International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 4 (2025)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 4, 2025

Research Paper

A ZERO TRUST–BASED OPERATIONAL DEPLOYMENT ROADMAP FOR RESPONSIBLE AND SECURE AI INTEGRATION

Dr. Aniket Deshpande (Research Scholar)
Department of Computer Science and Engineering
Sunrise University, Alwar, Rajasthan, India

ABSTRACT:
Generative Artificial Intelligence (AI) has the potential to transform decision making in multiple industries, but requires to be securely, ethically and trust-centric frameworks in order to ensure it is used or deployed responsibly. This paper proposes a Zero Trust-Based Operational Implementation Framework for safe and responsible integration of AI that considers aspects of accountability and resilience in Generative AI to mitigate significant risks across sectors. This research addresses a persistent need for transparency, auditability, and ethical projected governance in AI-based decision making. The research framework is intended to present a grounded operational governance framework for deploying and governance of AI that is consistent with Zero Trust principles: continuous verification, human in the loop (HITL) and guarding. The research is qualitative in nature and is primarily reliant on secondary research of examples of AI in academic journals and institutions, recognizing that the examples reflect on protocols. The findings indicate that there is significantly less risk of misuse where human accountability and verifiable trust is integrated into Zero Trust controls at the identity, data, inference and action borders. Human accountability and verifiable trust would therefore greatly influence the sustainable governance of AI. The hierarchy of risk by supervision of the integration of AI is regarding levels of automations or not, but but supervision verification and trust-based protocols, thus ensuring safe, secure and ethical integration of AI technologies in critical settings.

Keywords: Zero – Trust Based Operational Deployment Framework; Secure AI Integration; Human-in-the-Loop; Zero Trust principles; Artificial Intelligence.

Received: 26-10-2025 Accepted: 08-12-2025 Published: 15-12-2025

INTRODUCTION:

The increasing use of Generative Artificial Intelligence (GenAI) across industries can yield incredible efficiencies, decision support, and knowledge. The deployment and use of GenAI in core workflows, though, increases additional risk vectors – erroneous conclusions, data privacy incidents or breaches, and unauthorized activities, each of which can bring detrimental financial, clinical or operational consequences. Traditional security strategies are incapable of addressing these concerns, and a zero-trust based continuous verification methodology is necessary with human-in-the-loop (HITL) enforcement and accountable augmentation (Inaganti et al., 2020; Joshi, 2024). An operational methodology for deployment supports a systematic approach that allows acceleration of AI enablement of decisions conservatively and does not compromise human expertise, regulatory compliance, or system outcomes. Organizations will need methods for AI risk management, as well as to maintain human accountability at decision action boundaries with organizational governance, monitoring, and industry-specific risk protection provisions (Li et al., 2025; Tiwari, 2022). This paper, ZT-GAI-CSMM, proposes a zero-trust based operational methodology for the responsible, safe, and resilient integration of Generative AI in varied organizational settings.

LITERATURE REVIEW:

Recent research has focused on the convergence of Zero Trust (ZT) principles and Artificial Intelligence (AI) to enhance cybersecurity resilience in cloud-native and enterprise-critical environments. Chattopadhyay (2025) highlighted how artificial intelligence, blockchain technology, and cyber intelligence have roles in enhancing the protection of DevOps pipelines for cloud applications. Muthusamy (2025) proposed an artificial intelligence-driven proactive ZT framework to forecast and mitigate future network risk with an emphasis on adaptation. In the context of cloud security across federal government organizations, Ofili, Erhabor, and Obasuyi (2025) evidenced that threat intelligence utilizing AI could be an important tool for complying with regulations, such as with CISA. Ibitoye (2023) investigated applying AI to orchestrate policy enforcement across several audiences in the US while constantly verifying access with risk mitigation. Meanwhile, Xu et al. (2025) noted that generative AI has the potential to extend ZT principles while creating a new set of vulnerabilities as automated decision-making policy enforcement. Together, the articles demonstrate the operational promise of AI as a partner of ZT architecture and discuss challenges in governance and interpretability.

Research Gap:

While there has been considerable emphasis on frameworks for AI-supported Zero Trust security, previous research has concentrated on technical design and compliance rather than on operational deployment methods for high-consequence workflows. There have been few studies that focus on how HITL (human-in-the-loop) procedures, risk constraints relevant to the specific sector, and real-time governance may lead to responsible augmentation while maintaining the quality of decisions. The applicability of AI-ZT frameworks across various sectors, and especially in relation to their ability to sustain a balance across speed, scalability, and security in BFSI, healthcare, ICS/OT, and public sectors has remained under-studied. This study is a call to researchers in this field to draw attention to the need for a legitimized, operational deployment roadmap that gives attention to both technological design, organizational accountability, and resilience.

