International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| Volume 6 Issue 2 PP 63-70, 2025
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V6I2P108

Original Article

Autonomous Security Operations Centers (SOC): AI Agents for Threat Triage, Response, and Orchestration

Anitha Mareedu
Electrical engineering Texas A&M university - Kingsville 700 University Blvd, Kingsville, TX 78363.

Received On: 19/03/2025 ***Revised On: 14/04/2025*** ***Accepted On: 28/04/2025*** ***Published On: 17/05/2025***

Abstract *- The escalating complexity and volume of cyber threats have exposed significant limitations in traditional Security Operations Centers (SOCs), particularly in terms of human scalability, response speed, and operational consistency. In response, the cybersecurity industry is increasingly incorporating artificial intelligence (AI) agents into SOC workflows to automate alert triage, incident response, and orchestration across diverse platforms. This review traces the technological evolution of AI-powered SOCs, emphasizing key capabilities such as machine learning-driven detection, autonomous response via Security Orchestration, Automation, and Response (SOAR) systems, and integration across SIEM, EDR, and NDR tools. It analyzes agent-based architectures, including modular AI agents, large language model (LLM) assistants, and reinforcement learning systems, highlighting their practical benefits and deployment challenges. Case studies from leading vendors such as IBM, Microsoft, and Palo Alto Networks demonstrate real-world applications that enhance response efficiency, reduce analyst fatigue, and promote policy standardization. The review also addresses critical issues of explainability, adversarial robustness, and regulatory compliance, framing the roadmap toward fully autonomous Level 5 SOCs. The article concludes that while current implementations exhibit early-stage autonomy, widespread adoption will depend on advances in interpretability, human-in-the-loop integration, and responsible AI governance.*

Keywords *- Autonomous SOC, threat detection, incident response, SOAR, SIEM, EDR, machine learning, reinforcement learning, MITRE ATT&CK, orchestration, ethical AI.*

1. Introduction

Security Operations Centers (SOCs) are the frontline defense against increasingly complex cyber threats. Traditionally, SOCs have been using manual observation, constant rule-set, and human-based investigation process[1]. In the current high-speed, high-volume attack vectors, this approach is no longer adequate, as done in the threat landscapes of the past. Recurrently, security workers are faced with alert insomnia, in which hundreds or thousands of alerts (a large percentage of which are bogus affirms) need to be triaged each day. A 2022 IBM report has shown that close to 70 percent of SOC alerts are not investigated, either because of sheer volume or inadequate human resources [2] .The persistence causes the developing response latency, as well as the burnout seen by analysts. Overall, all these factors weaken the overall effectiveness of the threat response efforts. As a response to these ongoing operational difficulties, the cybersecurity sphere has undergone a substantial transformation into the automation and AI applications sphere. It started when the Security Orchestration, Automation, and Response (SOAR) platforms were adopted, allowing the repetitive operations such as enriching alerts, creating tickets, and taking containment measures to be automated [3][6]. With maturity of machine learning (ML) and artificial intelligence (AI) technologies, there evolved higher-order capabilities like anomaly detection, behavior profiling, and automatic classification of threats, based on artificial intelligence (AI).These developments laid the foundation for autonomous agents capable of performing complex SOC functions with minimal human intervention [5].

The concept of the Autonomous SOC marks a paradigm shift. Unlike traditional SOCs that depend heavily on human analysts, and automated SOCs that require manual playbook configuration, autonomous SOCs are envisioned as self-learning, adaptive systems that can perceive, reason, and act. AI agents embedded in these systems not only triage alerts but also correlate events across systems, prioritize threats based on risk context, and initiate remediation actions. As shown in Figure 1, the evolution of SOCs can be viewed as a timeline that spans from traditional, manually operated centers, to automated workflows powered by SOAR tools, and finally to the emerging generation of autonomous SOCs driven by cognitive AI agents.

