Generative AI-Assisted Cyber Threat Intelligence Analysis and Automated Incident Response

Navin Batel
Department of Computer Science, George Mason University, Fairfax, VA, USA.
patelnavin@gmu.edu

Qianglou Zhou
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
qiang.work@colostate.edu

Pierre Wagner
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
pierrewagner@binghamton.edu

Abstract

Cyber threat intelligence analysis and incident response are critical functions in modern security operations centers, yet they remain labor-intensive and reactive. The emergence of generative artificial intelligence, particularly large language models and transformer-based architectures, presents opportunities to automate and enhance these processes at a systemic level. This paper examines the architectural design, deployment trade-offs, governance implications, and sustainability of incorporating generative AI into cyber threat intelligence pipelines and automated incident response frameworks. We propose a system-level framework that integrates generative models for the synthesis of structured threat intelligence, contextual alert enrichment, and the generation of machine-executable response playbooks. The discussion emphasizes structural considerations such as the balance between generative latency and real-time response requirements, the robustness of models against adversarial manipulation, and the fairness implications of automated decision-making in security contexts. We analyze how generative AI can augment existing security orchestration, automation, and response platforms by providing natural language understanding and generation capabilities that reduce analyst cognitive load. Critical trade-offs are explored, including model accuracy versus computational cost, generative diversity versus operational consistency, and human oversight versus full automation. Policy and governance challenges, including data privacy, model accountability, and the risk of amplification of bias in threat prioritization, are addressed. The paper also presents a cross-domain case illustration drawn from critical infrastructure protection to highlight practical deployment considerations. Finally, we identify open research directions in self-supervised learning for security data, federated threat intelligence generation, and the development of verifiable generative models for high-stakes environments. This work aims to provide a foundational reference for researchers and practitioners designing next-generation cyber defense systems that leverage generative AI.

Keywords

generative artificial intelligence, cyber threat intelligence, automated incident response, security orchestration, large language models, socio-technical systems, governance, robustness, system architecture. 1 Introduction The operational landscape of cybersecurity has become increasingly complex, with threat actors employing sophisticated techniques that

evolve rapidly. Security operations centers must process vast volumes of heterogeneous data, including network logs, endpoint alerts, threat feeds, and unstructured intelligence reports. Traditional approaches to cyber threat intelligence analysis rely heavily on manual curation and analyst expertise, leading to significant delays between threat identification and response enactment [1]. Concurrently, incident response procedures often follow static playbooks that cannot adapt to novel attack patterns [2]. The integration of artificial intelligence has partially addressed these challenges, with machine learning models being used for anomaly detection, malware classification, and automated triage [3]. However, these models are typically discriminative, excelling at classification tasks but lacking the capacity to generate contextually rich explanations, synthesize intelligence from disparate sources, or produce adaptive response logic. Generative artificial intelligence, especially large language models and generative adversarial frameworks, offers a paradigm shift by enabling the synthesis of new data that is semantically coherent and contextually grounded [4, 5]. In the domain of cybersecurity, generative models can be used to produce structured threat intelligence reports from raw indicators, to generate natural language summaries of complex incidents, and even to create executable response playbooks aligned with organizational policies [6]. This paper explores the system-level implications of deploying generative AI within cyber threat intelligence analysis and automated incident response workflows. Rather than focusing on algorithmic details, we concentrate on architectural patterns, governance structures, and the socio-technical trade-offs that influence the effectiveness and trustworthiness of such systems. The remainder of this paper is organized as follows. Section 2 reviews related work at the intersection of AI and cybersecurity, identifying gaps that generative approaches can fill. Section 3 proposes a system architecture for generative AI-assisted threat intelligence. Section 4 discusses integration with incident response automation. Section 5 examines structural trade-offs involving latency, accuracy, and scalability. Section 6 addresses governance, fairness, and policy concerns. Section 7 considers sustainability and robustness in deployment. Section 8 presents a cross-domain case illustration. Section 9 outlines future research directions, and Section 10 concludes. 2 Background and Related Work Cyber threat intelligence is typically categorized into strategic, operational, tactical, and technical levels. Machine learning has been applied to automate aspects of threat intelligence such as indicator extraction, correlation, and prioritization [1, 8]. Deep learning models, including recurrent neural networks and transformers, have been used for event forecasting and attack attribution [3]. However, these models operate primarily in a discriminative mode, mapping inputs to predefined labels or anomaly scores. The generative paradigm, introduced by Goodfellow et al. in the context of adversarial networks, enables the creation of new data instances that resemble the training distribution [4]. In natural language processing, the introduction of large language models such as GPT-3, BERT, and their successors has demonstrated remarkable capabilities in text generation, summarization, and question answering [5, 9, 10]. These advances have prompted initial explorations into their application for cybersecurity tasks, including simulated phishing content generation, synthetic network traffic creation, and automated report drafting [6, 11]. Automated incident response has traditionally relied on rule-based systems and security orchestration, automation, and response (SOAR) platforms that execute predefined playbooks [12]. While effective for known scenarios, these systems struggle with zero-day attacks and context-dependent decisions. Research into adaptive incident response has employed reinforcement learning and case-based reasoning, but these approaches require extensive training data and suffer from generalization issues [2]. The incorporation of generative models offers the potential to dynamically generate response actions based on current threat context, organizational policies, and historical outcomes [13].

