International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

Leveraging Generative AI And Zero Trust Countermeasures In Critical Infrastructure: From Vulnerability To Resilience

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

ABSTRACT:
Generative artificial intelligence (GenAI) has had a transformative effect on the automation and decision-making responsibilities in numerous industry sectors, including healthcare, industrial control systems, and finance. However, the rapid expansion of GenAI has inherently added complexity to the global cybersecurity threat landscape. The objective of this research is to investigate how adversaries are leveraging the capabilities of GenAI to conduct scalable fraud, manipulate healthcare data, and subject security-sensitive operational processes to exploit. In order to support the objective, a qualitative analysis is performed that also utilizes sector-based risk mapping, adversarial scenario modeling, and safety system framework validation. The research found that continuous identity verification, context-aware Zero Trust controls, Human-in-the-Loop (HITL) monitoring, and data provenance assurance significantly mitigate vulnerabilities associated with natural-language processing (GenAI) systems. These contributions are significant for lawmakers, cybersecurity architects, and organizational leaders who want to construct reliable artificial intelligence governance to support critical infrastructure systems.

Keywords: *Generative AI; Zero Trust Measures; Critical Infrastructure; Vulnerability; Artificial Intelligence.*

INTRODUCTION:
Generative AI, or GenAI, is a seismic shift in artificial intelligence that redefines how machines generate, reason and interact with human systems. Rather than using symbolic reasoning and rescue of artificial intelligence expert systems, as has been the approach, we resorted to machine learning and deep neural networks, that then engage in autonomous adaptation (Haenlein & Kaplan, 2019). GenAI is a shift from job automation to creative synthesis, using latent representations of data to generate new language, graphics, programming code and designs (Sengar et al., 2024). As Bandi et al. (2023) comment, GenAI's architecture includes diffusion networks, transformer-based large language models (LLMs), and generative adversarial networks (GANs) that allow machines to simulate human-like processes and creativity. GenAI has a promising future in improving personalization, flexibility, and content origination in education, according to Ogunleye et al. (2024). The collaborative nature of GenAI, IoT, blockchain, and cloud ecosystems has also changed what is possible with data-driven automation and digital governance (Gill et al., 2019). Because of the rapid adoption of GenAI, there are concerns related to credibility, abuse, and safety of sensitive national domains that need governance to ensure the technology is reasonable, explainable, and ethical across a variety of critical infrastructure domains.

LITERATURE REVIEW:
The table that follows provides a detailed illustration of the previous literatures that are associated with this study.

Table 1: Related Studies

Authors and Year	Methodology	Findings (Connected to Paper Objective)
Ehtesham, A., Kumar, S., Singh, A., & Khoei, T. T. (2025)	SWOT analysis of Meta's Generative AI Foundation Model through qualitative and comparative evaluation in media and entertainment industries.	Identified GenAI's transformative power in digital content creation while exposing vulnerabilities in data authenticity and ethical governance.
Bender, S. (2025)	Qualitative sociocultural analysis examining GenAI's influence on human creative labor within media industries.	Found that AI automation erodes human oversight and accountability in creative workflows.

Dhoni, P., & Kumar, R. (2023)	Conceptual study integrating literature and policy reviews to explore GenAI's role in cybersecurity and governance.	Showed that GenAI can both strengthen and endanger cybersecurity.
Zarrar, H., & Kakar, S. A. (2024)	Comparative policy and defense analysis assessing GenAI's military applications in the U.S. and China.	Demonstrated how GenAI accelerates autonomous defense decision-making but increases geopolitical risk.
Göçen, A., & Asan, R. (2023)	Mixed-method literature review and risk-benefit evaluation of GenAI adoption in education.	Identified both opportunities and ethical risks related to data privacy and bias in AI-driven learning.

Research Gap

Even with considerable research on potential applications of Generative AI and ethical consequences within media, education, and defense, it remains unresolved how to enrich protection measures against AI-enabled exploitation to critical infrastructure. Although current research dominates functional benefits, creative disruption, and cybersecurity risks, there is not documentation on Zero Trust resilience frameworks. There is little empirical literature on how adaptive, human-in-the-loop governance may mitigate economic, therapeutic, and industrial risks. Useful context-based actionable models for the deployment of Generative AI in high-consequence settings are needed.

