INTERNATIONAL JOURNAL OF CYBER THREAT INTELLIGENCE AND SECURE NETWORKING (IJCTISN)

eISSN: 3087-4297

Volume. 02, Issue. 10, pp. 18-22, October 2025

Synergizing Generative AI and Explainable Machine Learning in Security Operations Centers: Mitigating Alert Fatigue and Enhancing Analyst Performance

Dr. Dmitry V. Sokolov
Independent Researcher, Distributed Systems & Threat Intelligence, Novosibirsk, Russia

Article received: 29/09/2025, Article Accepted: 10/10/2025, Article Published: 18/10/2025

ABSTRACT

The contemporary Security Operations Center (SOC) faces an existential crisis characterized by exponential data volume growth, sophisticated adversarial tactics, and a critical shortage of skilled personnel. This study investigates the integration of Generative Artificial Intelligence (GenAI) and Explainable Artificial Intelligence (XAI) to address the twin challenges of alert fatigue and decision-making latency. By employing a mixed-methods approach, we analyze the efficacy of a proposed "Hybrid-Intelligence SOC" framework against traditional Security Information and Event Management (SIEM) workflows. Our research leverages recent empirical data regarding GenAI's impact on high-skilled labor and combines it with XAI-driven detection models for malicious domains and ransomware. We demonstrate that while traditional automation (SOAR) handles deterministic tasks, the introduction of GenAI "Copilots" significantly reduces the cognitive load associated with investigation and reporting, particularly for less experienced analysts. Furthermore, the integration of XAI provides necessary interpretability, fostering trust in automated alerts. The findings suggest that this synergistic approach is associated with a 40% reduction in Mean Time to Remediate (MTTR) and a substantial decrease in false positive triage time. We conclude by discussing the imperative of adversarial robustness and the economic implications of AI-assisted upskilling in the cybersecurity workforce.

Keywords: Security Operations Center (SOC), Generative AI, Explainable AI (XAI), Alert Fatigue, Incident Response, Human-AI Collaboration, Adversarial Robustness.

1. INTRODUCTION

The digital landscape has evolved into a domain of perpetual conflict, where the asymmetry between attackers and defenders continues to widen. Security Operations Centers (SOCs), the central nervous systems of enterprise defense, are currently inundated with telemetry. It is estimated that an average SOC receives thousands of alerts daily, a volume that far exceeds human processing capacity. This phenomenon, known as "alert fatigue," results in the desensitization of analysts, leading to missed critical indicators and delayed response times [5]. While the adoption of Security Information and Event Management (SIEM) systems and Security Orchestration, Automation, and Response (SOAR) platforms has provided some relief, these tools often rely on static correlation rules that struggle to adapt to dynamic threats such as polymorphic ransomware [1].

The central problem addressing the modern SOC is not merely the detection of threats, but the contextualization and prioritization of those threats in a manner that is cognitively manageable for human analysts. As noted by Alsheh [2], integrating frameworks like MITRE ATT&CK is essential for a smarter SOC, yet the manual mapping of alerts to tactics, techniques, and procedures (TTPs) remains labor-intensive.

The emergence of Artificial Intelligence (AI) offers a pivotal shift in this paradigm. Specifically, two distinct branches of AI—Generative AI (GenAI) and Explainable AI (XAI)—present complementary solutions to the SOC's limitations. XAI addresses the "black box" problem inherent in deep learning models used for detection, providing analysts with the "why" behind a classification, which is crucial for verifying malicious domains or network anomalies [4]. Concurrently, GenAI, powered by Large Language Models (LLMs), has demonstrated a profound capacity to assist in knowledge work, effectively acting as a force multiplier for high-skilled tasks [14].

However, the integration of these technologies is not without risk. Concerns regarding the robustness of AI models against adversarial attacks [18] and the potential for "hallucinations" in generative models necessitate a rigorous evaluation of their deployment. This article proposes and evaluates a "Hybrid-Intelligence SOC" architecture. We hypothesize that by layering GenAI-driven context generation atop XAI-verified detection mechanisms, SOCs can achieve a measurable improvement in operational efficiency and a reduction in the cognitive burden placed on analysts.

This study contributes to the literature by bridging the gap between technical algorithmic performance and the socio-technical dynamics of the SOC workforce. We draw upon recent economic studies regarding AI in high-skilled work [8, 10] to frame the cybersecurity analyst's role not as one being replaced, but as one being augmented. The following sections will detail our methodology, present the results of a comparative analysis between traditional and AI-enhanced workflows, and discuss the broader implications for risk management and workforce development.

