International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 1, Jan-Feb-2026, ISSN (Online): 2395-566X

AI-Driven Zero Trust Security Architecture for Protecting U.S. Critical Infrastructure

Nagaraju Goshikonda

Abstract- The digitalization of critical infrastructure sectors of the U.S. economy such as energy, transportation, healthcare, and defense has expanded the cyber-attack surface at a rapid pace. The old models of perimeter-based security are no longer effective against complex attacks, like advanced persistent attacks (APTs), insider attacks and AI-assisted cyber-attacks. This paper will suggest AI-based Zero Trust Security Architecture (ZTSA) adapted to secure the critical infrastructure in the United States. The framework incorporates behavioral analytics, federated learning, and adaptive risk scoring, that allow one to continue verification and intelligent response to threats. The predictive and generative AI models are utilized to simulate the attack scenario, improve anomaly detection, and automate policy enforcement. Experimental assessment based on simulated critical infrastructure datasets is shown to have a higher detection rate of 95.8 and a 30% lower rate of false positives than traditional zero-trust systems. The outcomes show that AI-enhanced zero-trust models will be capable of mitigating critical infrastructure in the US to a considerably greater extent in terms of resilience, scalability, and mitigation of threats in real-time.

Keywords – Zero Trust Architecture, Critical Infrastructure Security, AI-Driven Cybersecurity, Federated Learning, Behavioral Analytics, Adaptive Access Control.

I. INTRODUCTION

The rapid digitalization of the critical infrastructure of the U.S. like its energy grids, transportation systems, healthcare systems, water systems and defense systems has significantly enhanced the cyber threat environment. Industrial control systems (ICS), interrelated cloud technologies, and Internet of Things (IoT) are the core of modern infrastructure settings and present advanced hybrid IT/OT settings, which are difficult to secure. Security models based on perimeter defense strategies become inefficient against sophisticated persistent threat (APT), ransomware attacks, and their interference with national infrastructure, less and less. Zero Trust Architecture (ZTA) has emerged as a potent template eliminating implied trust and unrelenting validation to users, devices, and applications [1]. ZTA reduces the horizontal flows and privilege escalation risks through the provision of identity-based controls as opposed to network-based controls [2]. However, hardened zero-trust implementations are often unable to implement adaptive intelligence to address AI-based attackers and dynamic patterns of attackers in worthwhile infrastructure environments [3].

Artificial intelligence has emerged as an immensely empowering facilitator of resilience in cybersecurity against critical sectors. Threat detection systems based on AI use machine learning algorithms to process large amounts of telemetry information to detect subtle anomalies that would be missed with signature-based systems [4]. In critical infrastructure settings, where business continuity is of utmost importance, AI is useful to improve real-time decision making and predictive risk management. Deep learning models used to derive behavioral analytics can profile both user and device usage and dynamically score trust and impose access controls based on risk [5]. Moreover, adaptive zero-trust models based on AI offer automatic policy adaptations to changing threats, lowering the need to intervene and reduce the response time [6]. In spite of these recent developments, integration problems still exist, especially with legacy OT systems where the constraints on computational power and the diversity of protocols limit the traditional AI implementation.

The concept of federated learning has become an accelerating solution to the privacy, sovereignty, and regulation issues in cross-sector infrastructure collaboration. Nationally regulated security requirements and industry compliance requirements often limit critical infrastructure operators to distribute raw cybersecurity information. Federated learning can be used to train distributed model without sharing sensitive data to maintain secrecy and enhance shared threat warning [7]. Federated systems promote collaborative defense of nodes geographically distributed by aggregating encrypted model parameters instead of raw data [8]. This would be especially applicable to sectors like energy and transportation where coordinated detection of attack can prevent a cascading failure in a region. The inclusion of federated learning into zero-trust models increases the scalability and preservation of privacy and does not decrease the detection performance [9].

Generative AI also brings about protection opportunities and new threats to zero-trust ecosystems. Generative adversarial networks (GANs) can be deployed on the defensive side to provide more efficient simulations of advanced attacks, to enable security systems to predict both zero-day exploits and lateral movement strategies that adversaries might use, in the future, before it is exploited [10]. Security orchestration, compliance mapping and policy refinement can also be automated with large language models (LLMs) in complex infrastructure environments [11]. Nevertheless, opponents can use generative AI to develop advanced phishing schemes, deepfakes social engineering, or automated reconnaissance instruments that should not be countered by conventional defenses [12]. This two-sidedity of generative AI requires a strictly controlled use in a zero-trust architecture to make it resilient without making it more vulnerable.

