World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

AI-powered cybersecurity innovations for protecting U.S. critical infrastructure

Emma Junior Emmanuel *

Department of Computer Science, Prairie View A and M University, Texas. USA.

World Journal of Advanced Research and Reviews, 2026, 29(02), 081-092

Publication history: Received on 20 December 2025; revised on 01 February 2026; accepted on 03 February 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.2.0279

Abstract

The critical infrastructure sectors, such as the energy, transportation, healthcare, and communications sectors, are now exposed to sophisticated cyber-attacks that are the result of increased connectivity and complexities associated with the use of technology. The recent cyber-attacks that involved ransomware, advanced persistent threats, and zero-day exploits are testimony to the inadequacies associated with the use of conventional mechanisms used to secure critical systems. This is primarily due to the lack of adaptability and the ability to provide real-time intelligence that conventional mechanisms do not possess. Artificial Intelligence (AI), as an advanced form of cybersecurity, has been increasingly recognized for its ability to support threat detection and response. The present paper intends to identify the deficiencies in existing cybersecurity solutions and present an advanced cybersecurity framework using artificial intelligence for protecting the critical infrastructure in the United States. The cybersecurity framework combines machine learning for anomaly detection and deep learning for threat classification and automated responses for effective cybersecurity. Contributions of this paper include: (i) a study of the state of cyber threats against critical infrastructure in the United States of America, (ii) a systematic investigation of available AI technologies that can safeguard critical infrastructure, and (iii) developing a cybersecurity system that uses AI capable of detecting and forecasting cyber threats in real time. This system shows how AI can greatly improve critical infrastructure cybersecurity systems.

Keywords: Artificial Intelligence (AI); Cybersecurity; Critical Infrastructure; Machine Learning; Deep Learning; Reinforcement Learning; Intrusion Detection; Automated Response

1. Introduction

The nation's critical infrastructure is vital for national security, economic well-being, and welfare of the United States' citizenry. Industries like energy, transportation, healthcare, water services, communication, and financial services have increasingly used networked digital technologies, Industrial Control Systems, and cyber-physical systems for monitoring and decision-making. Although the increased use of networked digital technologies has brought about greater efficiency, it also entails an increased attack surface available for cyber assault [1, p. 390]. Recently, critical infrastructure attacks by cyber threats have been on the rise both in terms of occurrence and sophistication. Threat actors are using tactics like ransomware attacks, advanced persistent threats (APTs), supply chain attacks, and zero-day attacks to cause disruption to essential services and steal valuable data. The potential risks and impacts of such attacks do not only include financial losses but also public safety and national security threats among others.

Conventional cybersecurity methods like firewalls, intrusion detection systems, and traditional malware pattern recognition solutions are generally of a defensive nature, relying on predefined policies or attack signatures. For this reason, conventional cybersecurity solutions find it difficult to effectively identify new threats, respond to dynamically changing attack techniques, or function properly in complex topologies of large infrastructures [1, p. 391]. In addition,

* Corresponding author: Emma Junior Emmanuel

traditional manual methods of threat detection/response are not of much help when it comes to protecting complex critical infrastructures in real time.

However, these challenges can be overcome by the application of artificial intelligence. Artificial intelligence can make use of ML and analytics for the analysis of anomalies, learning from patterns, and responding accordingly [1, p. 389]. The research attempts to highlight the use of AI-based advancements in order to improve the safety and security of critical infrastructure in the USA. The major contributions of this research work include the analysis of challenges, analysis of the application of AI for the enhancement of cybersecurity, and the development of an AI-based framework.

2. Background and related work

Protecting critical infrastructure has emerged as one of the important focuses of cybersecurity research activities triggered by the growth of convergence between information technology (IT) and operational technology (OT). The critical infrastructure environment relies on a high level of availability, backed by legacy systems, which turns out to be highly susceptible to cyber-attacks [2, p. 32]. The current literature has already confirmed that the fundamental approaches applied for securing informatization do not properly address the critical infrastructure environment.

Recent work underlines the widening chasm between cyberattack complexity and conventional security protection capabilities and points out that research increasingly targets intelligent and adaptive security tools able to work within real time and react to dynamically changing events [2, p. 33]. The following section reviews cyber threat trends against critical US infrastructure and summarizes work relevant to developing cyber protection based on artificial intelligence innovation and research.

2.1. U.S. critical infrastructure cyber threat landscape

In the US, the number of cyber incidents within critical infrastructure sectors has increased over the last decade with the adoption of increased connectivity as IT and OT converge [3, p. 26]. Notable attacks on critical infrastructure sectors such as the power grid, energy pipelines, and hospitals are indicative of the possible disruption that cyber-attacks could potentially pose to the general safety of communities, as attacks on the energy system via cyber infiltration led to disruptions, while attacks on hospitals led to delayed care. These acts can be blamed on various threat actors, from state-driven enemies to cybercrime gangs, as well as insiders. CI cyber-attacks are usually driven by long-term goals, including the concealment of their activities, which can be the aim instead of immediate data exfiltration.

Typical attack methods used in the context of critical infrastructure involve phishing and social engineering attacks, attacks through unpatched vulnerabilities, the distribution and execution of malware and ransomware, and supply chain attacks. Furthermore, the inherent nature of industrial control systems makes them susceptible to attacks related to communication protocols, unchecked access through remote connectivity, and sensor data manipulation in cyber-physical systems [3, p. 27]. The deployment and adoption of Internet of Things (IoT) devices and remote access solutions further widen the threat sphere.

2.2. Traditional cybersecurity techniques and limitations

Traditional cyber defense measures have been used for quite some time to ensure the protection of information systems using tools like firewalls, intrusion detection and prevention systems (IDPS), antivirus software, and access control policies. These tools work on either predefined policies or signatures for the detection of malicious activity [4, p. 30]. These tools have shown efficacy in the traditional information technology setup for previously known malicious activities. Nonetheless, critical infrastructure environments face specific working challenges that highlight key shortcomings of traditional security strategies. Legacy infrastructure hardware and software applications commonly found in infrastructure networks were not intended to have a focus on cybersecurity and hence cannot be easily updated or patched frequently. Again, critical infrastructure integrates both IT systems and other operational technology applications that do not have strong support from traditional security software tools.