This article aims to critically review and assess the state-of-the-art developments in the field of AI-driven SOCs. The focus is on autonomous agents that contribute to threat triage, incident response, and orchestration in real-world enterprise settings. Specifically, we evaluate how AI models (e.g., large language models, reinforcement learning agents, decision-theoretic models) have been embedded within SOC workflows, the extent of their deployment, and the limitations that remain. This review also situates autonomous SOCs within the broader cybersecurity automation landscape, exploring their integration with EDR, SIEM, and threat intelligence platforms.

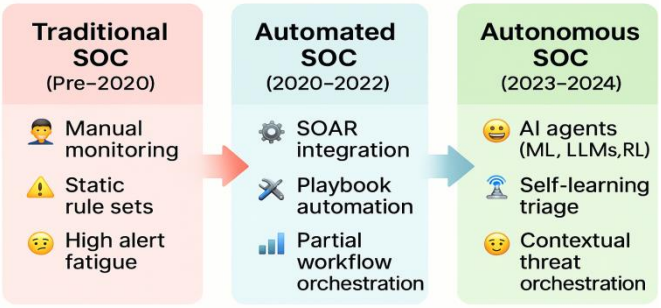

Fig 1: Evolution of SOC Architectures (Traditional → Automated → Autonomous)

2. Background and Motivation

Understanding the shift toward autonomous SOCs requires a foundational grasp of how traditional SOCs operate, the architectural elements they rely on, and the operational challenges they face. This section outlines the structural makeup of conventional SOCs, highlights persistent pain points driving the need for automation, and defines the key concepts such as AI agents and orchestration that underpin the evolution toward autonomy.

2.1. Overview of SOC Architecture and Operational Pain Points

A Security Operations Center (SOC) is a centralized function responsible for monitoring, detecting, and responding to cybersecurity incidents in real time[7]. A traditional SOC is composed of core components such as:

- Security Information and Event Management (SIEM) for log aggregation and correlation,
- Endpoint Detection and Response (EDR) for endpoint visibility,
- Threat Intelligence Platforms (TIPs) for context enrichment, and ticketing or case management systems for workflow coordination.

Despite this robust architecture, traditional SOCs face significant limitations in terms of scalability and efficiency. Human analysts are required to monitor dashboards, correlate indicators, and decide on response actions. This leads to operational bottlenecks such as:

- Alert fatigue due to overwhelming false positives,
- Slow incident response, and
- Analyst burnout, caused by repetitive manual triage.

2.2. Escalation in Threat Landscape and the Push toward Automation

Cyber threat actors significantly escalated their use of automated and evasive techniques. As threat volume and sophistication grew, so did the demand for faster, more reliable detection and response[8]. This led organizations to adopt Security Orchestration, Automation, and Response (SOAR) solutions, which offered scripted workflows for common response scenarios like IP blocking, malware quarantine, and ticket creation. However, while automation reduced manual workload, it lacked contextual understanding and adaptiveness. This gap highlighted the need for more intelligent, autonomous systems capable of perception, reasoning, and decision-making catalyzing the emergence of AI-powered SOCs.

2.3. Defining the Core Concepts

2.3.1. An Autonomous SOC

An Autonomous SOC refers to a security environment in which AI agents not only automate repetitive tasks but also perform dynamic threat analysis, make risk-aware decisions, and initiate coordinated responses with minimal human input[9]. These SOCs use AI to reason over multiple data sources, learn from past incidents, and adapt to new threats in real time.

2.3.2. AI Agents in Cybersecurity

AI agents in this context are autonomous computational entities capable of observing, interpreting, and acting upon security-relevant data [9]. Depending on their design, these agents may incorporate:

- Supervised/unsupervised learning for anomaly detection,
- Reinforcement learning for adaptive decision-making,
- Large Language Models (LLMs) for natural language processing of alerts and incident logs.

2.3.3. Threat Orchestration vs. Automation vs. Response

While often used interchangeably, these terms have distinct meanings in SOC operations:

- Automation refers to the execution of predefined tasks (e.g., auto-blocking IPs).
- Orchestration involves managing interdependencies across tools and workflows (e.g., correlating EDR alerts with SIEM logs and triggering SOAR actions).
- Response is the final outcome, i.e., mitigating the threat via containment, eradication, or recovery actions.