METHODOLOGY

This qualitative research study employs secondary data from peer-reviewed journals, conference proceedings, and institutional reports. A systematic review of the peer-reviewed literature on exploits of Generative AI, cybersecurity, and Zero Trust governance is undertaken. Case studies based on sectors such as the finance, healthcare, and industrial control sectors are then compared to identify weaknesses and resilience mechanisms specific to a sector. Thematic synthesis discusses the misuse of AI, the limitations of governance models through a theoretical approach, and protective solutions to data interpretation. The detailed area of interest for this qualitative publication reviews how adapted Zero Trust models and patterns of human behavior enhance resilience to GenAI exploitation and provides a basis for future development and practices.

RESULTS AND DISCUSSION

The application of the Zero Trust–Guided Generative AI Cybersecurity Maturity Model (ZT-GAI-CSMM) to Finance, Healthcare, and ICS/OT demonstrated unique yet interrelated resilience outcomes. Generative AI enhances operational efficiency and predictive capabilities, but its abuse poses threats to identity, data integrity, and physical safety. Each case study highlights the variations in adversarial exploitation patterns by sector, as well as modifications to their defences, based on a Zero Trust perspective.

Firstly, table 1 demonstrates that fraud, identity manipulation, and impersonation through deepfake technologies represent primary vulnerabilities in the BFSI domain. In a Zero Trust–based context, ZT-GAI-CSMM enhanced binding of identity, dynamic confidence scoring, and human in the loop (HITL) permissioning to the level of a transaction, thereby reducing the attack surface. The model also monitored and validated each financial decision ahead of implementation.

Table 1: BFSI Threats and ZT-GAI-CSMM Safeguards

Threat / Security Dimension	**GenAI Capability / Traditional Control**	**ZT-GAI-CSMM Enhanced Safeguard**	**Real-World Impact / Outcome**
Deepfake Impersonation	Text-to-Speech + Face Reenactment	Continuous identity confidence scoring	Prevents executive fraud and unauthorized transfers
Synthetic Identity Creation	AI-generated credentials & biometrics	Device/behavior-bound identity validation	Blocks KYC/AML evasion and false onboarding
Policy Evasion	Context-aware paraphrasing	Adaptive inference risk scoring	Stops compliance circumvention
AI Scam Operations	LLM-driven persona simulation	HITL gating with real-time trust checks	Reduces large-scale social engineering
Fraud Ring Coordination	Pattern-aware automation	Hallucination suppression + regulatory guardrails	Limits coordinated digital fraud attempts

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2, 2025

Table 2 illustrated the risk to patient safety in the healthcare and life sciences space created by GenAI abuse via manipulation of clinical data and hallucinated medical advice. The protection frameworks emphasized,in order: data provenance, ontology-based hallucination detection, and clinical oversight. As a result of these protections, clinician accountability, reduced diagnostic risk, and explainable AI in clinical decision-support systems would increase.

Table 2: Healthcare GenAI Threats and ZT-GAI-CSMM Safeguards

Threat / Dimension	**GenAI Capability / Pre-GenAI Condition**	**ZT-GAI-CSMM Safeguard Applied**	**Outcome / Real-World Impact**
Clinical Decision Manipulation	Prompt or inference subversion / Static CDS rules	Real-time inference risk scoring + justification traceability	Prevents misdiagnosis and unsafe treatment
Synthetic Medical Records	Template-based text generation / Manual data entry	Data lineage verification + authenticity validation	Stops identity fraud and data corruption
Bioinformatics Data Poisoning	Targeted dataset manipulation / Unverified research inputs	Secure data provenance + model integrity checks	Ensures valid research outcomes
Patient Interaction Misuse	Conversational AI / Scripted triage	Role-bound prompt guardrails + HITL review	Prevents unauthorized care guidance
Hospital Social Engineering	Clinical vocabulary phishing / Manual credential management	Access control reinforcement + anomaly detection	Protects EMR/EHR integrity and clinician credentials

Table 3 indicated that the inappropriate use of GenAI affecting critical infrastructure would lead to dangerous (or destructive) actuation or operational sabotage in ICS/OT. Authentication bound to hardware, simulation to validate actions, and defaults that were fail-safe allowed safety layers to not be bypassed in autonomous AI actions.