One major drawback of conventional cybersecurity infrastructure is that it is reactive in nature. Relying on signature-based detection, conventional cybersecurity infrastructure is inefficient against zero-day threats, and rule-based infrastructure faces difficulties in coping with changing approaches adopted by attackers. Additionally, conventional cybersecurity infrastructure tends to report high levels of alarms, which are time-consuming and overwhelming for security analysts to respond to [5, p. 502]. This is especially true in time-critical processes involving critical infrastructure. Another important challenge is scalability, as traditional approaches to analyzing massive amounts of varied data generated by today's critical infrastructure, such as network traffic, sensor data, and logs, don't work

effectively. However, the increasing levels of cyber threats in size and complexity make the need for intelligent and automated cyber security approaches that can safeguard America's critical infrastructures even more imperative.

2.3. AI in cybersecurity: existing research

However, recent studies have begun to investigate the ability to apply artificial intelligence in the realm of cybersecurity, in light of the limitations within current defense systems. Machine learning-based methods have been heavily utilized for intrusion detection, malware categorization, and anomaly identification. Supervised learning algorithms such as support vector machines, decision trees, and random forest learning algorithms have been utilized for network traffic classification and malicious activity identification using labeled datasets [6, p. 2417]. Unsupervised learning techniques such as clustering algorithms and statistical anomaly identification can also detect anomalies in system behavior, making them ideal for unknown threat identification.

Deep learning techniques have also furthered the development of AI-assisted cybersecurity systems by allowing complex features to be extracted from high-dimensional data. The structure of convolutional neural networks has been utilized for traffic analysis and visualization of malware, whereas recurrent neural networks and long short-term memory networks have proved successful in modeling attack behavior. More recent works have also explored the potential of attention models and transformer networks for higher accuracy rates of threat detection. Despite this progress, much work is still needed in this area. Currently, most AI solutions developed in the field of cybersecurity testing are carried out under constrained settings and not in real-world critical infrastructure scenarios [6, p. 2418]. Data quality, model generalization, and the concept of drift are still considered open research questions in the field of security research and developments. Meanwhile, the areas of explainability, adversarial robustness, and the integration of automated response systems are only marginally explored as research opportunities in this community of interest.

3. AI technologies for cybersecurity

The various technologies of artificial intelligence offer the means for developing intelligent and adaptable systems for defending against cyber threats. When considering the area of protecting critical infrastructures, artificial intelligence algorithms make possible the automated analysis of large amounts of varied data produced through network traffic, system logs, sensors, and control devices [7, p. 5600]. Through learning normal and malicious activity patterns, artificial intelligence systems make predictive protection against threats possible. Among different approaches in AI, machine learning has emerged as an essential component in cybersecurity. Machine learning algorithms can be trained to identify known threat patterns, detect anomalies that signal novel threats, and adjust to different system behaviors. These are particularly helpful in critical infrastructure environments where traditional rule-governed approaches tend to lack scalability.

3.1. Machine learning techniques

3.1.1. Supervised learning for threat classification

Techniques of supervised learning have found widespread applications in the realm of cybersecurity for tasks involving the classification of known threats. These methods make use of these labeled repositories of data containing illustrations of both harmless and malicious behaviors in order to learn distinguishing characteristics found within cyber threats. Some of the most commonly employed methodologies of supervised learning encompass support vector machines, k-nearest neighbors, decision trees, random forest classification, and gradient boosting. First, the benefit associated with supervised learning is the accuracy offered in identifying known types of attacks. But this method is very dependent on the availability of labeled data, which is not easily accessible in the critical infrastructure domain because such data is sensitive and some attacks may not frequently occur [8, p. 208].

3.1.2. Unsupervised learning for anomaly detection

Unsupervised learning methods can fill some gaps in supervised learning methods by detecting anomalies in normal system behavior without reliance on labeled samples. Unsupervised methods can be very effective in identifying zero-day attacks that have never been encountered or seen before. Popular unsupervised methods include cluster analysis techniques, principal component analysis, autoencoders, and statistical models of anomaly detection. Unsupervised learning is realized in the case of critical infrastructure. In such systems, unsupervised learning is utilized to detect unusual traffic and sensor data that indicate potential cyber threats [9, p. 4044]. Although unsupervised machine learning is useful in detecting unknown threats, it may produce false positives, which should be meticulously tuned to differentiate between both legitimate and malicious activities.

3.2. Deep learning models

There has been considerable interest in the application of deep learning models in the domain of cybersecurity research because of the potential of these models to learn hierarchical features from high-dimensional data without human interventions. Unlike traditional machine learning techniques, deep neural networks can learn relevant patterns directly from raw or lightly filtered input data, which is a desirable attribute in the context of detecting sophisticated cyber threats [10, p. 102749].

3.2.1. Neural networks for intrusion detection

The application of Artificial Neural Networks (ANNs) can be seen in the field of Intrusion Detection Systems as a model of non-linear relationships of the network traffic and system activity data. The neural network-based Intrusion Detection System can reach high accuracy of detection on the assumption of knowledge of typical and malicious behavior patterns due to the ability of ANNs to learn patterns of large and diverse sources of information [10, p. 102749]. But neural network-based intrusion detection is computationally intensive and requires vast amounts of training data, which can be challenging in real-time implementations within critical infrastructure environments where resource constraints are a concern.

3.2.2. CNNS, RNNS, and transformer-based models in cybersecurity

The application of Convolutional Neural Networks (CNNs) was successfully implemented for cybersecurity-related purposes dealing with structured and spatial data representation for instances like traffic flow matrices and malware samples represented as image formats. The primary capability with which CNNs function is recognizing spatial correlations.

