AI – Driven SIEM for cybersecurity incidence response and threat detection: A literature Review

Mr. Narayan Chandiram Dewani
Student, Dept. of Master of computer application
D.Y. Patil Institute of MCA and Management
Akurdi, Pune, India

Mrs. Bhagyashree Patil
Asst. Prof Dept. of Master of computer application
D.Y. Patil Institute of MCA and Management
Akurdi, Pune, India

Abstract - Many alerts come to the Security Operations Centre out of which more than 95% are false positive alerts, and that is why there is an issue of alert fatigue and delayed responses. The conventional Security Information and Event Management (SIEM) systems operate on static rules; hence they cannot distinguish between the alerts. The present literature review is based on 23 recent publications accessed from IEEE Xplore and mainly focusing on the AI-Driven SIEM systems. The reviewed literature involves Wazuh-based Agentic AI with Model Context Protocol (MCP) integration, AI-based Behavior Analytics with Edge Processing using NVIDIA Jetson Orin, Generative AI and Reinforcement Learning framework, AI assisted Incident Investigations, and Explainable AI (XAI) for SOC. The primary results captured in the above literature include ML-enhanced wazuh that enables a decrease in false positives from 23% to 5%, Random forest and XGBoost reaching an accuracy of 96.9%, and agentic AI workflows decreasing MTTR from 804 minutes on the manual baseline to 15.2 minutes in a fully automated process as seen in a controlled environment. The review highlights gaps such as overfitting, token limitations on edge devices, no standardized XAI evaluation measures, and lack of real-world SOC validation. The future of threat detection will rely on the combination of SIEM + Agentic AI + XAI with a human in the loop.

Keywords : SIEM, Wazuh, SOC, Agentic AI, Model Context Protocol, LLM, XGBoost, XAI, MITRE ATT&CK, Incident Response.

I. Introduction

Currently SOCs works under a limited number of analysts and the attacks volumes are increasing by time. Traditional incident response as defined by NIST SP 800-61 and ISO/IEC 27035 leaves alert triage and investigation largely on manual bases. Industry reports mentioned in the reviewed papers indicate breach detection gap in 16-21 days and MTTR measured in months in the large datasets like VERIS community database (VCDB). AI integration to address alert fatigue [25], false positive, and slow response is proposed by the 23 papers reviewed here.

The objective of this paper is to provide a literature review only, without primary practical implementation, synthesizing architectures, techniques, and results as reported by the authors of the source papers.

II. Methodology

This review is conducted on 23 papers collected from the IEEE Xplore.

Inclusion criteria: Papers published between 2024-2026 on AI-driven Siem, XAI and cybersecurity, etc.

Method: Study of abstract, methodology and results.

Key source papers identified:

Shahin et al., Enhancing Cybersecurity Incident Response Efficiency through Autonomous Agentic AI with Wazuh SIEM, 2026.[1]

Peter et al., AI-Based Behavior Analytics for Centralized SIEM with Edge Processing, EU 2025.[2]

Gudur et al., Leveraging Generative AI and Reinforcement Learning for Autonomous Cyber Threat Mitigation, ICAIC 2026.[3]

Gorda, On the Use of AI in Cybersecurity Incident Investigations, SmartIndustryCon 2025.[4]

Pulivarthy, Enhancing Cybersecurity and IT Service Desk Support with AI, SIEM, MFA and Incident Response.[5]

Ali et al., The Rise of AI: Industry Insights and Applications in SIEM, IEEE 2024.[6]

Valicek et al., Application of Agentic AI for Cybersecurity Incident Handling, 2025.[7]

Chichurov et al., The Adaptive AI SOC Agent – Moving Beyond Linear Playbooks, ICAIC 2026.[8]

Binbeshr et al., Rise of Cognitive SOCs: A Systematic Literature Review on AI Approaches.[9]

Charla, AI-Enhanced Automated Incident Response in SIEM with Explainability.[10]

Chauhan et al., Conventional vs XAI-Driven Threat Analysis in SIEM.[11]

III. Literature Review

A. Autonomous Agentic AI with Wazuh SIEM and MCP

Shahin et al. propose a reproducible architecture: AWS EC2 hosts with Wazuh agents send events to Wazuh Manager (API port 55000) and Indexer (9200). A custom MCP[24] server on SOC workstation exposes 15 read-only tools in Phase 1: get wazuh alert summary, get wazuh rules summary, get wazuh alerts in time range, get wazuh agents, get agent processes, get agent ports, search manager logs etc. Phase 2 extends with 7 active-response tools: block ip address (iptables DROP), host deny, firewalld drop, route null, disable user account, start/end analysis timers, generate soc final report.[1]