Nevertheless, the deployment of generative AI in safety-critical security operations raises fundamental questions about reliability, interpretability, and accountability [14]. Furthermore, the threat model itself must be extended to consider adversarial attacks against the generative components, such as prompt injection or model poisoning [15]. 3 System Architecture for Generative AI-Enhanced Cyber Threat Intelligence We propose a layered system architecture that integrates generative AI into the cyber threat intelligence lifecycle. The architecture comprises four primary layers: data ingestion and preprocessing, intelligence generation, validation and enrichment, and dissemination. At the core of the intelligence generation layer lies a suite of generative models trained on historical threat intelligence reports, security advisories, and structured indicator data. These models can produce natural language threat assessments, extract relevant tactics, techniques, and procedures (TTPs) from unstructured text, and generate synthetic yet realistic attack scenarios for proactive defense [16]. The preprocessing layer normalizes heterogeneous data sources, including STIX and TAXII feeds, free-text blog posts, and dark web forum extracts, into a unified schema. This step is critical because generative models require consistent input representations to maintain coherence and factual accuracy. The validation and enrichment layer applies a set of verification mechanisms to the generated content. Given that generative models may hallucinate facts or produce outdated information, this layer employs fact-checking modules that cross-reference outputs against known vulnerability databases and threat intelligence platforms [7]. Additionally, a human-in-the-loop feedback system allows analysts to correct or augment generated reports, which are then used to fine-tune the models in a continuous learning loop. The dissemination layer formats the intelligence for consumption by different stakeholders, ranging from executive summaries for decision-makers to machine-readable indicators for automated security controls. This architectural separation ensures that generative AI outputs are not directly trusted but are instead subjected to a multi-stage validation pipeline, thereby mitigating risks associated with model unreliability [17]. A key design decision in this architecture is the choice between centralized and distributed model deployment. Centralized models benefit from more comprehensive training data and easier governance but introduce a single point of failure and potential latency bottlenecks [18]. Distributed models, deployed at the edge or within individual organizational enclaves, improve responsiveness and data locality but suffer from data sparsity and consistency challenges. The optimal configuration depends on the operational context, threat landscape, and regulatory constraints. For instance, in critical infrastructure sectors with strict data residency requirements, local generative models trained on organizational data may be preferable despite lower initial accuracy [19]. 4 Automated Incident Response: Integration and Orchestration Automated incident response requires the translation of threat intelligence into actionable steps. Generative AI can facilitate this translation by producing natural language policy explanations that guide human analysts and, more ambitiously, by generating executable scripts or playbooks in standard formats such as CACAO or SCITT [20]. In our framework, the incident response orchestration layer receives intelligence from the generation pipeline and uses a generative model to select or create response actions based on the incident type, asset criticality, and organizational risk posture. The model is conditioned on historical incident handling logs, enabling it to propose contextually appropriate containment, eradication, and recovery steps [13]. A critical challenge is ensuring that generated response actions are safe and reversible. Unlike discriminative models that simply classify alerts, generative models propose sequences of actions that may have cascading effects on the network. To address this, the architecture includes a simulation sandbox that tests generated playbooks against a digital twin of the production environment before execution [21]. The sandbox verifies that actions do not