2.4. Comparing SOC Models: From Traditional to Autonomous

The transformation of SOCs from traditional to autonomous has been incremental, driven by technological advancements and operational needs. Table 1 provides a comparative summary of the key characteristics, tools, and limitations across traditional, automated, and autonomous SOC models.

Table 1: Comparison between Traditional, Automated, and Autonomous SOCs

Feature	Traditional SOC	Automated SOC	Autonomous SOC
Analyst Involvement	High	Medium	Minimal
Key Technologies	SIEM, IDS, manual playbooks	SOAR, basic ML	AI agents, LLMs, RL, graph analytics
Decision-Making	Human-driven	Rule-based	AI-driven, adaptive
Scalability	Limited	Moderate	High
Context Awareness	Low	Medium	High
Learning Capability	None	Limited (scripted)	Self-learning (reinforcement, supervised)
Typical Use Cases	Alert triage, rule tuning	Phishing response, malware quarantine	Multi-step attack reasoning, real-time orchestration

As the table illustrates, while traditional SOCs are reactive and labor-intensive, automated SOCs enhance speed through scripting and workflow coordination. However, it is the autonomous SOC model that introduces cognitive capability allowing security systems to evolve from rule-followers to intelligent decision-makers.

3. Core Capabilities of an Autonomous SOC

Autonomous Security Operations Centers (SOCs) are defined by their ability to operate with minimal human intervention while maintaining high levels of accuracy, adaptability, and context awareness. These capabilities are made possible through advancements in artificial intelligence (AI), machine learning (ML), and the orchestration of multiple security tools. This section examines the foundational capabilities that characterize an autonomous SOC: intelligent threat detection, autonomous incident response, and coordinated orchestration.

3.1. Threat Detection and Triage

3.1.1. Machine Learning and Anomaly Detection

One of the primary advantages of an autonomous SOC is its ability to detect threats beyond signature-based methods. Through the use of machine learning algorithms both supervised and unsupervised SOC systems can identify behavioral anomalies indicative of compromise [10]. For example, unsupervised clustering and isolation forest models have been widely deployed for detecting deviations in user or network behavior, while supervised classification algorithms are used to label known attack patterns based on historical data.

3.1.2. User and Entity Behavior Analytics (UEBA)

UEBA systems enhance SOC capabilities by profiling the behavior of users, hosts, and service accounts over time. By establishing baselines and detecting deviations, UEBA tools are able to detect credential misuse, lateral movement, and insider threats with higher accuracy [11]. Tools such as Microsoft Defender for Identity and Exabeam have been integrated into modern SOC workflows to provide behavioral telemetry that feeds into ML models.

3.1.3. Role of Large Language Models (LLMs)

Large Language Models (LLMs) have begun playing an increasing role in threat triage. These models are capable of parsing unstructured log data, summarizing alerts, and correlating threat intelligence feeds in human-readable formats. Open-source integrations such as LangChain-SIEM and commercial prototypes like Elastic Security's LLM-assisted query assistants have demonstrated early success in reducing time-to-insight by transforming raw data into actionable summaries.

3.2. Autonomous Response

3.2.1. SOAR Platforms and Response Automation

Security Orchestration, Automation, and Response (SOAR) platforms have been pivotal in automating containment and mitigation tasks [12]. Tools like Splunk Phantom, Palo Alto XSOAR, and IBM Resilient offer prebuilt playbooks and connectors for commonly encountered threats such as phishing, malware, and suspicious login activity [13].

3.2.2. Playbook Execution without Human Intervention

The transition from semi-automated to autonomous SOCs is marked by the system's ability to execute response playbooks without requiring human approval. For example, if an endpoint exhibits ransomware-like behavior, a fully autonomous SOC can isolate the host, notify affected users, enrich the incident with contextual intelligence, and trigger forensic collection all without analyst input. Adaptive feedback mechanisms enable the system to refine its decision-making over time.