II. LITERATURE REVIEW

Zero Trust for Critical Infrastructure

The concept of Zero Trust Architecture (ZTA) has become a cornerstone of cybersecurity paradigm in securing national critical infrastructure as a response to more advanced cyber-attacks. Opposite to the traditional perimeter-based models, the concept of zero trust presupposes that no user, device, or network segment can be trusted and promotes constant verification and access based on the least privilege. Ahammed noted that identity-based security and micro-segmentation are particularly useful in curbing horizontal movements in high value infrastructure setups [16]. The method is particularly noteworthy in the case of operational technology (OT) networks, where legacy systems frequently do not have in-built security controls. Bhaskaran also noted that industries like energy grids, military systems and transportation networks should adopt zero-trust as policies since attackers often circumvent the perimeter security in place by compromising credentials and through supply-chain attacks [17]. Also, Zanasi et al. showed that resilience can be improved since flexible zero-trust models can be used to achieve isolation of critical assets, as well as by enforcing policy granularity in industrial IoT ecosystems [18].

In spite of these benefits, there are numerous practical applications which are more or less rule-based and static. Mushtaq et al. found that many organizations implement zero trust as a policy overlay instead of an adaptive architecture that is fully dynamic, constraining its ability to respond to dynamic threats [19]. Conventional ZTA implementations tend to use canned access rules and signature-based detection, which are not effective in identifying new attack patterns and insider threats. Further, OT systems prevalent in the U.S. critical infrastructure in the past have integration problems because of the heterogeneity in protocols and the maintenance concerns. These restrictions suggest that although zero trust offers a solid conceptual basis, its full capabilities in critical infrastructure security need intelligent automation and adaptive analytics profiles that can be updated to keep up with the threat environment [20].

Role of AI in Zero Trust

Artificial intelligence has emerged as a central facilitator to boosting the effectiveness of zero-trust by injecting adaptive and data-driven security functionality. The AI-based systems are capable of analyzing user behavior, device posture, and network telemetry using dynamic risk scores to inform real-time access control. Ajish has shown that machine learning-based anomaly detection is much more effective than the traditional authentication systems in detecting the slightest behavioral differences that are evidence of compromise [21]. The latter ability is specifically useful in the context of critical infrastructure that attackers tend to be low-and-slow to be detected. Chokkanathan et al. also demonstrated that the implementation of AI into zero-trust processes enhances cyber-resilience through automated threat correlation and quick incident response [22]. Deep learning-based models have extra benefits in high-volume telemetry setup. According to Akinloye et al., neural network models have the ability to pick up multistage, intricate attack patterns that are often overlooked by traditional intrusion detection systems (IDS) in distributed national infrastructure networks [23]. On the same note, Joshi highlighted that AI-powered risk engines can be used to provide context-sensitive access control with trust decisions dynamically adjusting to real-time threat intelligence as well as behavioral baselines [24]. Nevertheless, certain operational issues are also pointed out by researchers, such as the explainability of models, their computational costs, and how AI systems could be manipulated adversarial. These issues underscore the need to develop open, resilient AI components in the implementation of intelligence on zero-trust architectures.

Federated and Privacy-Preserving Security

Sharing of information among critical infrastructure industries is necessary in the identification of coordinated attacks, but regulatory and national security issues tend to hinder the sharing of raw security information among organizations. Federated learning (FL) has become a promising privacy-saving mechanism that allows sharing the model training without revealing sensitive data. Faheem et al. showed that federated AI models could allow near-centralized detection accuracy, and data sovereignty among distributed infrastructure operators [25]. FL lowers the chances of the leakage of data and allows the development of collective threat intelligence by sending encrypted updates of the models. Ibitoye suggested a zero-trust cloud architecture designed and managed by AI, where federated learning can be considered to foster secure cooperation between critical sectors, and it can highlight its contribution to upholding adherence to data protection rules [26]. On the same note, Nangi et al. pointed out that multi-layered federated zero trust systems have the potential of

improving the scalability and resiliency of cloud-native enterprise environments by decentralizing the assessment of trust among nodes [27]. Although such advantages exist, federated deployments have also brought about new problems such as communication overhead, model poisoning, and synchronization latency. Mushtaq et al. have observed that federated zero-trust ecosystems require careful aggregation measures and mechanisms of differentiated privacy to ensure both quality and security of performance [19].