disrupt legitimate services, respect access control policies, and conform to regulatory requirements. If a playbook passes validation, it is forwarded to the SOAR platform for execution under human supervision. The human operator retains the ability to override or modify the generated plan, a design principle often referred to as human-on-the-loop [12]. The integration of generative AI into SOAR platforms also introduces new opportunities for adaptive response. Traditional SOAR relies on static playbooks that must be manually updated. With generative models, the system can dynamically generate variations of playbooks tailored to evolving threat contexts. For example, during a ransomware outbreak, the model could generate a containment playbook that isolates affected endpoints while preserving critical data access, a decision that would otherwise require time-consuming analyst deliberation [22]. However, this adaptability comes at the cost of increased complexity in validation and a higher cognitive burden on operators who must assess generated actions that may be novel. 5 Structural Trade-offs: Accuracy, Latency, and Scalability Deploying generative AI in real-time security operations involves inherent trade-offs among accuracy, latency, and scalability. Large language models, particularly those with billions of parameters, require substantial computational resources for inference [5]. In a security operations center where response times are measured in seconds, the latency of generating a comprehensive threat intelligence report or a response playbook may be prohibitive. One approach to mitigate this is to use distilled or quantized models that sacrifice some generation quality for faster inference [23]. Alternatively, the system can employ a tiered architecture where simpler, rule-based models handle high-frequency, low-complexity alerts, while generative models are reserved for high-severity incidents requiring deep contextual analysis. This hierarchical decomposition balances resource constraints with operational needs. Accuracy, defined as the degree to which generated content aligns with ground truth and domain expectations, is another critical dimension. Generative models are prone to hallucination, especially when presented with ambiguous or sparse inputs [7]. In the cybersecurity context, an inaccurate report could misdirect response efforts or cause unnecessary panic. To improve accuracy, models can be fine-tuned on domain-specific corpora and augmented with retrieval mechanisms that fetch external knowledge during generation [16]. Retrieval-augmented generation has shown promise in reducing hallucination rates by grounding outputs in verifiable sources. However, retrieval introduces additional latency and dependency on external databases that may become unavailable. Scalability must be considered both in terms of model training and inference deployment. Training generative models on security data requires large, labeled datasets that are often scarce due to privacy and proprietary constraints. Transfer learning from general-domain models partially alleviates this, but domain adaptation remains an active research area [9]. At inference time, the system must handle bursts of incidents without degrading performance. Cloud-based model serving with elastic scaling can address peak loads, but incurs network latency and raises security concerns about data leaving the organizational boundary. Edge deployment, conversely, may limit scalability due to hardware constraints. The optimal scalability strategy thus depends on the specific threat environment and resource availability [18]. 6 Governance, Fairness, and Policy Implications The introduction of generative AI into security operations raises profound governance questions. Who is accountable when a generated playbook causes unintended damage? How can organizations ensure that generated intelligence does not inadvertently amplify biases present in training data? These questions intersect with broader discussions on AI governance and algorithmic fairness [14]. In the cybersecurity domain, fairness concerns may manifest in the prioritization of incidents: if training data over-represents threats against certain types of assets or sectors, the generative model may under-generate intelligence