The recurrent neural networks (RNNs) and long short-term memory (LSTM) networks are particularly good at addressing sequential datasets. In the field of information security, RNNs are used as models in detecting time-series attack behavior, including multi-step attack behavior and low-and-slow attacks [11, p. 385]. More recently, transformer models have been introduced as effective models for cybersecurity tasks. Since transformers employ attention mechanisms, transformers have the capability to learn long-range relationships in extensive networks and systems. This makes transformers scalable models compared to RNN models and makes them attractive candidates for real-time threat intelligence tools in complex critical infrastructure setups.

3.2.3. Reinforcement learning

Reinforcement learning (RL) has recently proved to be an attractive method for designing adaptive and automated cybersecurity defense systems. Unlike other learning methodologies, such as supervised learning and unsupervised learning, RL allows an agent to learn the optimal defensive strategies through an interacting process with the environment [12, p. 546]. The agent can change its actions based on rewards and penalties received.

Adaptive defense mechanisms

In the context of cybersecurity, reinforcement learning can be used for developing adaptive defense systems that can react to changing attack behaviors. An agent learning through reinforcement learning can perceive system states such as network traffic characteristics, system performance parameters, and anomaly detections and choose appropriate defense actions such as filtering network traffic, denying access, and restructuring systems [12, p. 546]. The agent eventually learns system defense strategies that maximize threat protection while minimizing disruptions to normal system operations. This is especially true in a critical infrastructure domain where traditional security policies can prove inefficient against complex threats. These systems are able to adapt and learn attack scenarios and improve their defenses against persistent and joint cyberattacks using the Reinforcement Learning concept.

Automated response systems

Another area in which reinforcement learning is beneficial is the implementation of automated incident response systems. These systems can minimize the need for human response by selecting and executing responses automatically and in real-time [13, p. 952]. These responses can include actions such as isolating compromised system resources, adjusting control parameters, and executing system fail-safes. However, in regard to the potential advantages of RL-based self-response systems, there are issues associated with the assurance of safety, convergence, and explainability for autonomous agents trained in improper ways and taking potentially disrupting or suboptimal actions. Yet, reinforcement learning stands out in AI technologies for proactive, adaptive, and scalable cybersecurity solutions for U.S. critical infrastructure.

4. System architecture

The proposed framework for the use of AI in cybersecurity solutions comprises three main levels, including the data acquisition level, the level involving the use of AI analysis, and the response and mitigation level. All levels will make the process efficient end-to-end.

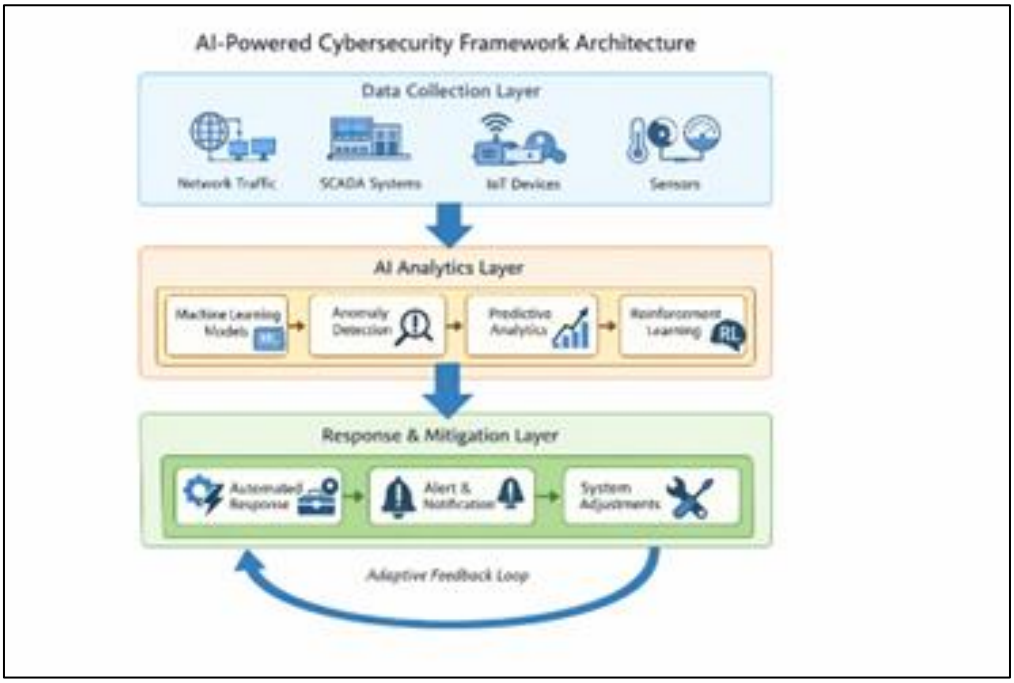

Figure 1 Overall, AI-powered cybersecurity framework architecture for protecting U.S. critical infrastructure

4.1. Data collection layer

The function of the data collection layer is related to the aggregation of data from various sources in a critical infrastructure sector. The sources may include network traffic monitors, system and application logs, industrial control systems, supervisory control and data acquisition systems, Internet of Things, and physical sensors. The data collection layer supports both IT and operational technology. Functions related to the preprocessing of data, including normalization, filtering, and extraction of features, can be handled at this level of the architecture. Infrastructure data is of high velocity and volume; hence, this layer of the architecture is built to function near real time while preserving the integrity of the information as well as its security [14, p. 1581].

4.1.1. AI analytics layer

AI Analytics Layer: The heart of the proposed architecture is the AI Analytics Layer. The AI Analytics Layer makes use of supervised learning models for classifying known threats, unsupervised learning models for identifying anomalies, as well as deep learning models for the recognition of complex patterns. Reinforcement learning agents are used for adaptive defense models. It constantly monitors the data and checks for suspicious events, the possibility of attack pathways, and system risk levels. It correlates data from various sources to improve the accuracy levels and minimize false positives [14, p. 1581]. This architecture supports the updating and redevelopment of the models without halting the system functionality.