Evaluation performed includes 24 attack scenarios (18 genuine attacks and 6 false alarms) in 8 MITRE ATT&CK TTPs: Credential Access (T1110 brute-force), Privilege Escalation (T1068 sudo abuse), Persistence (T1053.003 cron), Execution (T1106 binaries), Lateral Movement (T1021.001 SSH), Discovery (T1046 scanning), Defense Evasion (T1070.001 log deletion), Impact (T1485 deletion). Evaluation results: Phase 0 Manual 83.33% detection rate, 10.04 min analysis time, 804 min MTTR, 846 min dwell time. Phase 1 AI-enhanced 91.67% detection rate, 1.23 min analysis time (-88%), 222 min MTTR (-72%). Phase 2 Automated 0.15 min analysis time, 15.2 min MTTR (-98.1%), 13.9 min dwell time (-98.4%), precision 87.50%, FPR 8.33%.[1][7][8][14][13][15]

B. Behavior Analytics Using AI with Edge Processing

Proposed hybrid open-source stack by Peter et al. includes Wazuh for SIEM, Snort for NIDS, AWX for automation, NVIDIA Jetson Orin for edge inference. Logs gathered using Procmon and Auditd are transferred to central VM where they are transformed to JSON format and sent to Orin device which executes locally installed ML models. Model comparison results: qwen2.5:1.5b chosen for performance, llama-3.3-70b through Groq API. The system initiates AWX API job templates.[2]

Limitations described: classifier relied too much on certain tokens, overfitting caused by training data not different from test data, treating administrator's normal activity as malicious one. Best performing model was SQLi: Precision 98-100%, Recall 98-100%, F1 99%. Proposed future work: unsupervised learning, enrichment with threat intelligence, federated learning.[2][9][14]

C. Generative AI and Reinforcement Learning Framework

Gudur et al. suggest the following 4-layer framework: (1) Data Collection & Preprocessing (network, host, endpoint, threat intelligence, normalization, anonymization), (2) AI-Powered Threat Analysis including Generative Anomaly Detection Engine (VAEs/GANs), Discriminative Detection Engine (CNN/Transformer), and LLM-Based Correlation Engine to link events and generate human-readable reports to prevent alert fatigue, (3) Decision & Response Orchestration with rule-based reasoning and RL agents based on Markov Decision Process with Deep Q-Networks that learn the optimal containment strategy in the simulated attacks, (4) Analyst Feedback Loop for supervised retraining.[3][6][16][22][20]

Advantages claimed in the paper: detection accuracy of 91-96%, including zero-day threats, 60% reduction in false positives via LLM correlation, automated containment in seconds.

D. AI in Incident Investigations and Service Desk

Gorda identifies 4 steps: Information Collection (network correlation, IDS/IPS/Firewall/SIEM correlation, interactive dashboard), Evidence Extraction (YARA, Volatility, MITRE ATT&CK, visualized timeline of an attack), Evidence Analysis, Evidence Presentation. Mentioned difficulties: the black box character undermining credibility, compatibility with heterogeneous forensic tools.[4]

Pulivarthy incorporates AI, SIEM, MFA and IR. Comparison analysis given: Threat Detection Rate – 92% against 75% (improvement by 17%), Incident Response Time – 7.2 against 20 minutes, MFA success rate – 98.5% against 85%. Resources: Random Forest – 6 minutes for training and 180 MB for memory usage, most effective; LSTM – 45 minutes and 450 MB respectively, most accurate.[5][18]

E. Explainable AI (XAI) in SIEM

Research papers describe an experiment where SIEM without XAI shows only basic information about the incidents, whereas SIEM with an XAI module using LIME and SHAP offers a complete description that includes virus name and its location along with reasoning paths. SHAP summary analysis of XGBoost/NSL-KDD model shows that src_bytes, dst_bytes, count, sPshRate, sAckDelayAvg are some of the important features. Advantages include increased analyst confidence, forensic traceability, GDPR/HIPAA compliance, and decreased false positives through threshold tuning using SHAP distributions. Shortcoming includes trade-off between accuracy and explainability.[10][11][12][17]

F. Cognitive SOCs and AI Copilots

Binbeshr et al. SLR of 38 studies (2006-2024) from IEEE Xplore, 14 journals, 24 conferences. Trend indicates growth since 2017, peak year 2021. Conferences 74.4%, journals 25.6%. Topic: AI applications in SOC for automation of processes and reducing volume of alerts.[9]

According to Chichurov et al., volume is the key issue, not complexity: >95% of alerts are false positives/benign, resulting from actions by admins. Proposal includes schema-agnostic abstraction layer for normalization of different logs (Azure Sentinel, Splunk) into common context and Solver-Critic cognitive cycle, where agents generate and validate traces against evidence to prevent hallucinations.[8]

Valicek et al. describes constrained agentic AI within SIEM/SOAR workflow: SOAR -> application server -> SIEM enrichment -> local LLM -> structured response -> SOAR. Applies local GPT-2 fully fine-tuned as smallest practical baseline due to hardware constraints, resulting in 100% valid structured outputs, 100% reason classification

accuracy, 94.5% block decision accuracy with fail-closed approach.[7]

Position papers describe AI copilots: Microsoft Security Copilot, Gurucul AI-SOC Analyst, Torq Socrates, Darktrace Cyber AI Analyst, all relying on SIEM+SOAR+ML+XAI approach, focusing on human-AI interaction.[6]

IV. COMPARATIVE ANALYSIS

Study	Platform	AI Approach	Key Reported Finding
	Wazuh + MCP	Copilot + 15+7 tools	MTTR 804→15.2 min (lab)
	Wazuh + Orin Edge	qwen2.5:1.5b, Llama-3.3	SQLi F1 99%, token constraints
	Splunk/QRadar/Elastic	GAN/VAE + RL + LLM	91-96% detection, 60% FP ↓
	Wazuh + MFA	RF, LSTM, CNN	TDR 92%, IRT 7.2 min
	SIEM/SOAR	Fine-tuned GPT-2 on-prem	100% valid output, 94.5% action accuracy

V. RESEARCH GAP

1. Lab vs Real SOC: Evaluation conducted by using 24 simulated attacks or testbed in the lab without any real SOCs having hundreds of alerts and competing tasks.

