UDC 351.862.211.7:004.8

# FROM ALERTS TO ACTION: AI AGENTS IN MODERN SECURITY OPERATIONS

*Nasevich A.A.*

*Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus*

*Malikova I.G. – Senior Lecturer at the Department of Foreign Languages*

**Annotation.** With the rapid growth in telemetry volume, as well as the creation of new and increasingly sophisticated cyber threats, security operations centers (SOCs) are facing the problem of alert overload and reduced operational efficiency. This article examines the main architectural causes of this problem and analyzes the possibility of introducing AI agents into the SOC infrastructure. It describes the SOC architectural model in which AI agents have been implemented, as well as their advantages and limitations.

**Keywords:** AI agents, cybersecurity, security operation center, alert fatigue, artificial intelligence.

***Introduction*.** Due to the rapid development of information technology, cloud services, and distributed information systems, the nature of cybersecurity has changed significantly in recent years. Data comes from antivirus software, intrusion detection systems, and other monitoring systems. As a result, security operations centers (SOCs) have to process alerts almost continuously. Not all of these alerts indicate real threats; a significant portion of them are false positives and lead to a serious problem known as alert fatigue, which in turn can cause analysts to miss critical incidents. Traditional security systems based on rules and manual monitoring often prove outdated in the face of rapidly evolving cyber threats, including those that use artificial intelligence. Therefore, cybersecurity operations require more adaptive, modern, and intelligent approaches [3].

One of the most promising solutions is the use of artificial intelligence agents in security centers [1].

***Main part*.** Modern security monitoring centers use Security Information and Event Management (SIEM) platforms that centralize telemetry from various sources [4]. These sources can include cloud services, access control systems, servers, etc. The system generates alerts based on predefined rules, behavior and pattern analysis, and anomaly detection algorithms.

As digital infrastructure grows, the number of connected services and devices increases, with each new component adding data for processing. In large companies, the number of alerts can reach tens of thousands per day, with only a small fraction of them related to actual incidents. False positives require manual verification by level 1 analysts, and the constant processing of insignificant information increases the time to detect (MTTD) and respond (MTTR).

From an engineering perspective, the problem is a scaling imbalance: the volume of telemetry is growing very rapidly due to digital transformation, while SOC analytical resources are increasing at a slower rate. This creates a systemic bottleneck and leads to longer queues for event processing. As a result, the SOC is shifting from a proactive defense model to a reactive one, where the response occurs after an incident has already started or developed. Therefore, the problem of alert overload should be viewed as a systemic design problem that requires certain architectural changes.

The implementation of AI agents in SOC architecture significantly improves the efficiency of security incident handling [1]. Unlike standard methods, AI agents analyze data in context, are more flexible, and are able to adjust and adapt to changes in infrastructure.

The SOC architecture into which AI has been implemented is divided into several levels [4].

At the data level, telemetry from various sources, such as servers and cloud services, is processed.

At the analytical level, certain anomaly detection algorithms are applied, such as Isolation Forest and autoencoders. UEBA behavioral algorithms and graph neural networks (GNN) are also used.

At the orchestration level, AI agents can autonomously execute response scenarios, such as account blocking or node isolation.

At the same time, the analyst does not lose control of the situation; they perform the final verification and make the final decisions in a typical scenario (human-in-the-loop mechanism). Thus,

an AI-based SOC is a hybrid model where automation speeds up data processing and humans provide strategic control.

The use of AI agents in SOCs significantly improves the operational efficiency of security monitoring centers.

First, it reduces the time to detect and respond to incidents (MTTD and MTTR) [4]. Automatic prioritization and pre-processing of alerts reduces the burden on people and speeds up the escalation of critical events.

Second, it increases the scalability of the system; automation reduces the SOC's dependence on staffing levels, allowing monitoring to be scaled without hiring new employees [3].

Third, it reduces the number of false positives; machine learning models adapt to specific infrastructure and accurately distinguish between normal activity and potential threats.

As a result, fewer employees perform the same amount of work as in traditional security systems and can focus on more cognitively complex tasks, such as searching for hidden threats and investigating incidents. However, the effectiveness of AI implementation must be assessed taking into account the associated risks.

First, the accuracy of the models depends on the quantity and quality of the training data [2]. If the data contains errors or is insufficient, the agent may not work correctly, for example, it may miss an important attack.

Second, there are risks of adversarial attacks (notably data poisoning and evasion attacks), which can lead to classification errors that allow the security system to be bypassed [2].

Similarly, excessive automation without the necessary level of human control can lead to critical response errors.

***Conclusion*.** Thus, the implementation of AI agents requires not only technical configuration, but also a robust control system with audit and control mechanisms.

The growth of digital infrastructures and the increase in cyber threats require a modernized approach to SOC operations.

Traditional systems based solely on rules cannot cope with the growing volume of data and number of alerts. The integration of AI agents makes it possible to increase the speed and quality of monitoring, improve the incident prioritization system, reduce the number of level 1 analysts, and reduce the workload on the rest. At the same time, this implementation must be consistent and accompanied by quality control, risk management, and the preservation of the leading role of humans.

AI agents should be viewed as analytical tools within the human-in-the-loop model. The most promising approach is a hybrid SOC model, in which automation only speeds up data processing, while specialists provide strategic control and analysis of complex situations.

## *References*

*1. AI Agent Trends 2026 Report [Electronic resource]. Mode of access: https://cloud.google.com/resources/content/ai-agent-trends-2026. Date of access: 27.02.2026.*

*2. Maslej N., Fattorini L., Perrault R., Gil Y., Parli V., Kariuki N., Capstick E., Reuel A., Brynjolfsson E., Etchemendy J., Ligett K., Lyons T., Manyika J., Niebles J. C., Shoham Y., Wald R., Walsh T., Hamrah A., Santarlasci L., Lotufo J. B., Rome A., Shi A., Oak S. (2025) Artificial Intelligence Index Report 2025. Stanford Institute for Human-Centered Artificial Intelligence, Stanford University [Electronic resource]. Mode of access: https://hai.stanford.edu/ai-index/2025-aiindex-report. Date of access: 02.03.2026.*

*3. Milojicic D., Abedi A., Acero A., Ahamed S. I. et al. (2026) Technology Predictions 2026. IEEE Computer Society [Electronic resource]. Mode of access: https://ieeecs-media.computer.org/media/tech-news/tech-predictions-report-2026.pdf. Date of access: 03.03.2026.*

*4. Mohsin A., Janicke H., Ibrahim A., Sarker I. H., Camtepe S. (2025) A Unified Framework for Human-AI Collaboration in Security Operations Centers with Trusted Autonomy [Electronic resource]. Mode of access: https://arxiv.org/abs/2505.23397. Date of access: 05.03.2026.*