Generative AI and Emerging Risks

Generative AI is already transforming cybersecurity swiftly as it is both an advanced defense mechanism and a high-end offense method. Generative adversarial networks (GANs) can be used to generate synthetic data of attacks that enhance model resilience to uncommon and zero-day attacks on the defensive side. Yigit et al. have shown that the security agents based on the LLM can improve the security of critical infrastructure by automating the process of threat intelligence analysis and incident triage [28]. These capabilities facilitate offensive countermeasures of zero-trust space scenarios by enabling systems to preempt the emergent vectors of attacks instead of having to rely on historical signatures.

But there are also serious security issues with the emergence of generative AI. Xu et al. cautioned that the generative models might be used by the opponent to generate very persuasive phishing attacks, automated reconnaissance systems and polymorphic malware that can overcome conventional defense mechanisms [29]. Reed also reported that there was increased concern that advanced threat actors have perhaps already started using AI to penetrate critical infrastructure networks and extend their dwell time and operational risk [30]. Such trends highlight the dual-purpose character of generative AI and the necessity of powerful governance, surveillance, and validation strategies in the implementation of GenAI as a part of zero-trust structures.

Research Gap

Despite the fact that the previous studies prove that a significant advance in the field of zero-trust security, AI-driven analytics, and federated collaboration is achieved, a number of critical gaps still exist. First, the majority of the current research considers zero trust and AI as complementary technologies but loosely coupled, and not closely integrated entities of a single security architecture. Second, the problem of specific operational constraints of U.S. critical infrastructure, such as legacy OT integration, high-availability needs, and cross-sector interdependencies are given little attention. Third, existing literature does not provide well-developed adaptive trust scoring mechanisms that integrate behavioral analytics and contextual intelligence with federated insights into one decision model. Moreover, compliance automation at real-time has been under-researched yet the critical infrastructure operators have significant regulatory load.

III. METHODOLOGY

Proposed Framework Architecture

The intended AI-Driven Zero Trust Security Architecture (AI-ZTSA) is an intelligence-based, multi-layered framework that is associated with safeguarding the environments of the U.S. critical infrastructure that cut across hybrid IT/OT ecosystems. In contrast to the traditional deployment of zero trust where the main aim is to verify identities, the suggested architecture will incorporate dynamic AI analytics and privacy-conserving collaboration systems to facilitate ongoing and context-sensitive trust analysis. The framework is structured into five layers closely coupled with each other and offering end-to-end visibility, automated decision-making, and real-time threat mitigation.

Layer 1: Data Ingestion Layer

This base layer gathers heterogeneous security telemetry globally over the infrastructure environment. The network traffic flows and industrial control system (ICS) logs, event logs in identity and access management, and endpoint telemetry, as well as external threat intelligence feeds, are listed as data sources. Since critical infrastructure settings tend to have legacy OT protocols, the ingestion layer has protocol normalization and time-synchronization features to maintain the consistency of data.

Layer 2: AI Analytics Layer

The cognitive layer of the framework is made up of the AI analytics layer. It uses deep learning, behavioral profiling, and graph analytics to constantly assess the risk by the users, devices, workloads, and network segments. This layer produces near real-time dynamic behavioral baselines and anomaly scores. Through sequence modeling and graph-based correlation, the system is capable of identifying stealthy lateral movement, insider threats, as well as, multi-stage attacks which are frequently missed by traditional signature systems.

Layer 3: Federated Learning Layer

The architecture will integrate a federated learning (FL) layer to help facilitate cross-sector collaboration without infringing on the needs of data sovereignty. Involved infrastructure nodes learn local models using only their own telemetry and send only encrypted updates to the model to a secure aggregator. This allows sharing of threat intelligence whilst maintaining confidentiality which is a critical requirement of the national infrastructure operators.

Layer 4: Zero Trust Enforcement Layer

Continuous verification is operationalized by this layer. It takes risk measurements generated by the AI layer and implements dynamic access control decisions based on policy engines and software-defined micro-segmentation. Based on the continuously recalculated trust, it is possible to contain compromised identities or devices in a short period of time. Enforcement layer is used in conjunction with identity

providers, endpoint detection systems and network access controls to enforce least-privilege access throughout the environment.

Layer 5: Security Operations Dashboard
The uppermost layer presents integrated situational awareness to the security analyst and infrastructure operators. It displays the scores of trusts, anomaly warning, compliance posture, and federated intelligence via real-time dashboards. Response recommendations that are generated with the help of AI and automated playbooks aid in reducing mean time to detect (MTTD) and mean time to respond (MTTR). Audit reporting is also supported in the dashboard in line with the requirements by the critical infrastructure regulations.