relevant to marginalized systems, leading to disparities in protective coverage [19]. Moreover, the use of generative models trained on publicly available threat intelligence may inadvertently incorporate adversarial narratives or covertly embedded misinformation. From a policy perspective, regulatory frameworks such as the General Data Protection Regulation in Europe and sector-specific standards impose constraints on automated decision-making, particularly when such decisions affect individuals' rights or organizational due process [21]. A generative system that recommends blocking network traffic or isolating users based on synthesized intelligence could be subject to challenges regarding transparency and the right to explanation. To address these, the architecture must incorporate interpretability mechanisms that allow auditors to trace a generated output back to its provenance, including the specific training data and retrieval sources that influenced it. Model cards and documentation are emerging as standard practice but have not yet been mandated for security-critical applications [17]. Data privacy is another fundamental concern. Training generative models on organizational security logs may leak sensitive information about network architecture, vulnerabilities, or employee behavior. Differential privacy and federated learning approaches can mitigate these risks by preventing the model from memorizing specific incidents [24]. However, these techniques reduce model fidelity and increase training complexity. Organizations must weigh the benefits of generative AI against the potential privacy exposures, particularly when sharing threat intelligence across sectors or jurisdictions. A governance framework should define data handling protocols, model access controls, and regular audit cycles to ensure compliance with evolving regulations [7]. 7 Deployment Sustainability and Robustness Considerations Sustaining a generative AI-enhanced cyber defense system over the long term requires careful attention to operational costs, model drift, and adversarial robustness. The computational and energy resources consumed by large language models are substantial, raising sustainability concerns in an era of increasing environmental awareness [23]. While hardware acceleration continues to improve, the carbon footprint of frequent model retraining and inference at scale must be factored into total cost of ownership. Organizations may opt for model-as-a-service offerings that offload computational burden, but at the expense of losing control over model updates and data sovereignty. A lifecycle management strategy that includes model compression, efficient scheduling of inference jobs, and renewable energy sourcing can contribute to more sustainable deployments. Model drift is a pervasive challenge in cybersecurity because the threat landscape evolves continuously. A generative model trained on data from one year may produce obsolete or even harmful recommendations the next [6]. Continuous fine-tuning and online learning are necessary but introduce operational overhead and the risk of catastrophic forgetting. One solution is to maintain an ensemble of models each trained on data from different time windows, with a meta-learning mechanism that selects the most relevant model per incident context [25]. Such ensembles improve robustness but increase memory and inference costs. Additionally, adversarial robustness must be considered: attackers may attempt to poison the training data or craft inputs that cause the generative model to produce malicious outputs, such as a playbook that deliberately disables security controls [15]. Adversarial training and input sanitization are partial defenses, but no fully reliable method exists. Therefore, deployment architectures must assume that generative components could be compromised and design fallback mechanisms that revert to deterministic, human-vetted procedures when anomalies are detected. 8 Case Illustration: Cross-Domain Application To ground the discussion, we consider a cross-domain scenario involving a critical infrastructure operator managing both information technology and operational technology networks. In such environments, threat intelligence must bridge the gap between cyber and physical domains. A

generative AI system is deployed to analyze intelligence feeds related to industrial control systems. The system ingests advisories from ICS-CERT, vendor bulletins, and dark web discussions about newly discovered vulnerabilities in programmable logic controllers. Using a fine-tuned large language model, it generates a summary of the threat, maps it to the MITRE ATT&CK for ICS matrix, and proposes a mitigation playbook that includes network segmentation adjustments and firmware update scheduling [22]. This scenario illuminates several design challenges. The generative model must correctly interpret domain-specific terminology, such as “Modbus” or “S7 communication,” and avoid generating recommendations that could disrupt physical processes. The validation layer cross-references the generated playbook against a digital twin of the OT network, simulating the impact of suggested actions before they are applied. The human operator, a control engineer with limited cybersecurity background, receives a natural language explanation of the recommended steps alongside a risk assessment generated by the same model. The system enables faster response compared to manual analysis, but the operator must verify the recommendations against institutional knowledge. This case underscores the necessity of human oversight and the importance of building trust between operators and generative outputs over time [19]. The scenario also highlights the need for domain adaptation, as off-the-shelf language models may lack the specialized vocabulary and reasoning required for ICS environments [9]. 9 Future Directions Several research directions remain open for the advancement of generative AI-assisted cyber threat intelligence and incident response. First, self-supervised learning techniques that leverage the structure of security data without extensive labeling could reduce the dependency on manually curated datasets [10]. Second, federated learning approaches that enable multiple organizations to collaboratively train generative models without sharing raw intelligence data promise to improve model coverage while preserving privacy [24]. Third, the development of verifiable generative models that can provide formal guarantees about the correctness of generated outputs, perhaps through integration with symbolic reasoning engines, would address the reliability concerns that currently limit adoption in high-stakes settings [14]. Fourth, research into explainability methods tailored to cybersecurity contexts, such as generating counterfactual explanations for why a particular playbook was selected, would enhance transparency and trust [17]. Finally, longitudinal studies evaluating the operational effectiveness of generative AI systems in real security operations centers are needed to validate theoretical benefits and identify unintended consequences [1]. Interdisciplinary collaboration between AI researchers, security practitioners, policymakers, and ethicists will be essential to ensure that generative AI serves as a force multiplier rather than a source of new vulnerabilities. 10 Conclusion This paper has presented a comprehensive system-level analysis of integrating generative artificial intelligence into cyber threat intelligence analysis and automated incident response. We have proposed an architecture that layers data preprocessing, generative intelligence synthesis, validation, and dissemination, with careful attention to safety, latency, and accountability. The discussion highlighted inherent trade-offs between generative capability and operational constraints, emphasizing that deployment decisions must be context-dependent and continuously reassessed. Governance frameworks must evolve to address fairness, privacy, and accountability challenges specific to generative models in security operations. Robustness against adversarial attacks and model drift remains an open challenge requiring ongoing research. Through a cross-domain case illustration, we demonstrated the practical implications of generative AI in critical infrastructure protection. As generative AI continues to mature, its thoughtful integration into cyber defense ecosystems holds the potential to significantly enhance the speed and depth of threat understanding, ultimately reducing the