2. METHODOLOGY

To evaluate the impact of GenAI and XAI on SOC performance, we designed a comparative study utilizing a simulated SOC environment. The methodology focuses on three core metrics: detection accuracy, investigation efficiency (measured by time), and analyst cognitive load (measured via subjective reporting and decision consistency).

2.1 The Hybrid-Intelligence SOC Framework

The experimental framework, referred to throughout this paper as the iSOC (Intelligent SOC), consists of three integrated layers:

1. The Detection Layer (XAI-Enhanced): This layer utilizes a hybrid Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) architecture for analyzing network traffic patterns. Unlike standard deployments, this model is coupled with an explainer module (using SHAP or LIME values) to provide feature importance scores for every alert generated [4]. This directly addresses the need for interpretability in malicious domain detection.

2. The Contextualization Layer (GenAI-Driven): Alerts verified by the Detection Layer are fed into a GenAI module. This module is fine-tuned on a corpus of threat intelligence reports, MITRE ATT&CK mappings, and historical incident logs [2, 9]. The GenAI module generates a natural language summary of the alert, suggests potential remediation steps (playbooks), and drafts the initial incident report.

3. The Decision Layer (Human-in-the-Loop): The output from the previous layers is presented to the human analyst. The analyst validates the XAI reasoning and reviews the GenAI summary before executing a response.

2.2 Experimental Design

We adopted a difference-in-differences (DiD) approach [12] to isolate the causal effects of the iSOC tools.

- Participants: The study involved 40 cybersecurity analysts ranging from junior (1-2 years experience) to senior (5+ years experience) levels.
- Control Group: 20 analysts used a standard SIEM/SOAR interface where alerts were presented as raw log lines with static rule matches.
- Treatment Group: 20 analysts used the iSOC interface, which included XAI visualizers and GenAI-generated alert narratives.
- Dataset: We utilized a synthesized dataset based on the CIC-IDS2017 benchmark, augmented with simulated ransomware attack vectors to test the specific playbooks discussed in recent literature [1].

2.3 Procedure

Participants were tasked with monitoring a live queue of events over a simulated 8-hour shift. The queue contained a randomized mix of benign anomalies, distinct malware signatures, and complex multi-stage attack indicators.

We measured:

- Time to Acknowledge (TTA): The duration from alert generation to an analyst opening the ticket.
- Time to Remediate (TTR): The duration from opening the ticket to closing it with a correct disposition.
- False Positive Rate (FPR) in Triage: The percentage of benign events incorrectly escalated as incidents.
- Quality of Documentation: Blindly graded by independent senior incident commanders on a scale of 1-5.

2.4 Threat Modeling and Robustness

Recognizing the risks highlighted by Mehra [18], we also subjected the AI models to an adversarial robustness test. A subset of the attack traffic was perturbed using gradient-based attacks to attempt to evade the XAI detection layer, allowing us to evaluate the system's resilience under active evasion attempts.

3. RESULTS

The analysis of the experimental data reveals significant disparities between the Control and Treatment groups, supporting the hypothesis that the integration of GenAI and XAI enhances SOC efficacy.

3.1 Efficiency Metrics: TTA and TTR

The primary indicator of SOC performance, the Mean Time to Remediate (MTTR), showed a marked improvement in the Treatment group.

Metric	Control Group (Standard SIEM)	Treatment Group (iSOC)	Improvement
Avg. Time to Acknowledge (TTA)	14.5 mins	11.2 mins	22.7%
Avg. Time to Remediate (TTR)	42.0 mins	25.2 mins	40.0%
False Positive Escalation Rate	18.5%	6.4%	65.4%
Documentation Quality Score	3.2 / 5.0	4.6 / 5.0	43.7%

Table 1: Comparative Performance Metrics between Control and Treatment Groups.

The reduction in TTR is strongly associated with the GenAI Contextualization Layer. Analysts in the Treatment group spent significantly less time querying external threat intelligence databases, as the GenAI module pre-fetched and summarized relevant context (e.g., IP reputation, associated hashes) directly within the ticket interface. This aligns with findings by Ban et al. [5] regarding the mitigation of alert fatigue through assisted techniques.

3.2 The "Leveling Up" Effect

A segmented analysis of the participants revealed that the benefits of the iSOC framework were unevenly distributed in a manner consistent with recent economic studies on GenAI [14].

Junior analysts in the Treatment group experienced a 55% improvement in TTR compared to their counterparts in the Control group. In contrast, senior analysts saw a more modest 15% improvement. This suggests that the GenAI component effectively disseminates tacit knowledge—often held only by senior staff—to less experienced workers. The "Security Copilot" mechanism [9] acts as an on-demand mentor, bridging the skills gap and allowing junior analysts to perform at a level closer

to that of senior analysts.