Mathematical Formulation
1. Behavioral Risk Score

For each entity e(user, device, or workload), the behavioral risk score is computed as:

$$R(e) = \sum_{i=1}^{n} w_i \cdot \frac{| f_i(e) - \mu_i |}{\sigma_i}$$

Where:
- f_i (e)= observed behavioral feature of entity e
- μ_i= historical mean of feature i
- σ_i= standard deviation of feature i
- w_i= importance weight of feature i

n= total number of behavioral features

This formulation quantifies deviation from baseline behavior using standardized feature distances. Higher values of R(e)indicate anomalous activity requiring stricter verification.

2. Dynamic Trust Score
The dynamic trust score for user uis defined as:

$$S(u) = \alpha C(u) + \beta B(u) + \gamma D(u) + \delta L(u)$$

Where:
- $C(u)$= credential strength score
- $B(u)$= behavioral consistency score
- $D(u)$= device posture trust score
- $L(u)$= location/contextual risk score
- $\alpha, \beta, \gamma, \delta$= weighting coefficients

Subject to:

$$\alpha + \beta + \gamma + \delta = 1$$

Access is granted only if:

$$S(u) > \tau$$

Where τrepresents the adaptive trust threshold determined by real-time threat intelligence.

Federated Aggregation
The global model update at iteration t+1is computed as:

$$w_{global}^{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} w_k^{t+1}$$

Where:
- K= total participating nodes
- n_k= number of samples at node k
- $N = \sum_{k=1}^{K} n_k$
- w_k^{t+1}= updated local model weights

This weighted aggregation preserves proportional contribution while maintaining decentralized data storage.

Generative Threat Simulation
The generative adversarial network (GAN) objective function is:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right]$$

Where:

G= generator model
D= discriminator model
x= real attack samples
z= random noise vector

This adversarial learning process enables the synthesis of realistic ICS/SCADA attack scenarios, improving robustness against zero-day threats.

C. System Implementation
Data Sources
The prototype implementation incorporates heterogeneous telemetry common to the U.S. critical infrastructure settings in order to provide a complete situational awareness both in the IT and OT realms. A network telemetry including NetFlow traces and packet metadata, ICS/SCADA operational logs, enterprise identity provider identity and authentication events, external threat intelligence feeds and OT sensor and controller data are all collected in the data ingestion pipeline.

Since critical infrastructure systems frequently face extremely distributed as well as latency sensitive conditions, the pipeline is created to permit batch as well as real-time streaming ingestion. Further processing of normalization and time-alignment are carried out to manage protocol heterogeneity and historic system format which are common in industrial networks.

AI Models
The AI analytics layer is a hybrid combination of complementary learning models to perform strong and flexible threat detection. A sequence behavioral analysis of the user sessions and device activity is carried out with Long Short-Term Memory (LSTM) networks that allow the system to capture temporal patterns related to slow-moving intrusions and insider threats. Graph Neural Networks (GNNs) are users,

hosts, processes and network flows relationships that enable the identification of lateral movement and multi-hop attack paths which are hard to detect in traditional approaches. Generative Adversarial Networks (GANs) are also used to generate rare and zero-day attack events to deal with class imbalance and enhance model generalization on ICS threat environments.

Zero Trust Controls

The implementation of the zero-trust enforcement layer realizes continuous verification through combining dynamic risk scoring with real-time policy enforcement policies. The architecture deploys the concept of constant authentication and re-verification of sessions to prevent the creation of a situation in which the users can always trust the identity of a user once they have initially successfully logged in. Software-defined micro-segmentation separates important assets and limits the east-west traffic flow, which minimizes the blast radius of possible attacks by a significant margin. Least-privilege access control provides users and devices with the minimum permissions needed to perform their roles and device posture checks continuously determines endpoint health and patch status and security configuration.

Compliance Automation

To deal with the complicated regulatory framework of the critical infrastructure within the United States, the framework uses an NLP-based compliance automation engine that translates policy requirements into machine-enforceable controls. Language models in the form of transformers are fed regulatory documents and automatically transform the obligations into formalized security rules, eliminating errors in interpretation. The system is currently operational with large regulatory frameworks such as the NIST cybersecurity guidelines, CISA operational directives and FERC/NERC CIP standards to protect the energy sector.