4.1.2. Response and mitigation layer

The responsibility for carrying out actions for defending against a threat rest with the response and mitigation layer. It facilitates both automated as well as semi-automated responses, which include isolating the network, enforcing access, throttling the traffic, system reconfiguration, as well as alerting the human operator [14, p. 1581]. The decision module powered by the concept of reinforcement learning allows for adaptable responses to be selected with optimized mitigation strategies, as well as minimal operational disruptions. Within the critical infrastructure domain where availability is considered of utmost importance, the response component integrates policy and safety related to disruptive actions.

Threat detection and prediction

Not only must there be a detection capability for real-time cyberattacks, but there must also be a prevention capability that can look forward and anticipate potential future attacks. The proposed AI-based cybersecurity solution combines real-time detection and prediction capabilities.

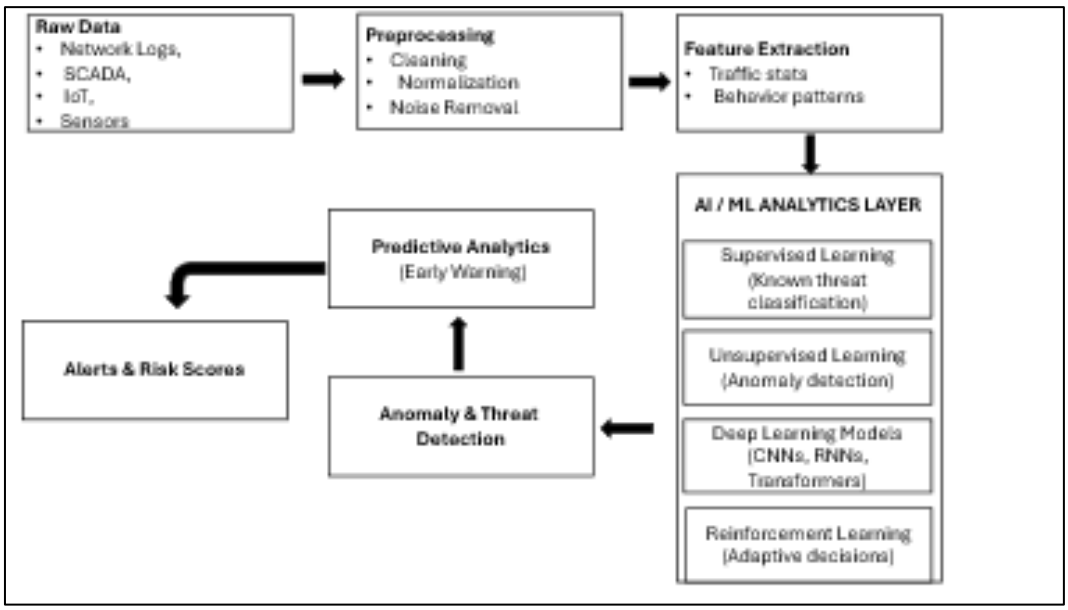

Figure 2 AI analytics pipeline for threat detection and prediction in critical infrastructure systems

4.1.3. Real-time monitoring

Real-time monitoring is done through the analysis of data streamed from various sources within the network, such as devices, systems, sensors, and logs. Real-time processing allows for quick detection of suspicious and malicious traffic [14, p. 1581]. Supervised learning is used to classify threats based on known patterns, while unsupervised learning is used to detect abnormal traffic based on system behavior that is within the established data baselines. Moreover, the use of deep learning models improves the process of real-time monitoring by identifying non-linear relations in the data. This ensures that the model can detect the signatures of attacks that cannot be detected by other security systems. In the context of correlating different data sources, the model ensures there is proper event correlation, which improves the situational awareness process.

4.1.4. Predictive analytics for early warning

Apart from real-time detection, the proposed framework includes the use of predictive analytics techniques as a warning system against probable cyber threats before the attack actually happens. Machine learning and deep learning models are used in this process to analyze historical and contextual information and make predictions based on attack trends and system vulnerabilities [15, p. 479]. The use of predictive analytics allows for proactive approaches in defending against threats, including preemptive hardening of systems, dynamic resource management, and the implementation of policy changes. Through the issuance of predictive signals, it is now possible for infrastructure system managers to implement preventive measures, thus avoiding the potential negative effects of infrastructure disruption and improving system robustness. Moving away from reactive approaches in cyber security is an essential leap in the protection of U.S. critical infrastructure.

4.2. Automated incident response

Cyber events that occur in critical infrastructure must be responded to effectively and promptly by minimizing disruptions caused by such events. The proposed AI-assisted cybersecurity paradigm includes automated response systems that use AI-driven decision-making capabilities for dealing with events and minimizing human intervention.

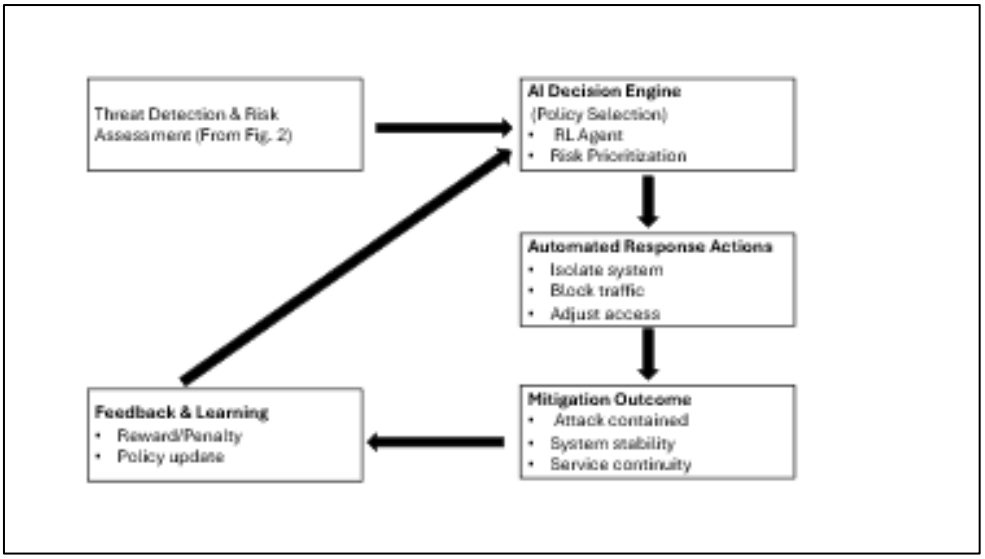

Figure 3 AI-driven automated incident response framework with reinforcement learning–based feedback loop

4.2.1. AI-driven decision making

The response and mitigation layer involves the implementation of machine and reinforcement learning algorithms in order to support decision making through AI. After an incident is detected and forecasted, the AI analytics layer examines the magnitude and extent of the threat. Depending on this analysis, the system applies corresponding measures for response based on policy and defense strategies. Reinforcement-learning agents are able to dynamically change their strategies for responding to a system based on learning from previous events and system responses to such events [16]. These agents seek to optimize their responses to achieve security as well as system availability and respond well to both known and unknown attack scenarios.