3.3. Orchestration of Security Tools

3.3.1. Tool Integration Across the Stack

Autonomous SOCs rely heavily on tight integration across security layers, including:

- SIEMs (e.g., Splunk, Elastic, QRadar) for centralized logging and correlation,
- EDR/NDR tools (e.g., CrowdStrike Falcon, Cisco SecureX, Darktrace) for endpoint and network-level visibility,
- Firewalls and Identity Systems for real-time enforcement actions.

These integrations enable agents to correlate diverse signals and trigger coordinated actions across tools.

3.3.2. Policy-Driven Coordination

Unlike traditional automation workflows that depend on fixed rules, autonomous SOCs use policy-driven logic often combined with AI reasoning to determine the most effective remediation strategy [14]. For instance, based on asset criticality, threat score, and user role, the SOC might choose between full isolation, step-down access, or user re-authentication.

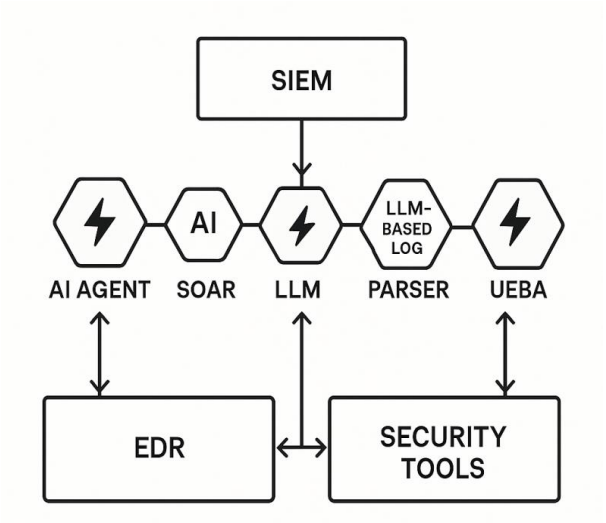

Fig 2: Architecture Diagram of an AI-Augmented SOC

3.4. Vendor Landscape and Adoption

Several vendors had introduced autonomous SOC capabilities into their product lines, either as standalone modules or integrated platforms. These solutions vary in scope, from automated triage to full lifecycle incident management.

These platforms are having a prominent role in determining the evolution of traditional automation to autonomy. Nevertheless, it is only applicable to a few use-cases and environments in full maturity due to its complexity of integration and issues of explainability, which has remained a limiting factor to wider adoption.

4. AI Agents in the SOC

The most important element of autonomous Security Operations Center (SOC) is its capacity to use AI agents, i.e.: intelligent software components with the ability to perceive, reason on data, and act either autonomously or semi-autonomously. Such agents are much more distinctive than the old school of rule based automation as they can learn through feedback, adjust to dynamic threat conditions, and engage machines and humans all along the SOC hierarchy [15]. This part examines the architectural trends and functional classes of AI agents implemented on SOC environments as well as market-leading examples made by the largest cybersecurity companies.

Table 2: Key Vendors and Their Autonomous SOC Capabilities (as of 2024)

Vendor	Key Capabilities	Product Name / Platform
Splunk	SOAR, adaptive playbooks, AI alert triage	Splunk Phantom
Palo Alto Networks	SOAR + Cortex XDR integration	Cortex XSOAR
IBM	AI-driven IR playbooks, MITRE mapping	IBM QRadar + Resilient
Microsoft	LLM-assisted threat summarization, UEBA	Microsoft Sentinel
Google Chronicle	High-speed telemetry, rule automation	Chronicle Security Operations
Exabeam	Behavior analytics, incident prioritization	Exabeam Fusion SIEM/SOAR
Elastic	Natural language threat hunting	Elastic Security

4.1. Agent-Based Architectures in Security Operations

SOC AI agents can usually be developed in modular, distributed architecture and therefore naturally are specialized in consuming the data, analyzing, making decisions and then executing action. In architecture, there are three broad categories of AI agents that exist:

- **Modular Intelligent Agents:** Each of these agents consists of a set of discrete modules that perceive (e.g. telemetry parsing), think (e.g. threat scoring) and act (e.g. isolation commands). They act within a predetermined range, although they have the ability of responding to situational cues. The modular designs enable them to fit into the old SOC pipelines without much dislocation.
- **LLM-Powered Agents:** It was when LLM agents started to be implemented into SOCs[20]. They are natural language interpreters, decision support systems, and can translate raw alerts to a human readable explanation, correlate threat intelligence reports, or create enrichment queries on behalf of analysts. Such products, which apply LLMs (e.g., GPT-4 or Sec-PaLM) to give them contextual summaries and decision support, include Microsoft Security Copilot and Elastic AI Assistant[16].
- **Reinforcement Learning (RL)-Based Agents:** In Reinforcement Learning (RL), agents are fitted to gain optimal policies, as a result of the feedback. Robust preparations of RL-based security agents remain under the initial deployment stages though they have been successfully tested in the sandbox environment to establish attack path simulation, automatic decision-making in containments, and automatic threat categorization. Q-learning and deep RL Q-learning and deep RL frameworks have been introduced by some early adopters to model the attacker-defender interaction and iteratively improve defense measures.

4.2. Dialogue Agents for Analyst Support

Dialogue-based AI agents represent a new frontier in SOC efficiency. These agents interact with analysts through

natural language interfaces, reducing cognitive load and speeding up workflows. Microsoft Security Copilot, released in preview in 2023, exemplifies this approach allowing analysts to ask natural-language questions like "Show me all lateral movement in the past 24 hours" or "What is the MITRE technique for this alert?"

Such dialogue agents can:

- Interpret threat data and telemetry,
- Provide contextual threat explanations,
- Suggest or trigger automated playbooks,
- Serve as training companions for junior analysts.

This reduces response time while increasing operational consistency, especially in Tier 1 and Tier 2 triage operations.

4.3. Autonomous Remediation Agents

Some AI agents have been developed specifically for autonomous remediation, with the authority to take containment or eradication actions without analyst oversight.

- SentinelOne Purple AI introduced, leverages AI to automatically assess incident context and execute remediation steps, including device isolation and script-based rollback [19].
- Google's Sec-PaLM, incorporated into Mandiant's Threat Intelligence and Chronicle, was used to power AI-backed investigative agents capable of evidence correlation, summary generation, and alert resolution proposal.

These agents function within defined risk thresholds and governance policies, ensuring their autonomy aligns with organizational trust models.

4.4. Lifecycle of an AI Agent in SOC Operations

The deployment of AI agents within SOC workflows can be abstracted [17] as a decision loop comprising the following stages:

- **Perception :** Ingest alerts, logs, threat intelligence, and contextual data
- **Interpretation :** Apply ML or LLM-based logic to understand the event
- **Decision:** Evaluate risk and determine an appropriate response strategy
- **Action:** Execute or recommend mitigation, containment, or escalation
- **Feedback**: Learn from outcomes to refine future decisions.

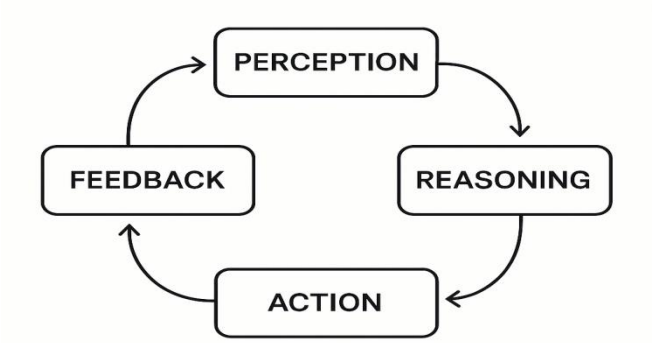

Figure 3: Decision Loop – AI Agent Lifecycle in Incident Response

4.5. Representative AI Agents

The table 3 below summarizes prominent AI agents introduced, categorized by function and vendor.