Continuous monitoring engines compare the activities in the system with these codified policies in real time and produce automated audit artifacts, compliance dashboards and exception alerts. The architecture, by integrating compliance checking into the zero-trust workflow, provides both a constant state of regulatory compliance and a much lower cost of audit preparation and the human cost overhead of audit preparation, which is especially important in large, highly available infrastructure environments.

IV. RESULTS AND DISCUSSION

Experimental Setup

The AI-Driven Zero Trust Security Architecture (AI-ZTSA) was tested on a large-scale simulated dataset of U.S. critical infrastructure to simulate the conditions of hybrid IT/OT infrastructures that are common in energy, transportation, and water utility industries. It was a collection of 10.2 million security event logs, such as authentication logs, ICS telemetry logs, network flow logs, endpoint activity logs, and policy enforcement logs. Out of these events, 0.9% were verified attack events which is realistic class imbalance conditions that occurs in critical infrastructure networks. The experimental setup had 18 distributed infrastructure nodes which were geographically diverse facilities interconnected via a hybrid cloud-OT architecture.

Three baseline strategies were used to be compared: (1) a fixed rule-based security engine, (2) a machine learning-based intrusion detection system (ML-IDS) as well as (3) a traditional Zero Trust Architecture (ZTA) without adaptive AI improvement. The accuracy, precision, recall, F1-score, false positive rate, detection latency, bandwidth overhead, and compliance validation time were the performance evaluation metrics. The Federated learning synchronization was implemented after every 150 local iterations to trade off convergence and communication efficiency. Stratified cross-validation was used to train and test all models so that they could be statistically robust.

Threat Detection Performance

Table 1: Detection Performance

Method	Accuracy	Precision	Recall	F1	False Positive
Rule-Based	82.4%	68.1%	60.2 %	64.0 %	11.9%
ML-IDS	89.7%	78.5%	75.2 %	76.8 %	7.4%
Standard ZTA	91.3%	83.1%	79.6 %	81.3 %	6.1%
Proposed AI-ZTSA	95.8%	90.2%	88.9 %	89.5 %	4.3%

According to the results of the experiment, the proposed AI-ZTSA beats the baseline models by a significant margin in all evaluation metrics. The framework had a 95.8% detection rate, which was 4.5% higher than standard ZTA and 13.4 increased compared to the rule engines left unmodified. Precision was also enhanced to 90.2, which minimized unnecessary investigations of incidents. The decreasing false positive rate (to 4.3) was most noticeable, which is more of a 30 percent reduction over traditional ZTA and almost 64 percent over rule-based detection systems. The model was very robust against attacks of lateral movement because its correlation engine is represented by a graph neural network. Behavioral deviation modeling also enhanced the capability to detect insiders' threats significantly. The implementation of real-time trust recalculation ensured decision latency of less than 120 milliseconds, which was adequate to meet the requirements of operational continuity of critical infrastructure systems whose downtime sensitivity is low.

Federated Learning Performance

Table 2: Federated Learning Analysis

Configuration	Nodes	Accuracy	Privacy Level	Bandwidth Reduction	Convergence Time
Centralized	1	96.0%	Low	0%	9.1 hrs
Federated	6	95.6%	High	41%	12.4 hrs
Federated	12	95.3%	High	55%	15.8 hrs
Federated	18	95.1%	High	62%	18.2 hrs

The federated learning setup over 18 infrastructure nodes has a 95.1% accuracy, which is only a 0.9% decrease as compared to centralized training and a significant improvement in privacy protection. The amount of bandwidth decreased by 62%, since the raw security logs were still kept on the local machine and only encrypted model parameters were sent. This outcome confirms that federated AI has the potential to facilitate cross-sector data sharing of intelligence without violating data sovereignty or regulatory responsibility. The minimal convergence time increment refers to coordination overhead between the distributed nodes but the trade-off can be tolerated in the context of national infrastructure collaboration. Model poisoning risks were reduced by secure aggregation mechanisms, and the confidentiality of the models was ensured by noise injection of the differentiation of privacy, without affecting the model stability.

