Generative AI for cybersecurity

Improving SOC teams productivity

Sergei Lebedev

Current state of cybersecurity

The technological landscape of modern enterprise changes constantly, and organisations become more and more dependent on digital technology in general and cloud solutions in particular. While it allows for economical growth, Security Operations Center (SOC) teams report that it also increases the attack surface and the number of vulnerabilities that affect organisations (Vectra AI, Inc., 2023).

At the same time, SOC teams are overworked as is: the report by Vectra AI (2023) shows that 83 % of alerts that analysts receive are false positives, and 67 % of them are never processed. As a result, 67 % of them are considering or planning on leaving their jobs.

What could be improved?

Analysts surveyed by Morning Consult and IBM (2023) include automation and the use of AI for it as a possible solution for reviewing the large quantities of alerts, as well as more advanced tools that integrate better. 86 % of analysts believe AI-driven analytics would save some or a lot of time.

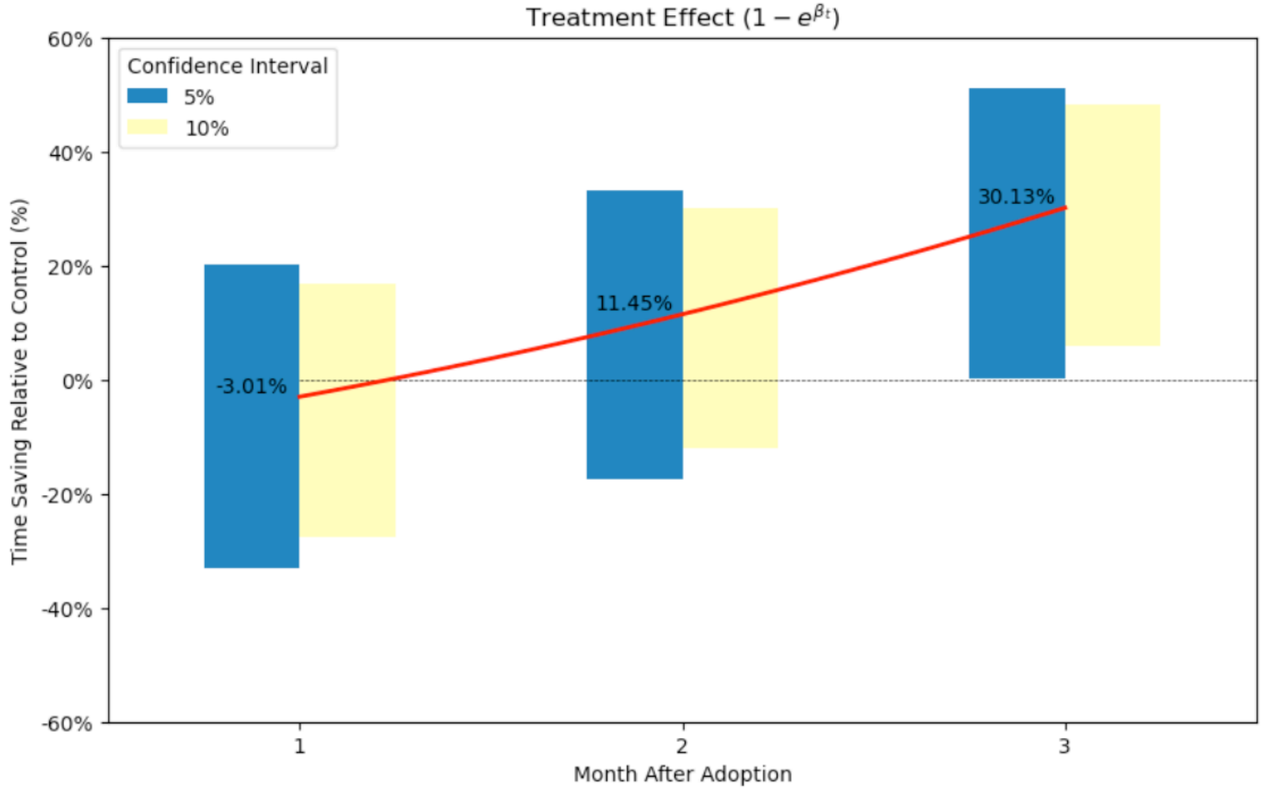
With what tasks can GenAI help?

- Analyse large amounts of threat data and gain insights from them
- Supplement monitoring and identify anomalies
- Detect phishing
- Automatically deploy countermeasures

Sources: Dhoni and Kumar (2023), Saddi et al. (2024)

Is there evidence?

Research by Bono, Grana, and Xu (2024) shows that the teams that adopted the AI-powered cybersecurity tool Microsoft Security Copilot saw an increase in their productivity, wherein time to resolution reduced by a third.



Time savings after introduction of GenAI-aided security tool (Bono, Grana and Xu, 2024)

Are there downsides?

GenAI is susceptible to prompt injection attacks where adversaries supply crafted input into the model to facilitate an incorrect response (Palo Alto Networks, no date). The risks become more pronounced when AI agents control the security infrastructure and can be instructed by adversaries to ignore real threats or raise the severity level for false positives, thus drawing the attention of the SOC team from an actual attack.

Protecting GenAI from prompt injection attacks

- Incorporating examples of attacks in the training process
- Data sanitisation: filtering out malicious inputs
- Modifying model training to prioritise robustness against the attacks

Source: Bathala and Babu (2024)

Conclusion

As any technology, GenAI is a tool that can both be useful when it comes to automating processes and improving workflows, and detrimental if applied without caution and knowing its limitations.

With the current state of corporate cybersecurity with overworked and burned out analysts that don’t have time or other resources to address the avalanches of alerts, introducing AI can hardly make the situation worse. Introducing AI for analysis, monitoring, and possibly managing some security infrastructure can bring relief to the SOC teams allowing them to work more productively and proactively.

References

Bathala, N.K. and Babu, G.V.R. (2024) ‘Adversarial Attacks on Large Language Models (LLMs) in Cybersecurity Applications: Detection, Mitigation, and Resilience Enhancement’, International Research Journal of Modernization in Engineering Technology and Science [Preprint]. Available at: https://doi.org/10.56726/IRJMETS61937.

Bono, J., Grana, J. and Xu, A. (2024) ‘Generative AI and Security Operations Center Productivity: Evidence from Live Operations’. arXiv. Available at: https://doi.org/10.48550/arXiv.2411.03116.

Dhoni, P. and Kumar, R. (2023) ‘Synergizing Generative AI and Cybersecurity: Roles of Generative AI Entities, Companies, Agencies, and Government in Enhancing Cybersecurity’. Available at: https://doi.org/10.36227/techrxiv.23968809.v1.

Morning Consult and IBM (2023) Global Security Operations Center Study Results. Available at: https://www.ibm.com/downloads/documents/us-en/10c31775a05401a5 (Accessed: 12 August 2025).

Palo Alto Networks (no date) What Is a Prompt Injection Attack?, Palo Alto Networks. Available at: https://www.paloaltonetworks.com/cyberpedia/what-is-a-prompt-injection-attack (Accessed: 4 August 2025).

Saddi, V.R. et al. (2024) ‘Examine the Role of Generative AI in Enhancing Threat Intelligence and Cyber Security Measures’, in 2024 2nd International Conference on Disruptive Technologies (ICDT), pp. 537–542. Available at: https://doi.org/10.1109/ICDT61202.2024.10489766.

Vectra AI, Inc. (2023) 2023 State of Threat Detection. Available at: https://www.vectra.ai/resources/2023-state-of-threat-detection (Accessed: 12 August 2025).