4.2.2. Reducing response time

It saves time since there is no waiting involved due to human analysis and intervention. It becomes possible to carry out actions intended for near real-time mitigation functions involving malicious component isolation, malicious traffic blocking, changing access rights, and recovery actions. In this respect, the framework closes the time gap between threat recognition and reaction to the threat and thereby helps prevent the spread of an attack and is very crucial for time-sensitive critical infrastructure sectors, where the eventual consequences of delay are likely to cause a cascade effect due to which critical infrastructure may suffer significantly [16]. In this respect, providing critical infrastructure with automation through AI increases levels of efficiency and protection against cyber-attacks.

4.3. Scalability and resilience

The critical infrastructure is large in nature and needs a continuous service, hence the importance of scalability and resilience in the design of the cybersecurity system. The proposed cybersecurity framework is scalable and resilient as it is aided by AI technology.

4.3.1. Handling large-scale infrastructure systems

The framework has a modular and distributed design. It is suited for managing complexity and scale associated with national critical infrastructures. The data harvesting and artificial intelligence analysis modules of the framework can be hosted within different segments of the infrastructures. It allows for parallel data processing. The method can process large amounts of data originating from network traffic, sensors, and control systems. The scalability of this system is further promoted by the implementation of cloud and edge computing concepts, which help maximize computational workloads dynamically according to the requirements [17]. The edge-based analysis feature facilitates quick reactions, whereas global correlation and coordination can be carried out by centralized elements. The result is a system that scaling efficiently as the complexity of infrastructure grows.

4.3.2. Fault Tolerance

Failures and attacks that affect the system are an essential resistance criterion that the proposed framework addresses by including fault tolerance to ensure the system can still function optimally. Redundancy data pathways and analytic

modules, response components to the failover, and analytic components are responsible for ensuring the system remains functional. Moreover, the models will work under suboptimal conditions and still have partial functionality when data sources become unavailable and/or when they become compromised. Automatic recovery and monitoring of the system also further improve resilience by allowing quick recovery of security services when they fail [17]. The resilience techniques ensure that the proposed solution framework is able to offer reliable cybersecurity support for U.S. critical infrastructures despite attacks and disruptions.

5. Case Study: Application To U.S. Critical Infrastructure

5.1. Energy sector: smart grid protection

The energy industry, and specifically the electrical power utility network, is one of the most important segments of the national infrastructure of the United States. Smart grids combine conventional power delivery networks and sophisticated cyber monitoring and control systems. Despite their benefits, they face unprecedented threats from cyberattacks, which can be categorized as unauthorized entry, malware attacks, data tampering, and denial of service [18]. The new AI-driven cybersecurity solution is able to increase the security of smart grids using multi-layered monitoring capabilities. Data gathering involves SCADA, smart meters, substation controllers, as well as network activity logs. Various sources of data are preprocessed and then used for the AI-driven analytics solution, where supervised models help identify existing patterns of attack, unsupervised models identify anomalies, and deep models identify hidden patterns of an attack.

Agents for the response and mitigation layer employ reinforcement learning. For instance, in the scenario where an intrusion is detected in the substation controller, the system can automatically remove the compromised device from the network and notify the control center about the incident without stopping the reinforcement learning process. Predictive analysis is used to warn systems about the possibility of an intrusion, and modifications can be made beforehand. With the incorporation of such an AI-based framework, the smart grid is able to benefit from improved resilience, lower latency in the detection-to-response cycle, as well as accuracy within the threat detection phase [18]. It is thus feasible to implement an AI-based cybersecurity framework to secure such mission-critical energy assets.

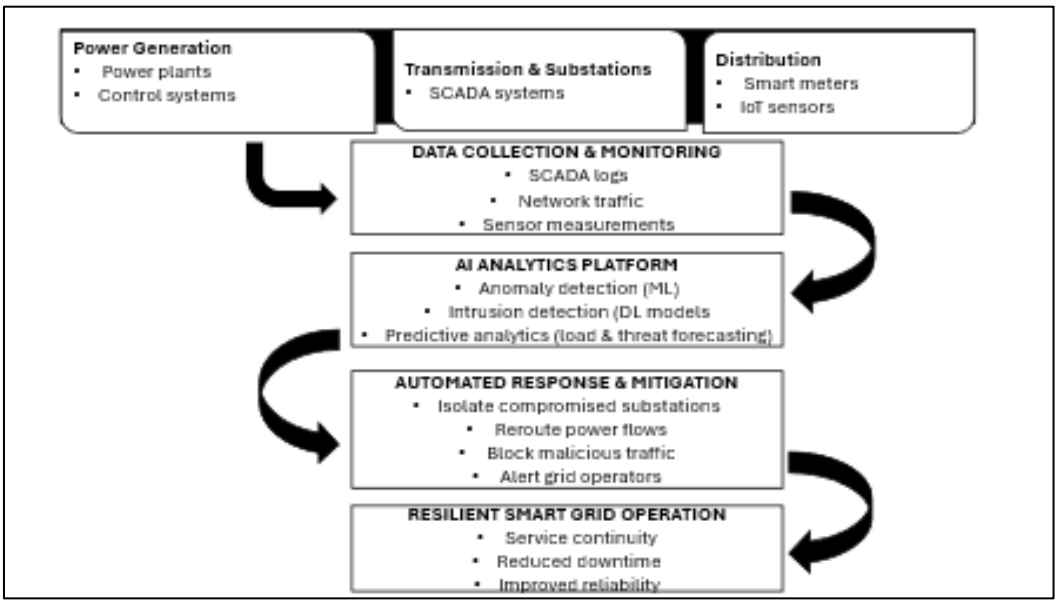

Figure 4 Application of the AI-powered cybersecurity framework to smart grid infrastructure in the U.S. energy sector