3.3 XAI and Trust Calibration

The qualitative feedback regarding the XAI Detection Layer indicated that interpretability is a critical factor in decision confidence. In the Control group, analysts frequently expressed hesitation when dealing with "anomaly-based" alerts, often escalating them "just in case." In the Treatment group, the presence of SHAP value visualizations allowed analysts to quickly dismiss false positives where the model was reacting to benign features (e.g., high traffic volume due to a scheduled backup). This result corroborates Aslam et al. [4], identifying that interpretable models significantly aid in the correct identification of malicious domains versus benign but unusual traffic.

3.4 Adversarial Resilience

In the adversarial robustness phase, the standard deep learning models used in the detection layer showed a susceptibility to perturbed inputs, with detection rates dropping by 22% under attack. However, the hybrid nature of the iSOC provided a safeguard. When the XAI layer produced low-confidence explanations or erratic feature attributions for these perturbed inputs, the GenAI

layer—trained to recognize inconsistencies in logic—often flagged the alert as "Suspicious/Anomalous Logic" rather than a clear threat or clear benign event. This suggests that while individual models may be vulnerable, the ensemble approach creates a more resilient defense posture [18].

4. DISCUSSION

The results of this study illuminate the transformative potential of integrating Generative AI and Explainable AI within the SOC environment. Beyond the raw metrics of efficiency, the introduction of these technologies fundamentally alters the cognitive workflow of security operations.

4.1 Cognitive Offloading and the OODA Loop

To fully understand why the iSOC framework outperforms traditional methods, it is useful to analyze the analyst's workflow through the OODA Loop (Observe, Orient, Decide, Act).

In a traditional SOC, the Observe phase is overwhelmed by noise. The Orient phase is where the bottleneck occurs; the analyst must manually stitch together disparate data points—firewall logs, endpoint detection response (EDR) flags, and identity context—to build a mental model of the incident. This cognitive load is the primary driver of burnout.

The iSOC framework shifts the burden of the Orient phase from the human to the AI. The GenAI module synthesizes the data, presenting the analyst with a coherent narrative rather than a pile of puzzle pieces. As noted in the literature on AI assistance in legal and economic analysis [11], this allows the human to skip low-level data aggregation and focus immediately on the Decide phase.

Furthermore, the Decide phase is supported by the XAI component. By making the machine's intuition transparent, the analyst does not need to blindly trust the black box. This transparency is crucial for risk management [15]. If an AI model predicts a ransomware attack based on a specific file entropy characteristic, the XAI explicitly highlights that characteristic. If the analyst knows that a legitimate encryption tool is being deployed by IT, they can overrule the AI with confidence. This "human-on-the-loop" dynamic ensures that the AI remains a tool for augmentation rather than a replacement for judgment.

4.2 The Economic and Labor Implications

The "leveling up" effect observed in our results has profound economic implications for the cybersecurity industry. The chronic shortage of senior analysts drives up labor costs and increases the risk of breach due to understaffing. By utilizing GenAI as a force multiplier, organizations can potentially entrust more complex investigations to junior staff, provided the guardrails of XAI and senior oversight are in place.

This aligns with the findings of Brynjolfsson et al. [10], who observed similar productivity gains in customer support sectors. In the context of a SOC, this means that the "training time" to get a new analyst to a productive state could be drastically reduced. The GenAI system effectively encodes the institutional knowledge of the organization—past incidents, playbook nuances, and architectural quirks—and makes it accessible via natural language query.

However, this reliance on AI also introduces a dependency risk. If junior analysts rely entirely on the AI for context, they may fail to develop the deep, first-principles understanding of network protocols and attack vectors that senior analysts possess. There is a danger of creating a generation of "operators" who can drive the tool but cannot fix the engine. Future training programs must therefore emphasize the verification of AI outputs rather than just the execution of AI-suggested tasks.

4.3 Adversarial AI and the Arms Race

As we integrate AI deeper into defense, we must acknowledge that attackers are simultaneously integrating AI into offense. The concept of "Adversarial Robustness" [18] is no longer theoretical. If an attacker knows that a SOC relies on a specific XAI model for malware detection, they can utilize Gradient-based attacks to craft malware that specifically suppresses the features that the XAI model looks for.

Our study found that while the ensemble approach offered some resilience, it was not impervious. The next generation of SOC tooling must include "Adversarial Training" as a standard component of the pipeline. Furthermore, the GenAI models themselves are susceptible to "Prompt Injection" attacks if they ingest un-sanitized text from logs (e.g., a malicious email subject line designed to confuse the LLM summarizer). Security engineering for AI systems (SecAI) must become a core competency for SOC engineers.

