18] Do Xuan C, Dao MH. A Novel Approach for APT Attack Detection Based on Combined Deep Learning Model. Neural Comput Appl. 2021;33:13251-13264.

[19] Cascavilla G, Catolino G, Sangiovanni M. Illicit Darkweb Classification via Natural-language Processing: classifying Illicit Content of Webpages based on Textual Information. In: Proceedings of the 19th international conference on security and cryptography. SCITEPRESS - Science and Technology Publications. 2022;1:620-626.

[20] Deshpande A. Cybersecurity in Financial Services: Addressing AI-Related Threats and Vulnerabilities. In: International Conference on Knowledge Engineering and Communication Systems (ICKECS). IEEE. 2024:1-6.

[21] Mahalle A, Yong J, Tao X, Shen J. Data Privacy and System Security for Banking and Financial Services Industry Based on Cloud Computing Infrastructure. In: 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD) IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE. 2018:407-413.

[22] Vasan D, Hammoudeh M. Enhancing Resilience Against Adversarial Attacks in Medical Imaging Using Advanced Feature Transformation Training. Curr Opin Biomed Eng. 2024;32:100561.

[23] https://openurl.ebsco.com/contentitem/gcd:181692734?sid=ebsco:plink:crawler&id=ebsco:gcd:181692734

[24] Bompally SD. AI-Driven Incident Response for Digital Forensics and Incident Response: A Comprehensive Framework. J Comput Sci Technol Stud. 2025;7:467-472.

[25] Manoharan A, Sarker M. Revolutionizing Cybersecurity: Unleashing the Power of Artificial Intelligence and Machine Learning for Next-Generation Threat Detection. Int Res J Mod Eng Technol Sci. 2024;4:2151-2164.

[26] Kinyua J, Awuah L. AI/ML in Security Orchestration, Automation and Response: Future Research Directions. Intell Autom Soft Comput. 2021;28:527-545.

[27] https://www.proquest.com/docview/3101178417/abstract/C31E9DDB260B445DPQ/1

[28] Mahboubi A, Luong K, Aboutorab H, Bui HT, Jarrad G, et al. Evolving Techniques in Cyber Threat Hunting: A Systematic Review. J Netw Comput Appl. 2024;232:104004.

[29] Jain J. Artificial Intelligence in the Cyber Security Environment. In: N. Bhargava, R. Bhargava, P.S. Rathore and R. Agrawal, editors. Artificial intelligence and data mining approaches in security frameworks. Wiley. 2021:101-117.

[30] RizwanBasha A, Annamalai R. Transforming Crime Scene Investigations Through the Integration of Artificial Intelligence in Digital Forensics. In: IEEE International Conference on Communication Computing and Signal Processing (IICCCS). IEEE. 2024:1-6.

[31] Kalogiannidis S, Kalfas D, Papaevangelou O, Giannarakis G, Chatzitheodoridis F. The role of artificial intelligence technology in predictive risk assessment for business continuity: A case study of Greece. Risks. 2024;12:19.

[32] Rahmani AM, Azhir E, Ali S, Mohammadi M, Ahmed OH, et al. Artificial Intelligence Approaches and Mechanisms for Big Data Analytics: A Systematic Study. PeerJ Comput Sci. 2021;7:e488.

[33] Hoseini SV, Suutala J, Partala J, Halunen K. Threat Modeling AI/ML With the Attack Tree. IEEE Access. 2024;12:172610-172637.

[34] Javaid HA. Ai-Driven Predictive Analytics in Finance: Transforming Risk Assessment and Decision-Making. Adv Comput Sci. 2024;7:1-9.

[35] Shinkle GA, Gujarati C, Sharry P. Scenario Analysis in the AI Era: Redefining Human Involvement. Rochester NY: Social Science Research Network. 2025. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5239542

[36] Babu CV, Andrew Simon PA, Manohoran S. Ai-Powered Defenses Against Ransomware: Mitigating Emerging Threats to Critical Infrastructures. In: Kumar R, Peng SL, Jain P, Elngar AA, editors. Deep learning innovations for securing critical infrastructures. IGI Global Scientific Publishing. 2025:579-606.