5.2. Healthcare sector: hospital cybersecurity

Hospitals and healthcare networks are now heavily dependent on a chain of digital devices, electronic health records, and medical networks. Such dependencies make hospitals vulnerable to cyberattacks like ransomware attacks, unauthorized patient record access, and disruption of patient care functions. With patient care risks at stake, hospitals must consider early detection and response [19]. The proposed AI-based cybersecurity system advances the state of the art in healthcare organization defenses against cyber-attacks by monitoring incoming and outgoing network, access, medical device telemetry, as well as user activity. While supervised models based on AI are used for classifying known

attacks, others based on unsupervised models are used for recognizing unusual patterns associated with access attempts and medical device usage.

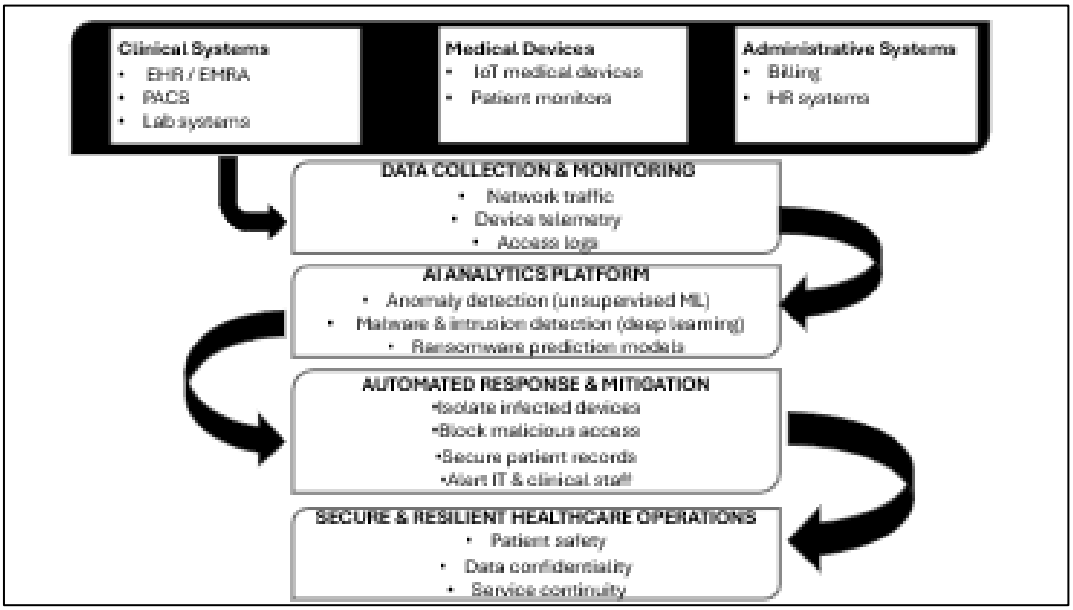

Figure 5 Application of the AI-powered cybersecurity framework to healthcare infrastructure and hospital systems

Using reinforcement learning techniques in automated response helps achieve adaptability and awareness of context while responding to attacks. Automated response in the system can limit the access of compromised machines to networks and notify administrators of attacks after identifying any malicious activity on servers or medical machines in a hospital setup. Analytics can anticipate attacks by ransomware or insider threats [20]. Thus, through the application of this framework, the hospital is able to reap the benefits of faster responses as well as improved accuracy of threat detection while being proactively protected against risks. This particular case study illustrates the flexibility of the framework even in an environment as technology-driven as the healthcare setting.

6. Evaluation and performance analysis

The efficiency of the proposed AI-based cybersecurity framework can be ascertained by using common performance metrics in the cybersecurity domain, viz. accuracy, precision, recall, and false-positive rates. Accuracy tests the total proportion of correct identifications of both malicious as well as harmless events. Precision defines the proportion of detected threats that are valid threats, which gives the efficiency of the system in terms of correctly identifying valid threats [21]. Recall or sensitivity tests the total efficiency of the framework in identifying all the valid threats. False-positive rates define the total proportion of harmless events that are wrongly classified as threats, which affects efficiency in terms of operation cost.

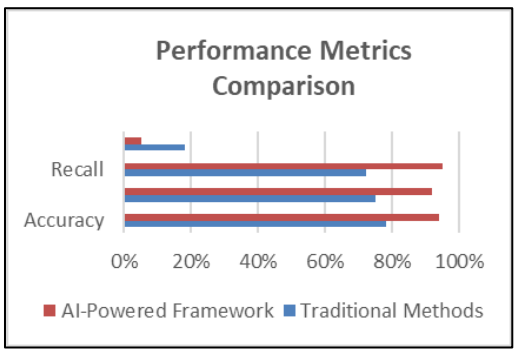

Figure 6 Performance comparison of the proposed AI-powered cybersecurity framework and traditional security methods

Compared to traditional cybersecurity strategies, including signature-based intrusion detection and rule-based security policies, the proposed system demonstrates several enhancements. The traditional systems lack the ability to detect unknown or emergent attacks, come up with higher levels of false positives, and require human response, leading to delayed reactions [21]. The AI-based system boasts better detection performance and lower levels of false positives and provides near-real-time automated response, especially in environments as complex as the ones in the critical infrastructure domain.