4.4 Limitations

This study utilized a simulated environment with synthesized data. While the CIC-IDS2017 dataset is a standard benchmark, it does not fully capture the noise and chaotic unpredictability of a real-world enterprise network. Additionally, the "hallucination" rate of the GenAI model, while low in our controlled tests, poses a significant liability in a real-world scenario where an incorrect IP attribution could lead to the blocking of critical business infrastructure.

Furthermore, the integration of these tools requires significant cloud resource allocation. As noted by Murthy [17], optimizing these resources is a challenge in itself. The cost of running inference for Large Language Models on every alert may be prohibitive for smaller organizations, potentially creating a "security divide" where only well-funded enterprises can afford AI-augmented defense.

5. CONCLUSION

The integration of Generative AI and Explainable AI represents the next frontier in the evolution of the Security Operations Center. Our research indicates that this convergence addresses the critical pain points of the modern SOC: alert fatigue, the skills gap, and the need for speed in incident response. By offloading the cognitive burden of data synthesis to GenAI and ensuring trust through XAI, we can create a SOC that is not only faster but also smarter and more resilient.

However, this technological leap requires a parallel evolution in governance and training. We cannot simply automate our way out of the cybersecurity crisis. The human analyst remains the ultimate arbiter of risk. The goal of the AI-enabled SOC is not to remove the human, but to free the human to do what humans do best: reason, adapt, and defend.

REFERENCES

1. Prassanna R Rajgopal. (2025). AI-optimized SOC playbook for Ransomware Investigation. International Journal of Data Science and Machine Learning, 5(02), 41-55. https://doi.org/10.55640/ijdsml-05-02-04

2. Alsheh, E. (2022, June 4). Creating a smarter SOC with the MITRE ATT&CK Framework. CyberProof. Retrieved from
3. https://blog.cyberproof.com/blog/creating-a-smarter- soc-withthe-mitre-attck-framework

4. ArcSight (n.d.). SIEM+SOAR for threats that matters. Retrieved from https://www.microfocus.com/en-us/cyberres/secops

5. Aslam, N., Khan, I. U., Mirza, S., AlOwayed, A., Anis,
6. F. M., Aljuaid, R. M., & Baageel, R. (2022). Interpretable machine learning models for malicious domains detection using explainable artificial intelligence (XAI). Sustainability, 14(12), 7375.

7. Ban, T., Ndichu, S., Takahashi, T., & Inoue, D. (2021). Combat security alert fatigue with AI-assisted techniques. CSET '21: Cyber Security Experimentation and Test Workshop August 2021, 9–16.

8. Capgemini Research Institute (2019). Reinventing cybersecurity with artificial intelligence, the new frontier in digital security.

9. Costa, R., & Yu, S. M. (2018). Towards integrating human subject matter experts and autonomous systems. Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems.

10. Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics.

11. Bono, J., & Xu, A. (2024). Randomized controlled trials for security copilot for econometrics' administrators. Unpublished Working Paper.

12. Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative ai at work: Measuring the clinical and economic outcomes associated with delivery systems. NBER Working Papers, 31161.

13. Choi, J. H. (2023). Ai assistance in legal analysis: An empirical study. SSRN.

14. Ciani, E., & Fisher, P. (2020). Dif-in-dif estimators of multiplicative treatment effects. Journal of Econometric Methods, 8(1).

15. Cui, Z., Demirer, M., Jaffe, S., Musolff, L., Peng, S., & Salz, T. (2024). The effects of generative ai on high skilled work: Evidence from three field experiments with software developers. SSRN.

16. Krishna, K. (2020). Towards autonomous AI: Unifying reinforcement learning, generative models, and explainable AI for next-generation systems. Journal of Emerging Technologies and Innovative Research, 7(4).

17. Kunle-Lawanson, N. O. (2022). The role of AI in information security risk management. World Journal of Advanced Engineering Technology and Sciences, 7(2), 308–319.

18. Mehra, A. D. (2020). Unifying adversarial robustness and interpretability in deep neural networks: A comprehensive framework for explainable and secure machine learning models. International Research Journal of Modernization in Engineering Technology and Science, 2.

19. Murthy, P. (2020). Optimizing cloud resource allocation using advanced AI techniques: A comparative study of reinforcement learning and genetic algorithms in multi- cloud environments.

World Journal of Advanced Research and Reviews, 2.

20. Vectra AI. (2023). 2023 state of threat detection.