Compliance Monitoring Results

Table 3: Compliance Monitoring Performance

Framework	Checks	Violations Detected	Detection Rate	False Alerts	Avg Response Time
NIST Controls	84,500	2,940	97.8%	3.2%	2.6 sec
CISA Directives	51,200	1,780	98.5%	2.4%	2.1 sec
FERC/NERC CIP	39,600	1,230	98.0%	2.7%	2.9 sec
Overall	175,300	5,950	98.1%	2.8%	2.5 sec

The automated compliance module recorded a total identify rate of 98.1% and this indicates high accuracy in detecting policy violations within regulatory frameworks. The enforcement of the policy in real-time allowed blocking the misconfigurations and the unauthorized control actions to be automated in the OT systems. False alerts were less than 3% and this greatly minimized audit fatigue and cost of manual investigation. The policy translation engine was an NLP system which dynamically translated changing regulatory updates into controls which could be executed by the machine. The presence of continuous monitoring guaranteed by persistent adherence to NIST cybersecurity standards and industry standards. The average response times were less than three seconds, and it was possible to validate compliance almost in real-time without facing the operational processes.

V. DISCUSSION

Adaptive Trust Evaluation

The findings indicate that adaptive trust evaluation is highly effective to boost zero-trust capabilities in critical infrastructure settings. The AI-ZTSA is able to calculate the level of trust continuously through behavioral analytics and contextual risk scoring as opposed to relying on permanent validation of credentials. This solution is consistent with the results that identity-focused, dynamic verification models are better in the challenging settings, as compared to the traditional access controls [16]. The decrease in false positives means that risk-weighted scoring does not deteriorate the decision accuracy at the expense of the recall. Moreover, adaptive scoring allowed the system to notice minor changes in the behavior of operators, which is also consistent with the findings of research highlighting AI-based anomaly detection in infrastructure systems [23]. Ongoing trust recalibration was a factor that restricted the opportunities to move laterally as well, which further substantiated earlier arguments that micro-segmentation coupled with intelligence-based monitoring is an effective way to curb attacks propagation [18]. Generally speaking, adaptive evaluation mechanisms enable the zero-trust enforcement to adapt to the threats alongside the threat behavior, enhancing the resilience to advanced adversaries trying to attack the assets of national infrastructure.

Proactive Threat Discovery

Generative models' integration helped to proactively discover threat by simulating zero-day attack patterns and ICS exploitation scenario. This goes in line with the new studies that propose that generative AI can be used to better predictive threat modeling in cybersecurity [22]. The system was able to enhance the recall rates through the addition of synthetic adversarial examples to the training datasets, without overfitting the system to attack signatures of the past. Simulation module based on GAN enhanced the detection of multi-stage intrusion and credential abuse schemes, as previously noted that generative learning enhances the robustness of classifiers [20]. Proactive simulation also minimized dwell time as it was able to detect the unusual patterns at earlier stages of attack. This predictive ability is especially valuable in critical infrastructure industries where the failure of a single industry can have domino effects on the nation. The experimental results indicate that generative AI has

the capability to transform zero-trust designs into proactive defense systems that do not rely on reactive models of enforcement, but instead attempt to model the up-coming adversarial strategies in ways that they can be stopped prominently before weakening the system is prevalent and exploited in large numbers.

Privacy-Preserving Collaboration

The results of federated learning indicate that privacy-preserving cooperation between distributed infrastructure operators is possible. The results of 95.1% detection rate on 18 nodes can justify the idea that decentralized intelligence sharing can be highly effective without centralization of sensitive information. This observation is consistent with the previous studies that show the efficacy of federated AI in the case of critical infrastructure [25]. The high bandwidth savings also confirms the arguments that distributed aggregation systems are scalable to national implementations [20]. Sovereignty Federated zero-trust models preserve data sovereignty and allow collective defense to develop a collaborative security posture across sectors. This kind of cooperation is critical towards identifying the coordinated and cross-regional attacks which individual operators might not detect alone.

Automated Regulatory Enforcement

Automated compliance monitoring scored highly in meeting the national regulatory needs with a 98.1% detection rate across the frameworks. Integrating NLP-based policy translation into zero-trust enforcement had a major impact on minimizing compliance overhead on the humans involved, and the same system ensured real-time verification. This is in line with literature that AI can be used to automate regulatory mapping and policy enforcement in complex environments [25]. Continuous compliance validation means that the operator of the infrastructure remains aligned to the changing standards without delaying the audits every so many periods. Operational governance is also enhanced by the possibility of dynamically blocking non-compliant actions, which reduces the risk exposure. Having compliance logic built into the trust evaluation process, the architecture aligns security and regulatory enforcement, resolving a long-term gap between cybersecurity operations and governance models.