Table 3: Consolidated ICS/OT GenAI Safeguards and Control Enhancements

Safeguard / Dimension	**Pre-GenAI Condition**	**ZT-GAI-CSMM Enforcement**	**Outcome / Impact**
Identity & Access Trust	Basic login, manual control access	Hardware-bound authentication + operator locality verification	Blocks unauthorized prompts and control misuse
Data & Model Integrity	Unsigned sensor data, open training	Signed telemetry + isolated digital twin environments	Prevents unsafe learning or manipulation
Model Interaction Governance	Free-form operator reasoning	Approved prompt templates only	Ensures safe, compliant operational decisions
Output / Command Execution	Human-initiated SCADA actions	Two-person rule + model audit + HITL review	Eliminates single-point unsafe actions
Safety Interlock Management	Operator discretion	Dual authorization + traceable decision logs	Reinforces accountability and safety control
Fail-Safe Defaults	Manual emergency handling	Automatic protected shutdown when trust unproven	Maintains plant safety under attack or model fault

This research corroborates the conclusions of Bender (2025), Dhoni and Kumar (2023), and Zarrar and Kakar (2024). Bender (2025) explains the loss of human creative control in respect to automation using Generative AI, but this research restores human oversight through the Zero Trust–Guided Generative AI Cybersecurity Maturity Model. Dhoni and Kumar (2023) describe Generative AI's dual role for enhancing and violating cybersecurity, which is a specific exploitation that commonly identifies areas for both security and violations. Zarrar and Kakar (2024) characterize the geopolitical and defense vulnerabilities introduced by AI, which, like cyber vulnerabilities, must be protected against, without stifling adaptation to operational and consequences examples. The

comparison demonstrates across each sector, the combining of adaptive trust verification, accountability, and human judgment at the critical action border delivers sustainable AI resilience, without inhibiting Generative AI innovation.

CONCLUSION

To ensure the resilience of generative AI across vital sectors, adaptive Zero Trust enforcement and continual validation of identity, as well as mandated human oversight, are essential, according to the report. The ZT-GAI-CSMM methodology applies cybersecurity principles to operational governance by contextualizing safeguards for financial, healthcare, and ICS/OT use cases to facilitate trustworthy, accountable, and secure AI deployment in those sectors.

REFERENCES

[1]. Sengar, S. S., Hasan, A. B., Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: a systematic review and applications. *Multimedia Tools and Applications*, 1-40.

[2]. Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The power of generative ai: A review of requirements, models, input–output formats, evaluation metrics, and challenges. *Future Internet*, *15*(8), 260.

[3]. Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). A Systematic Review of Generative AI for Teaching and Learning Practice. *Education Sciences*, *14*(6), 636.

[4]. Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, *61*(4), 5-14.

[5]. Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., ... & Garraghan, P. (2019). Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, *8*, 100118.

[6]. Ehtesham, A., Kumar, S., Singh, A., & Khoei, T. T. (2025, January). Movie Gen: SWOT Analysis of Meta's Generative AI Foundation Model for Transforming Media Generation, Advertising, and Entertainment Industries. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 00189-00195). IEEE.

[7]. Bender, S. (2025). Generative-AI, the media industries, and the disappearance of human creative labour. *Media Practice and Education*, *26*(2), 200-217.

[8]. Dhoni, P., & Kumar, R. (2023). Synergizing generative ai and cybersecurity: Roles of generative ai entities, companies, agencies, and government in enhancing cybersecurity. *Authorea Preprints*.

[9]. Zarrar, H., & Kakar, S. A. (2024). Generative artificial intelligence & its military applications by the US and China–lessons for South Asia. *Journal of Computing & Biomedical Informatics*.

[10]. Göçen, A., & Asan, R. (2023). Generative artificial intelligence: Risks and benefits for educational institutions. *Center for Open Science. https://doi. org/10.31219/osf. io/mvcb5*