METHODOLOGY:

A Zero Trust-based Generative AI operational deployment framework is established in this study, utilizing a design and implementation methodology. To facilitate full integration of AI, adopt Zero Trust security measures, and account for industry-specific safeguards, a comprehensive literature review is undertaken (Chattopadhyay, 2025; Xu et al., 2025; Muthusamy, 2025). Assessing five distinct organizational workflows in BFSI, healthcare, ICS/OT, government, and SMEs defines action boundaries, human-in-the-loop decision points, and evaluate governance. The framework provides an eight-step deployment structure that forces monitoring, escalation, and resilience cycles. Scenario-based testing and expert validation evaluate flexibility, accountability, and operational feasibility across sectors.

RESULTS AND DISCUSSION:

The deployment of the Zero-Trust Generative AI Consequential Safeguarding Maturity Model (ZT-GAI-CSMM) demonstrates responsible augmentation (vs. full automation) is the only means to maintain human accountability at the action boundaries (Figure 1). This methodology ensures that Generative AI can augment decision quality without compromising ethical, operational, or safety requirements for BFSI, Healthcare, ICS/OT, government, and SMEs.

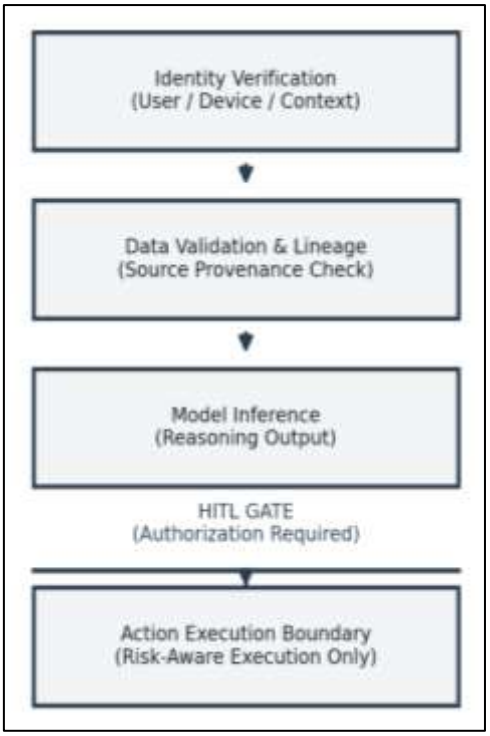

Figure 1: Action Boundary HITL Enforcement Model

The Universal 8-Step Rollout (Table 1) demonstrated that model inventory visibility, action boundary definitions, and HITL procedures were the most effective protections.

Table 1: Universal ZT-GAI-CSMM Deployment Blueprint (8 Steps)

No.	Step	Objective	Key Outputs	Accountable Owner
1	Model & Use Case Inventory	Establish visibility across all GenAI use	Model catalog, risk classification	Model Risk Management Lead
2	Define Action Boundaries	Identify where outputs influence real consequences	HITL thresholds, escalation criteria	Domain COO / BU Leadership
3	Data Lineage & Consent Mapping	Ensure data legitimacy and auditability	Source-of-truth lineage logs	Chief Data Officer
4	Guardrail & Prompt Policy Configuration	Restrict unsafe inference behavior	Prompt firewall rules, domain constraints	AI Platform Lead
5	Drift & Monitoring Infrastructure	Detect behavioral changes in real time	Drift dashboards, alerting paths	MLOps / Observability Lead
6	Workforce Training by Role	Ensure correct interpretation & use	Role-specific competency standards	Learning & Capability Lead
7	HITL Procedure Integration	Bind human authorization to action	Approval workflows tied to risk	BU COO / Clinical / Ops Supervisor
8	Assurance & Post-Incident Learning Loop	Convert incidents into structural improvement	RCA, updated guardrails, governance reports	Audit + Risk Council

Dual authorization at transaction points and strong identity verification reduced BFSI fraud (Table 2).

Table 2: Guiding Principles for Advancement Between Maturity Levels

To Move From	You Must Demonstrate	Evidence Required
Level 1 → Level 2	Visibility and intended-use control	Model inventory, purpose classification, owner assigned
Level 2 → Level 3	Governance is operating at the action boundary	HITL logs, escalation records, approval trails
Level 3 → Level 4	Resilience against drift and adversarial pressure	Drift monitoring dashboards, red-team reports, retraining audit records

To protect patients, clinical validation and hallucination suppression were prioritized in healthcare rollout, with doctor justification at every stage. CS/OT required two-person permission and fail-safe interlocks for operational adjustments' physical safety. SMEs prioritized guardrails, HITL, and basic monitoring to reduce critical risk and increase efficiency.

Maturity assessment showed businesses followed a behavioral curve rather than a technological one (Figure 2). Figure 3 shows how sector-specific safeguarding priorities targeted important controls by risk surface, potential impact, and accountability. Daily HITL checks, weekly guardrail evaluations, and quarterly governance assessments in the normal operating rhythm increased traceability and accountability (Table 8.9). Governance is operational hygiene, not a particular event, and exception handling protocols guaranteed events were studied, corrected, and turned into structural changes.