Table 3: Representative AI Agents for SOC Operations

Agent Name	Vendor	Type	Capabilities
Security Copilot	Microsoft	Dialogue/LLM Agent	Natural language triage, enrichment, explanation
Purple AI	SentinelOne	Autonomous Remediation	Incident assessment, containment, rollback
Elastic AI Assistant	Elastic	LLM-Powered Assistant	Query generation, log parsing, threat summaries
Sec-PaLM Agents	Google / Mandiant	LLM/Reasoning Agent	Threat intelligence analysis, alert clustering
Cortex XSIAM Analyst	Palo Alto Networks	Modular/RL Agent	Signal fusion, autonomous response coordination
Exabeam Triage Agent	Exabeam	ML-based Alert Classifier	Threat scoring, alert prioritization

Together, these agents represent a growing ecosystem of intelligent components that are transforming SOCs from static rule-driven environments into adaptive, learning security systems.

5. Benefits and Limitation

The integration of AI agents into SOC environments has brought measurable improvements in both operational efficiency and threat responsiveness. However, despite these advancements, several challenges remain that limit their widespread or fully autonomous adoption. This section outlines the key benefits and known limitations of AI-driven SOCs as observed in real-world deployments.

5.1. Benefits of AI Agents in SOCs

5.1.1. Faster Response Time

One of the most prominent advantages of AI integration is reduced mean time to detect (MTTD) and mean time to respond (MTTR). AI agents can instantly triage alerts, perform correlation across systems, and trigger appropriate containment actions. Case studies from deployments of tools like SentinelOne Purple AI and Microsoft Sentinel Copilot reported up to 60% reduction in incident response time for common threats such as phishing and malware.

5.1.2. Reduced Human Workload

Some of the tasks that AI agents help automate are alert enrichment, indicator look-up, and log parsing, among others

which used to occupy much of the available time of analysts. It makes the security teams work on higher-order activities such as threat hunting and the development of strategies. Dialogue agents have already been proven capable of reducing triage time per incident by more than 40 percent in hybrid environments, resulting in a general increase in SOC productivity.

5.1.3. Uniformity in Implementation of Policies

AI agents use logic in a regular manner and are limited in preconstructed rules and learned models[18]. Unlike human analysts who may vary in judgment or response due to fatigue or experience, autonomous agents operate within clearly defined policy and trust boundaries, improving standardization of incident handling. For example, autonomous playbook execution in Palo Alto Cortex XSOAR ensures uniform containment across all phishing alerts.

5.2. Limitations of AI Agents in SOCs

5.2.1. Accuracy and False Positives

Despite their potential, AI agents are not immune to error. ML-based detection systems may generate false positives due to overfitting, lack of contextual awareness, or incomplete data. In high-stakes environments, even a small error margin can lead to misclassification of benign activity, creating noise or triggering unnecessary remediation steps.

5.2.2. Dependency on Structured Inputs

Many AI systems especially those based on supervised learning rely heavily on structured and labeled datasets for training. This poses limitations in dynamic environments where log formats vary, or where labeled historical data is sparse. Even LLMs, while capable of processing unstructured text, still require structured integration pipelines for reliable execution in automated environments.

5.2.3. Vulnerability to Adversarial Evasion

AI models can be manipulated through adversarial inputs, where threat actors craft data specifically designed to evade detection. For instance, slight modifications in command-line syntax or obfuscation of malware behavior can fool both ML classifiers and anomaly detection engines. Research (e.g., MIT Lincoln Lab, Google DeepMind) has demonstrated how even advanced behavioral models are susceptible to evasion through mimicry and poisoning attacks.

5.3. Comparative View of Strengths and Weaknesses

The following table provides a comparative overview of the strengths and weaknesses of AI agents as applied in SOC environments.