Implementation Challenges and Computational Overhead

Implementation challenges are still present even with the performance improvements. Computational overheads in AI models, especially graph neural networks and GANs, have the potential to induce stress on legacy OT environments with limited processing capacity. This issue is in line with reports of resource limitation within industrial systems [19]. Secure communication channels and coordination infrastructure are also needed in model training and federated synchronization, which may add complexity to operations. Moreover, the issue of explainability is also critical; the transparency of decision-making is a significant requirement of regulatory audits and trust among operators. According to pre-existing studies, opaque AI decisions can impede their implementation in safety-critical settings [24]. These issues need to be solved by optimization of lightweight models, explainable artificial intelligence, and hardware acceleration approaches to fit industries.

Adversarial AI Risks and Future Considerations

The generative AI has emerging risks that should be averted due to its dual use nature. The opponents can take advantage of AI to create avoidance malware or/and attempt to control the signals of trust as reported in earlier studies on cybersecurity [25], [20]. Poisoning attacks over federated systems are also a threat that could occur in case aggregation protection is weak. As such, the integrity is required to be maintained with strong adversarial training, secure aggregation, and anomaly validation layers. Explainable trust scoring, quantum-resistant authentication, and lightweight edge-deployable AI modules should be improved in the future.

VI. CONCLUSION

This paper introduced a comprehensive AI-based Zero Trust Security Architecture (AI-ZTSA) that fits a critical infrastructure setting in the U.S. and responds to the increasing shortcomings of traditional zero-trust frameworks and perimeters. The proposed framework, based on behavioral analytics, federated learning, generative threat simulation, and automated compliance enforcement, exhibits significant gains in accuracy of detection, reduction of false-positive, and adaptive evaluation of trust in real-time. Experimental findings affirm that integrating AI into zero-trust decision loops can proactively detect insider threats, horizontal flows, and multistage attacks in addition to maintaining the sovereignty of data across the distributed infrastructure nodes. Regulatory preparedness is also promoted by the architecture based on ongoing policy translation in accordance with key cybersecurity standards. Future research will involve optimization of the lightweight edge AI, interpretable trust scoring models and strong adversarial defenses to enhance autonomous zero-trust systems of protecting mission-critical national infrastructure.

Acknowledgment

The authors are deeply grateful to the works of researchers, practitioners, and organizations whose previous research on the topic of zero trust architecture, artificial intelligence, and critical infrastructure security provided a strong foundation and background behind this paper. Another important aspect that the authors like to point out is the assistance of the cybersecurity and critical infrastructure research community to provide open datasets, technical advice, and standards that facilitated the creation and testing of the proposed framework. This would not have been possible without special thanks to

colleagues and peer reviewers who have ensured that this work was more lucid, rigorous, and technically filled. The views, results and conclusion presented in this paper are personal to the authors and not always representative of the institutions and other groups that they are affiliated to.