The results from the evaluation clearly show that incorporating supervised and unsupervised learning approaches, deep learning models, and reinforcement learning provides a substantial increase in both threat detection capabilities and response capabilities. While supervised and unsupervised learning is responsible for precise identification and recognition of both known and unknown threats, deep learning incorporates minute and high-dimensional patterns usually unidentified by conventional approaches. On the other hand, reinforcement learning increases efficiency and effectiveness in adaptive and context-driven automated response actions and helps reduce possible damage caused by both identification and response activities [21]. The suggested framework yields enormous advancements over conventional approaches in terms of efficiency, accuracy, adaptability, and robustness, which confirms its suitability for application at a nationwide level for critical infrastructure within the US.

7. Challenges and ethical considerations

Although the use of AI-based cybersecurity frameworks enhances the security of critical infrastructure in the United States, there are various challenges that must be considered for the safe use of these technologies. Data privacy remains the first issue that must be considered. This is because critical infrastructure generates enormous quantities of data, which may include operational data, user interactions, and even identifiable information of individuals [22]. Ensuring the use of AI that does not violate the privacy of this information is paramount.

Another challenge that exists for the use of AI technology is the issue related to adversarial attacks. These are usually perpetrated by malicious users who try to misuse data input to the system or make use of vulnerabilities that may lead to inappropriate responses by models used by the system. Reinforcement and deep learning models tend to have vulnerabilities that make them susceptible to these types of attacks [22]. Data bias and transparency in AI decision-making processes are other factors. AI models trained with biased, incomplete, and representative data can lead to biased outcome generation. The application of explainable AI methods can help address this challenge. Regular auditing of models can ensure improved accountability.

Finally, government standards must be complied with regarding AI adoption in critical infrastructure sectors in the United States. Frameworks must conform to cybersecurity standards and guidelines prepared and issued by relevant United States' government bodies like NIST Cybersecurity Frameworks, Department of Homeland Security guidelines, among others. This is important not only from a legal perspective but also helps to improve system integrity and interoperability for AI systems [22]. It is important for these challenges to be addressed for AI-assisted cybersecurity solutions to be used effectively and responsibly. Through thoughtful analysis of data privacy, adversarial robustness, bias, transparency, and regulatory requirements, this framework has the potential to ensure that benefits are reaped while risks associated with these challenges have been mitigated.

8. Future research directions

Though the potential applications of AI-based cybersecurity frameworks appear to be tremendous, there are still several aspects that have been left untouched and require further research and development to improve both efficacy as well as trustability. One such area that requires attention is the use of explainable AI. Although the current state-of-the-art models based on AI have been found to be highly accurate and efficient for threat detection as well as automated response actions, the underlying decision-making processes remain opaque [23].

Another significant area of research would be the integration of security primitives resistant to a potential quantum threat with AI-based cybersecurity solutions. This comes against the backdrop of the growing threat of quantum computing, which could render current cryptographic solutions vulnerable to attacks in the future [23]. Cross-sector infrastructure protection is an area that has recently gained importance. There has been growing dependence among critical infrastructure sectors, which means that cyber risks in one infrastructure area can create a domino effect in other infrastructure areas. The development of AI frameworks that can address risks in more than one interconnected infrastructure area will improve the protection of the country against systemic risks.

Lastly, policy and governance of AI are critical elements required to ensure responsible and safe uptake of AI technology. Research should thus focus on creating standardized policies, ethical standards, and compliance structures that ensure a harmonious relationship between technological development and other factors. This will ensure that cybersecurity tools based on AI are used in an ethical and nationally secure way [23]. Taken together, the future research themes make it clear that the need for innovation and careful management exists in order to ensure that the potential of AI-powered cybersecurity systems gets harnessed effectively while limiting the risks that come along with their use.

9. Conclusion

AI-based innovations in cybersecurity are presented as a vital tool in securing critical infrastructure in the U.S. Having surveyed the current state of cybersecurity threats, conventional security strategies are noted to be ineffective in identifying and defending against advanced cyber-attacks [24]. This background discussion has equally shed light on the capabilities of machine learning, deep learning, as well as reinforcement of learning processes in cybersecurity. The proposed framework involving the use of AI to ensure cybersecurity encompasses the various technologies mentioned above. The use of case studies in the energy and healthcare industries demonstrated the viability of the framework and the different ways that adaptability, automation, and predictability can improve resilience and response times. The results obtained during the evaluation process further highlighted improvements in accuracy, precision, recall, and false positives over conventional security systems.

The importance of AI in securing US critical infrastructure against the emergence of advanced cyber threats becomes even clearer as a result of the research undertaken. AI-based solutions provide the opportunity to improve security by enabling pro-active detection and subsequent response adapted according to the threat scenario, and as a result, the conclusion of the paper is that while the potential of AI in cybersecurity is high, considerations of privacy and ethics increase its reliability [24].

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] O. S. Ndibe, "AI-Driven forensic systems for Real-Time anomaly detection and threat mitigation in cybersecurity infrastructures," International Journal of Research Publication and Reviews, vol. 6, no. 5, pp. 389–411, May 2025, doi: 10.55248/gengpi.6.0525.1991.

[2] H. M. Rodriguez-Casavilca, D. Mauricio, and J. M. M. Villanueva, "Evolution of Artificial Intelligence-Based OT cybersecurity models in energy infrastructures: services, technical means, facilities and algorithms," Energies, vol. 18, no. 19, p. 5163, Sep. 2025, doi: 10.3390/en18195163.

[3] O. Obioha-Val, T. M. Kolade, M. O. Gbadebo, O. Selesi-Aina, O. O. Olateju, and O. O. Olaniyi, "Strengthening cybersecurity measures for the defense of critical infrastructure in the United States," Asian Journal of Research in Computer Science, vol. 17, no. 11, pp. 25–45, Nov. 2024, doi: 10.9734/ajrcos/2024/v17i11517.

[4] R. S. Perera and D. Dharmasooriya, "Analyzing new and emerging cyber threats in industrial control systems and their impact on critical infrastructure," DIVA, 2025. https://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Altu%3Adiva-115031

[5] N. Khalaf, Al Barazanchi, A. Radhi, S. Parihar, P. Shah, and R. Sekhar, "Mesopotamian Journal of Cyber Security," Development of Real-time Threat Detection Systems With AI-driven Cybersecurity in Critical Infrastructure, vol. 5, no. 2, pp. 500–513, Dec. 2022, doi: 10.58496/mjcs.