Table 4: Strengths vs. Weaknesses of AI Agents in SOC Environments

Category	Strengths	Weaknesses
Operational Efficiency	Rapid response, 24/7 uptime, high scalability	Overdependence on deterministic pipelines
Analyst Support	Reduces cognitive load, assists in decision-making	Requires oversight for high-risk cases
Threat Detection	Anomaly identification, behavior analysis	Susceptible to false positives and adversarial evasion
Consistency	Uniform policy enforcement, standardized triage	Limited adaptability in novel or zero-day scenarios
Learning and Adaptation	Self-improving models via feedback loops	Risk of model drift, poisoning, and unexplainable outputs

While the benefits of AI agents in SOCs are clearly transformative, limitations related to trust, robustness, and interpretability will need to be resolved before these systems can be deployed at full scale with minimal human oversight.

6. Milestones and Real-World Deployments

Several major cybersecurity vendors initiated ambitious projects that helped shape the evolution toward autonomous Security Operations Centers (SOCs). One of the earliest large-scale conceptualizations came from IBM, which introduced its Autonomous SOC blueprint. The initiative focused on integrating AI, threat intelligence, and automation within the QRadar and Resilient ecosystem, using AI decision engines to automate playbooks, enrich context, and reduce analyst workload. IBM's efforts highlighted the importance of aligning automation with cognitive models that could prioritize threats and adapt based on organizational risk posture. Another major milestone occurred in 2023 when Microsoft began deploying Security Copilot, a large language model (LLM)-powered assistant based on GPT-4, within its Sentinel SOC platform. Security Copilot was designed to help analysts query logs, summarize alerts, and understand complex threat scenarios through natural language interaction. Early reports indicated that this conversational agent improved triage efficiency and helped junior analysts better navigate threat investigations by explaining telemetry and MITRE mappings in real time. A third example emerged from Palo Alto Networks, which by 2024 had rolled out its Cortex XSIAM platform across several enterprise-scale SOCs. XSIAM fused telemetry from endpoint, network, and identity layers into a centralized data fabric, and applied machine learning and behavioral analytics to drive fully automated detection and response actions. Unlike traditional SOAR tools, XSIAM emphasized autonomous correlation and proactive remediation, reducing mean time to respond (MTTR) by automating multi-step containment across distributed environments. These deployments illustrate the tangible progression of SOCs from static, manual operations to adaptive and intelligent systems capable of autonomous threat defense.

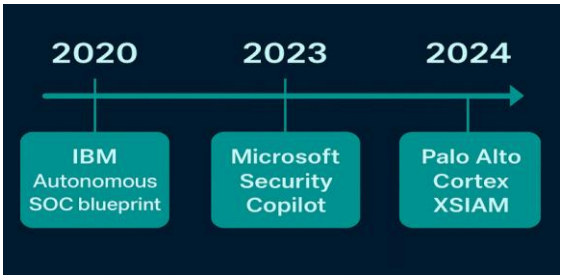

Fig 4: Timeline of Milestones toward Autonomous SOC

Together, these case studies demonstrate not only the technological feasibility of autonomous SOCs but also the diversity in architectural approaches from LLM-based assistants and modular automation to full-stack telemetry-driven intelligence. While adoption remains uneven across industries, these real-world implementations signal a broader industry shift toward AI-powered cybersecurity operations.

7. Future Outlook

As organizations continue to integrate AI deeper into their cybersecurity infrastructure, the vision of a fully autonomous, Level 5 SOC, a security environment requiring no human intervention appears increasingly plausible, though still aspirational. The journey from current semi-autonomous systems to true end-to-end autonomy will likely unfold over several phases, with improvements in AI reasoning, real-time orchestration, and agent collaboration playing central roles. However, achieving such autonomy involves more than technical refinement; it also raises critical questions about trust, accountability, and governance. One of the most pressing challenges on the horizon is ethical oversight. Autonomous agents making high-stakes decisions such as isolating a medical device or terminating user access must operate within strict organizational and legal boundaries. As AI agents assume greater control, the potential for unintended consequences, algorithmic bias, or over-enforcement becomes a significant concern. This creates a need for transparent decision-making frameworks that include mechanisms for auditing, redress, and rollback. Another core issue is explainability.