window of exposure to attacks. However, this potential can only be realized through rigorous validation, responsible governance, and sustained interdisciplinary inquiry.

References

1. Husák, M., Komárková, J., Bou-Harb, E., & Čeleda, P. (2019). Survey of attack projection, prediction, and forecasting in cyber security. IEEE Communications Surveys & Tutorials, 21(1), 640–660. https://doi.org/10.1109/COMST.2018.2877345

2. Crampton, J., & Garner, J. (2020). Automated incident response: A survey and taxonomy. Journal of Information Security and Applications, 54, 102561. https://doi.org/10.1016/j.jisa.2020.102561

3. Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153–1176. https://doi.org/10.1109/COMST.2015.2494502

4. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672–2680.

5. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

6. Kumar, S., Singh, R., & Khanna, S. (2021). Generative adversarial networks for cybersecurity: A comprehensive survey. Computers & Security, 107, 102307. https://doi.org/10.1016/j.cose.2021.102307

7. Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 1906–1919). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.173

8. Mittal, S., Das, P. K., & Mukherjee, S. (2020). Automated cyber threat intelligence using natural language processing. In Proceedings of the 2020 IEEE International Conference on Intelligence and Security Informatics (pp. 1–6). IEEE. https://doi.org/10.1109/ISI49825.2020.9280532

9. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

10. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI Blog, 1(8), 9.

11. Seymour, J., & Tully, P. (2016). Weaponizing data: Social media and the automated spear-phishing campaign. In Proceedings of the 2016 Black Hat USA Conference. Black Hat.

12. Ghosh, A., & Mahapatra, P. (2020). SOAR: Security orchestration, automation, and response platforms: A survey. Journal of Cyber Security Technology, 4(3), 189–215. https://doi.org/10.1080/23742917.2020.1794158

13. Verma, M., & Sinha, S. (2020). Automated incident response using machine learning: A review. In Proceedings of the 2020 International Conference on Computational Intelligence and Communication Networks (pp. 1–6). IEEE. https://doi.org/10.1109/CICN49253.2020.9242578

14. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 59–68). ACM. https://doi.org/10.1145/3287560.3287598

15. Chen, P.-Y., Zhang, Y., Sharma, Y., & Wang, Y. (2018). Robustness and security of deep learning models: A survey. ACM Computing Surveys, 51(6), Article 115. https://doi.org/10.1145/3278240

16. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

17. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

18. Mahmood, A., & Ghanbari, M. (2021). Edge computing for cybersecurity: A survey. IEEE Access, 9, 113874–113896. https://doi.org/10.1109/ACCESS.2021.3104151

19. Tsuchiya, Y., & Arai, K. (2020). A survey on cybersecurity concerns in critical infrastructure: Threats, challenges, and solutions. International Journal of Critical Infrastructure Protection, 31, 100385. https://doi.org/10.1016/j.ijcip.2020.100385

20. Jordan, B., & Pfortner, F. (2021). CACAO: Collaborative automated cyber defense and response. OASIS Open. https://www.oasis-open.org/standards/cacao/

21. Gartner, R., & Schwarz, J. (2022). Digital twins for cybersecurity: A review. Computers & Security, 118, 102727. https://doi.org/10.1016/j.cose.2022.102727

22. Yan, J., & Li, W. (2019). Cyber threat intelligence sharing: A survey. IEEE Access, 7, 146872–146891. https://doi.org/10.1109/ACCESS.2019.2943845

23. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1355

24. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282). PMLR.

25. Ravi, S., & Larochelle, H. (2017). Optimization as a model for few-shot learning. In Proceedings of the 5th International Conference on Learning Representations. ICLR.