REFERENCES

1. M. F. Ahammed, "Zero-Trust architectures for securing U.S. critical infrastructure," Frontiers in Computer Science and Artificial Intelligence, vol. 4, no. 2, pp. 71–78, 2025. doi: 10.32996/fcsai.2025.4.2.7.
2. M. F. Ahammed and M. R. Labu, "AI-Driven adaptive Zero-Trust models for U.S. defense networks," Journal of Computer Science and Technology Studies, vol. 7, no. 6, pp. 485–493, 2025. doi: 10.32996/jcsts.2025.7.6.56.
3. S. Ahmadi, "Autonomous identity-based threat segmentation for zero trust architecture," Cyber Security and Applications, vol. 3, p. 100106, 2025. doi: 10.1016/j.csa.2025.100106.
4. G. A. Ajimatanrareje and J. S. Agbesi, "AI-Powered Zero Trust Architectures for Critical Infrastructure Protection: A comprehensive framework for next-generation cybersecurity," International Journal of Scientific Research and Modern Technology, pp. 40–56, 2025. doi: 10.38124/ijsrmt.v4i9.792.
5. D. Ajish, "The significance of artificial intelligence in zero trust technologies: A comprehensive review," Journal of Electrical Systems and Information Technology, vol. 11, no. 1, 2024. doi: 10.1186/s43067-024-00155-z.
6. Akinloye, S. Anwansedo, and O. T. Akinwande, "AI-driven threat detection and response systems for secure national infrastructure networks: A comprehensive review," International Journal of Latest Technology in Engineering Management & Applied Science, vol. 13, no. 7, pp. 82–92, 2024. doi: 10.51583/ijltemas.2024.130710.
7. M. Alonge, "Securing National Critical Infrastructure: AI-Driven approaches to cyber threat detection," SSRN Electronic Journal, 2025. doi: 10.2139/ssrn.5389559.
8. D. Bhaskaran, "Zero Trust Architecture: Securing America's critical infrastructure," SSRN Electronic Journal, 2025. doi: 10.2139/ssrn.5145800.
9. S. Chitimoju, "The impact of AI in Zero-Trust security architectures: Challenges and innovations," 2024.
10. K. Chokkanathan et al., "AI-driven zero trust architecture: Enhancing cyber-security resilience," in 2024 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), 2024, pp. 1–6. doi: 10.1109/csitss64042.2024.10816746.
11. M. Faheem, M. Awais, A. Iqbal, and H. Zia, "Adaptive AI-driven cyber threat detection system for U.S. critical infrastructure protection," World Journal of Advanced Research and Reviews, vol. 26, no. 3, pp. 2282–2291, 2025. doi: 10.30574/wjarr.2025.26.3.2333.
12. J. S. Ibitoye, "Zero-trust cloud security architectures with AI-orchestrated policy enforcement for U.S. critical sectors," International Journal of Science and Engineering Applications, 2025. doi: 10.7753/ijsea1212.1019.
13. H. Joshi, "Emerging technologies driving zero trust maturity across industries," IEEE Open Journal of the Computer Society, vol. 6, pp. 25–36, 2025. doi: 10.1109/ojcs.2024.3505056.
14. Kolawole, "Leveraging cloud-based AI and zero trust architecture to enhance U.S. cybersecurity and counteract foreign threats," World Journal of Advanced Research and Reviews, vol. 25, no. 3, pp. 006–025, 2025. doi: 10.30574/wjarr.2025.25.3.0635.
15. M. Mangla and D. Kumar, "AI-Driven Zero Trust Architecture: A scalable framework for threat detection and adaptive access control," 2023. doi: 10.56127/ijst.v2i3.22.
16. E. Mmaduekwe, "AI-Driven cyber threat detection for securing national critical infrastructure," Asian Journal of Research in Computer Science, vol. 18, no. 6, pp. 424–431, 2025. doi: 10.9734/ajrcos/2025/v18i6711.
17. S. Mushtaq, M. Mohsin, and M. M. Mushtaq, "A systematic literature review on the implementation and challenges of zero trust architecture across domains," Sensors, vol. 25, no. 19, p. 6118, 2025. doi: 10.3390/s25196118.
18. K. Muthusamy, "Harnessing AI-Powered Zero Trust Architectures for proactive cyber defense: A comprehensive framework for future-ready network security ecosystems," International Journal of AI BigData Computational and Management Studies, vol. 6, pp. 22–29, 2025. doi: 10.63282/3050-9416.ijaibdcms-v6i1p103.
19. P. R. Nangi, C. K. R. N. Obannagari, and S. Settipi, "A multi-layered zero-trust security framework for cloud-native and distributed enterprise systems using AI-driven identity and access intelligence," International Journal of Emerging Trends in Computer Science and Information Technology, vol. 4, pp. 144–153, 2023. doi: 10.63282/3050-9246.ijetcsit-v4i3p115.
20. J. Reed, "Are attackers already embedded in U.S. critical infrastructure networks?," IBM, 2025.
21. Wolf, "The role of artificial intelligence in zero trust cybersecurity frameworks," Arctic Wolf, Feb. 3, 2026.
22. D. Xu et al., "The erosion of cybersecurity zero-trust principles through generative AI: A survey on the challenges and future directions," Journal of Cybersecurity and Privacy, vol. 5, no. 4, p. 87, 2025. doi: 10.3390/jcp5040087.
23. Y. Yigit et al., "Generative AI and LLMs for critical infrastructure protection: Evaluation benchmarks, agentic AI, challenges, and opportunities," Sensors, vol. 25, no. 6, p. 1666, 2025. doi: 10.3390/s25061666.

24. K. Zakhmi et al., “Evolving Zero Trust architectures for AI-driven cyber threats in healthcare and other high-risk data environments: A systematic review,” Cureus, vol. 17, no. 6, e85446, 2025. doi: 10.7759/cureus.85446.
25. C. Zanasi, S. Russo, and M. Colajanni, “Flexible zero trust architecture for the cybersecurity of industrial IoT infrastructures,” Ad Hoc Networks, vol. 156, p. 103414, 2024. doi: 10.1016/j.adhoc.2024.103414.