[37] Cai L, Zhu Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Sci J. 2015;14:2.

[38] Whang SE, Roh Y, Song H, Lee JG. Data Collection and Quality Challenges in Deep Learning: A Data-Centric AI Perspective. VLDB J. 2023;32:791-813.

[39] Dalalah D, Dalalah OM. The False Positives and False Negatives of Generative AI Detection Tools in Education and Academic Research: The Case of ChatGPT. Int J Manag Educ. 2023;21:100822.

[40] Zbrzezny AM, Grzybowski AE. Deceptive Tricks in Artificial Intelligence: Adversarial Attacks in Ophthalmology. J Clin Med. 2023;12:3266.

[41] Akhtar ZB, Tajbiul Rawol AT. Enhancing Cybersecurity Through Ai-Powered Security Mechanisms. IT J Res Dev. 2024;9:50-67.

[42] Ahmad A, Rehman AU, Ghani MU, Nasim F, Naseem S. An In-Depth Comparative Analysis of Traditional vs Ai-Enhanced Encryption Algorithms. al-Aasar. 2025;2:294-305.

[43] Chen H, Hussain SU, Boemer F, Stapf E, Sadeghi AR, et al. Developing Privacy-Preserving AI Systems: The Lessons Learned. In: 57th ACM/IEEE Design Automation Conference (DAC). IEEE. 2020:1-4.

[44] Wei K, Li J, Ding M, Ma C, Yang HH, et al. Federated Learning With Differential Privacy: Algorithms and Performance Analysis. IEEE Trans Inf Forensics Sec. 2020;15:3454-3469.

[45] Abid N. Advancements and Best Practices in Data Loss Prevention: A Comprehensive Review. Glob J Univers Stud. 2024;1:190-225.

[46] Lal A, Prasad A, Kumar A, Kumar S. Data Exfiltration: Preventive and Detective Countermeasures. Rochester NY: Social Science Research Network. 2022:7. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4031852

[47] Sabir B et al. Machine Learning for Detecting Data Exfiltration: A Review. ACM Comput Surv. 2021;54:1:47.

[48] Liu Z, Huang Y, Yu X, Zhang L, Wu Z, et al. Deid-Gpt: Zero-Shot Medical Text De-identification by GPT-4. 2023. ArXiv preprint: https://arxiv.org/pdf/2303.11032

[49] Sudaryono S, Pratomo R, Ramadan A, Ahsanitaqwim R, Fletcher E. Artificial Intelligence in Predictive Cybersecurity: Developing Adaptive Algorithms to Combat Emerging Threats. J Comput Sci Technol Appl. 2025;2:1-3.

[50] Pekaric I, Groner R, Witte T, Adigun JG, Raschke A, et al. A Systematic Review on Security and Safety of Self-Adaptive Systems. J Syst Softw. 2023;203:111716.

[51] Khayat M, Barka E, Adel Serhani M, Sallabi F, Shuaib K, et al. Empowering Security Operation Center With Artificial Intelligence and Machine Learning – A Systematic Literature Review. IEEE Access. 2025;13:19162-19197.

[52] Roshanaei M, Khan MR, Sylvester NN. Navigating AI Cybersecurity: Evolving Landscape and Challenges. J Intell Learn Syst Appl. 2024;16:155-174.

[53] Choi WH, Kim J. Unsupervised Learning Approach for Anomaly Detection in Industrial Control Systems. Appl Syst Innov. 2024;7:18.

[54] Smolen T, Benova L. Comparing Autoencoder and Isolation Forest in Network Anomaly Detection. In: 33rd Conference of Open Innovations Association (FRUCT). IEEE. 2023:276-282.

[55] Ren B, Tang Y, Wang H, Wang Y, Liu J, et al. A Multiagent Deep Reinforcement Learning Autonomous Security Management Approach for Internet of Things. IEEE Internet Things J. 2024;11:25600-25612.