[6] A. Pinto, L.-C. Herrera, Y. Donoso, and J. A. Gutierrez, "Survey on Intrusion Detection Systems based on Machine learning Techniques for the protection of Critical Infrastructure," Sensors, vol. 23, no. 5, p. 2415, Feb. 2023, doi: 10.3390/s23052415.

[7] A. S. Adepoju, "Adaptive program management strategies for AI-Based cyber defense deployments in critical infrastructure and enterprise digital transformation initiatives," International Journal of Research Publication and Reviews, vol. 6, no. 7, pp. 5599–5615, Jul. 2025, doi: 10.55248/gengpi.6.0725.2651.

[8] E. E. Abdallah, W. Eleisah, and A. F. Otoom, "Intrusion Detection Systems using Supervised Machine Learning Techniques: A survey," Procedia Computer Science, vol. 201, pp. 205–212, Jan. 2022, doi: 10.1016/j.procs.2022.03.029.

[9] C. Sánchez-Zas, X. Larriva-Novo, V. A. Villagrá, M. S. Rodrigo, and J. I. Moreno, "Design and evaluation of unsupervised machine learning models for anomaly detection in streaming cybersecurity logs," Mathematics, vol. 10, no. 21, p. 4043, Oct. 2022, doi: 10.3390/math10214043.

[10] D. Akgun, S. Hizal, and U. Cavusoglu, "A new DDoS attacks intrusion detection model based on deep learning for cybersecurity," Computers & Security, vol. 118, p. 102748, May 2022, doi: 10.1016/j.cose.2022.102748.

[11] J. T. Santoso, B. Hartono, F. D. Silalahi, and M. Muthohir, "Transformers in Cybersecurity: Advancing Threat Detection and Response through Machine Learning Architectures," Journal of Technology Informatics and Engineering, vol. 3, no. 3, pp. 382–396, Dec. 2024, doi: 10.51903/jtie. v3i3.211.

[12] E. Cadet, E. D. Etim, I. A. Essien, J. O. Ajayi, and E. D. Erigha, "The role of reinforcement learning in adaptive cyber defense mechanisms," International Journal of Multidisciplinary Research and Growth Evaluation, vol. 2, no. 2, pp. 544–559, Jan. 2021, doi: 10.54660/. ijmrge.2021.2.2.544-559.

[13] S. Ren, J. Jin, G. Niu, and Y. Liu, "ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization," Applied Sciences, vol. 15, no. 2, p. 951, Jan. 2025, doi: 10.3390/app15020951.

[14] M. Ragab and A. Altalbe, "A Blockchain-Based architecture for enabling cybersecurity in the Internet-of-Critical infrastructures," Computers, Materials & Continua/Computers, Materials & Continua (Print), vol. 72, no. 1, pp. 1579–1592, Jan. 2022, doi: 10.32604/cmc.2022.025828.

[15] [15] N. K. D. O. Ofoegbu, N. O. S. Osundare, N. C. S. Ike, N. O. G. Fakeyede, and N. A. B. Ige, "Real-Time Cybersecurity threat detection using machine learning and big data analytics: A comprehensive approach," Computer Science & IT Research Journal, vol. 4, no. 3, pp. 478–501, Dec. 2023, doi: 10.51594/csitrj. v4i3.1500.

[16] [16] J. Yu, A. V. Shvetsov, and S. H. Alsamhi, "Leveraging Machine learning for Cybersecurity Resilience in Industry 4.0: Challenges and future directions," IEEE Access, vol. 12, pp. 159579–159596, Jan. 2024, doi: 10.1109/access.2024.3482987.

[17] "Edge Computing, 5G, and cloud Security convergence: Strengthening USA's critical infrastructure resilience," International Journal of Computer Applications Technology and Research, Mar. 2025, doi: 10.7753/ijcatr1209.1003.

[18] A. Idowu, I. Ismaila, and J. A. Ojeniyi, "Machine Learning-Driven Cybersecurity Solutions for Enhanced smart grids and Critical Infrastructure: A review," NIPES Journal of Science and Technology Research, vol. 7, no. 3, pp. 159–184, Jul. 2025, doi: 10.37933/nipes/7.3.2025.11.

[19] Md. T. Hasan and I. Ahmed, "AI-driven anomaly detection for data loss prevention and security assurance in electronic health records," ai-driven anomaly detection for data loss prevention and security assurance in electronic health records, vol. 04, no. 03, pp. 35–67, Sep. 2025, doi: 10.63125/dzyr0648.

[20] M. Sewak, S. K. Sahay, and H. Rathore, "Deep reinforcement learning in the advanced Cybersecurity Threat Detection and Protection," Information Systems Frontiers, Aug. 2022, doi: 10.1007/s10796-022-10333-x.

[21] N. Mohamed, "Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms," Knowledge and Information Systems, vol. 67, no. 8, pp. 6969–7055, Apr. 2025, doi: 10.1007/s10115-025-02429-y.

[22] C. Mohitkar and D. Lakshmi, "Explainable AI for transparent Cyber-Risk assessment and Decision-Making," in Advances in computational intelligence and robotics book series, 2024, pp. 219–246. doi: 10.4018/979-8-3693-7540-2.ch010.

[23] N. S. A. Daniel and N. S. S. Victor, "Emerging trends in cybersecurity for critical infrastructure protection: a comprehensive review," Computer Science & IT Research Journal, vol. 5, no. 3, pp. 576–593, Mar. 2024, doi: 10.51594/csitrj. v5i3.872.

[24] A. S. Adebayo, N. Chukwurah, and O. O. Ajayi, "Artificial intelligence and machine learning algorithms for advanced threat detection and cybersecurity risk mitigation strategies," Engineering and Technology Journal, vol. 10, no. 03, Apr. 2025, doi: 10.47191/etj/v10i03.18.