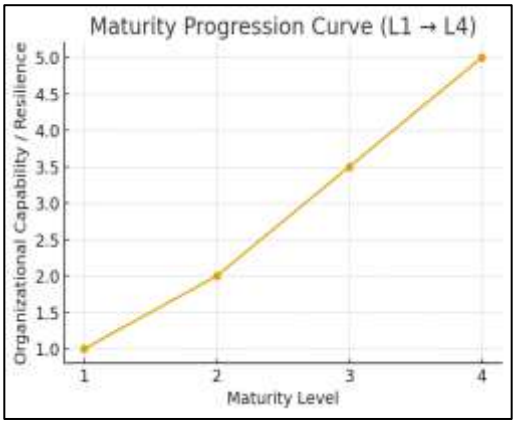

Figure 2: Generative AI Safeguarding Maturity Progression Curve (L1 → L4)

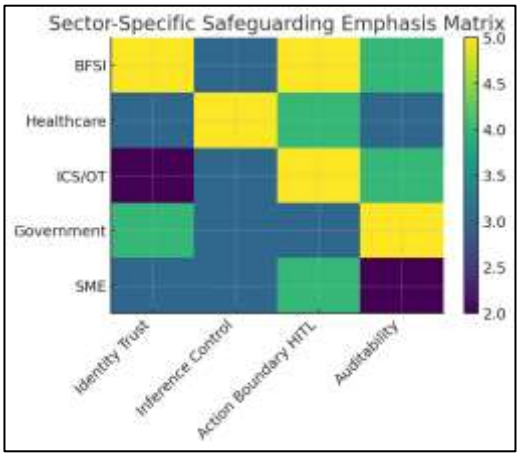

Figure 3: Sector-Specific Safeguarding Emphasis Matrix

The results show that ZT-GAI-CSMM is scalable, sector-agnostic, and uses AI to improve efficiency and insight while maintaining human accountability, operational safety, and organizational resilience.

ZT-GAI-CSMM confirms that AI-powered zero-trust architectures improve accountability, traceability, and security (Muthusamy, 2025; Ofili, Erhabor, & Obasuyi, 2025; Ibitoye, 2023). ZT-GAI-CSMM improves human-AI decision augmentation and organizational resilience by emphasizing sector-specific HITL enforcement, operational cadence, and maturity progression, unlike network- or cloud-focused frameworks.

CONCLUSION:

ZT-GAI-CSMM confirms that AI-powered zero-trust architectures improve accountability, traceability, and security (Muthusamy, 2025; Ofili, Erhabor, & Obasuyi, 2025; Ibitoye, 2023). ZT-GAI-CSMM improves human-AI decision augmentation and organizational resilience by emphasizing sector-specific HITL enforcement, operational

cadence, and maturity progression, unlike network- or cloud-focused frameworks.

REFERENCES:

[1]. Li, K., Li, C., Yuan, X., Li, S., Zou, S., Ahmed, S. S., ... & Akan, Ö. B. (2025). Zero-trust foundation models: A new paradigm for secure and collaborative artificial intelligence for internet of things. *IEEE Internet of Things Journal*.

[2]. Rosén, A. (2025). Achieving Zero Trust: A Strategic Roadmap for Phased Implementation of Zero Trust Architecture.

[3]. Inaganti, A. C., Sundaramurthy, S. K., Ravichandran, N., & Muppalaneni, R. (2020). Zero Trust to Intelligent Workflows: Redefining Enterprise Security and Operations with AI. *Artificial Intelligence and Machine Learning Review*, *1*(4), 12-24.

[4]. Tiwari, S., Sarma, W., & Srivastava, A. (2022). Integrating artificial intelligence with zero trust architecture: Enhancing adaptive security in modern cyber threat landscape. *International Journal of Research and Analytical Reviews*, *9*, 712-728.

[5]. Joshi, H. (2024). Emerging technologies driving zero trust maturity across industries. *IEEE Open Journal of the Computer Society*.

[6]. Chattopadhyay, B. C. (2025). Secure DevOps in Cloud-Native Systems: Integrating Cyber Intelligence, Blockchain, and AI for Zero-Trust Enterprise Applications. *International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management*, *1*(03), 44-50.

[7]. Xu, D., Gondal, I., Yi, X., Susnjak, T., Watters, P., & McIntosh, T. R. (2025). The erosion of cybersecurity zero-trust principles through generative ai: A survey on the challenges and future directions. *Journal of Cybersecurity and Privacy*, *5*(4), 87.

[8]. Muthusamy, K. (2025). 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*, *1*(1), 24-32.

[9]. Ofili, B. T., Erhabor, E. O., & Obasuyi, O. T. (2025). Enhancing Federal Cloud Security with AI: Zero Trust, Threat Intelligence, and CISA Compliance. *World Journal of Advanced Research and Review*.

[10]. Ibitoye, J. (2023). Zero-Trust cloud security architectures with AI-orchestrated policy enforcement for US critical sectors. *International Journal of Science and Engineering Applications*, *12*(12), 88-100.