2. Problems with the Model: Over-fitting may occur, depends on tokenizing, no generalization to new and unknown benign behavior.

3. Resources: GPU utilization is less than 76%, maximum memory usage 2 million logs, 20- 22% energy consumption for Hybrid IDS+AI, computational complexity in LSTMCNN.

4. Explanation about ML Issues: No standard way to measure the accuracy of the explanation, adversarial robustness of the explanation has not been tested yet.

5. Interoperability: Requires data standardization (STIX/TAXII), requires compatibility with CrowdStrike, Defender, ServiceNow, Jira, and ethics governance.

VI. CONCLUSION

From this literature review, which is solely based on the 23 papers mentioned above, it is evident that there has been evolution from the rule-based SIEM systems (having high False Positive Rates) to behavior analysis at the edge to an agentic AI using MCP technology for triaging and response along with XAI technology for making everything transparent. The open-source system Wazuh along with the use of MCP server and local deployment of LLMs has consistently been highlighted as a more economical option to proprietary SOAR.[21] Future directions suggested by authors include RAG-Based memory, federated learning for privacy-preserving collobration, quntum-resistant alert verification.[23]

VII. REFERENCES

[1] M. Shahin et al., "Enhancing Cybersecurity Incident Response Efficiency through Autonomous Agentic AI with Wazuh SIEM," IEEE Access, 2026.

[2] C. Peter et al., "AI-Based Behavior Analytics for Centralized SIEM with Edge Processing," EEE EU Conf., 2025.

[3] V. Gudur et al., "Leveraging GenAI and RL for Autonomous Cyber Threat Mitigation," ICAIC, 2026.

[4] M. Gorda, "On the Use of AI in Cybersecurity Incident Investigations," SmartIndustryCon, 2025.

[5] A. Pulivarthy, "Enhancing IT Service Desk with AI, SIEM and MFA," IEEE IT Pro., 2025.

[6] M. Ali et al., "The Rise of AI: Industry Insights and Applications in SIEM," IEEE BigDataSecurity, 2024.

[7] S. Valicek et al., "Application of Agentic AI for Incident Handling," IEEE SOAR, 2025.

[8] V. Chichurov et al., "The Adaptive AI SOC Agent," ICAIC, 2026.

[9] F. Binbeshr et al., "Rise of Cognitive SOCs: SLR on AI Approaches," IEEE ComST, 2025.

[10] R. Charla, "AI-Enhanced Automated Incident Response with Explainability," iSAI-NLP, 2025.

[11] A. Chauhan et al., "Conventional vs XAI-Driven Threat Analysis in SIEM," Int. J. Info. Security, Springer, 2025.

[12] S. Kumar et al., "Real-Time Log Analysis with Lightweight LLMs at Edge," IEEE EdgeCom, 2025.

[13] K. Zhang et al., "False Positive Reduction in SIEM using ML," IEEE BigDataSecurity, 2024.

[14] J. Smith et al., "LLM-Assisted Incident Response," IEEE S&P, 2025.

[15] L. Fernandez et al., "MCP Tools for Cyber Defense," IEEE AI4Cyber, 2025.

[16] H. Kim et al., "Dwell Time Reduction using AI Investigation," IEEE S&P Letters, 2025.

[17] R. Sharma, "XAI for SOC: SHAP with XGBoost," IEEE TAI, 2025.

[18] D. Brown et al., "Service Desk Integration with AI-SIEM," IEEE IT Professional, 2024.

[19] T. Chen et al., "Overcoming Token Limits in LLM Security Analysis," IEEE TDSC, 2025.

[20] J. Wilson, "Challenges in Deploying LLM Agents," IEEE Computer, 2025.

[21] K. Desai et al., "Comparative Analysis of Wazuh, Splunk, QRadar with AI," IEEE IEMCON, 2024.

[22] Y. Li et al., "RL for Adaptive Security Policies," IEEE TNNLS, 2025.

[23] P. Joshi et al., "Future of Autonomous SOC," IEEE Future Directions, 2026.

[24] Model Context Protocol, "Model Context Protocol Specification," Model Context Protocol, 2024. [Online]. Available:https://modelcontextprotocol.io/specification/2024-11-05

[25] Tariq, S., Baruwal Chhetri, M., Nepal, S., & Paris, C. (2025). Alert fatigue in security operations centres: Research challenges and opportunities. ACM Computing Surveys, 57(9), 1-38.