Many of today's most effective AI techniques, particularly those involving deep learning or large language models are notoriously difficult to interpret. In cybersecurity, where accountability and precision are paramount, stakeholders will increasingly demand human-understandable rationales for AI-driven actions. This will necessitate advances in XAI (eXplainable AI) methodologies and tighter integration of human-in-the-loop (HITL) models, where analysts oversee or validate decisions made by AI agents before critical actions are executed. Finally, as AI capabilities evolve, regulatory landscapes are expected to mature in parallel. Future SOC designs will need to comply with frameworks that address data privacy, automated decision-making accountability, and international threat attribution standards. Governments and industry bodies may begin to mandate certifications or operational transparency for AI-driven security systems, especially in sectors like finance, healthcare, and critical infrastructure. All in all, the road to fully autonomous SOCs will need to incorporate multidisciplinary developments not only technological but also ethical design, clarification of laws, and humane control. Such dimensions will become the measure of the next era of cybersecurity with AI agents not only acting as the tool but also as its responsible digital partner.

8. Conclusion

The security operations centers are undergoing the paradigm shift with the integration of the AI component into them and becoming a critical phase in cyber defense. SOC has raised the outstanding potential of autonomous SOCs, although they are still coming of age, to deliver dramatic increases in efficiency, resiliency, and scale in identifying and rectifying threats. The emergence of AI agents, in the form of LLMs-based assistants all the way up to autonomous remediation modules, has enabled them to support an ever-larger proportion of triaging and decision-making workloads, freeing human analysts to adapt to less procedural roles and responsibilities in managing the process. Nonetheless, this transition does not come with no restrictions. False positives, required structured inputs, and AI models susceptibility to adversarial evasion represent warnings not to implement AI recklessly and with insufficient regulation. Besides, the absence of transparency in most machine learning systems highlights the relevance of explainability and human responsibility especially when machine learning systems are deployed to a high-stakes setting or in a regulated setting. An example of the IBM, Microsoft, and Palo Alto Networks case study reveals the potential feasibility of AI agents in practice, and the results of such algorithms show an increase in the speed of response, the stability of work, and the automatization of tasks of the analyst.

However, these deployments also disclose the truth that present-day implementations are still not fully autonomous, usually depending upon human endorsements or narrowly predetermined decision-making rules. That puts into perspective the fact that autonomy is something desirable, but it should be sought gradually, with a solid basis in trust, transparency, and supervision. Looking ahead, the trajectory toward a Level 5 autonomous SOC will require advancements not only in AI technology, but also in system integration, policy design, and ethical governance. A fully autonomous SOC must not only be technically capable of acting independently but also be context-aware, policy-aligned, and auditable. Success in this domain will depend on developing hybrid models that integrate AI-driven speed and scale with human reasoning and adaptability. Ultimately, AI agents in the SOC represent more than just tools for automation; they are precursors to a broader shift in how cybersecurity operations are conceptualized and executed. With the right safeguards, explainability models, and human-in-the-loop design principles, autonomous SOCs have the potential to become trusted, intelligent defenders that not only respond to threats, but also learn, adapt, and protect proactively in real time.

References

[1] A. Basta, et al., Open-source Security Operations Center (SOC): A Complete Guide to Establishing, Managing, and Maintaining a Modern SOC, John Wiley & Sons, 2024.

[2] T. R. Kim III, Reducing Entropy Through Targeted Information Sharing: An Exploratory First Principles Approach to Closing the Gaps in Modern Security Operations Centers, Ph.D. dissertation, Marymount University, 2024.

[3] J. Kinyua and L. Awuah, "AI/ML in security orchestration, automation and response: Future research directions," Intell. Autom. Soft Comput., vol. 28, no. 2, 2021.

[4] R. A. Hammed and K. Sheriffdeen, "Revolutionizing SOC efficiency: Adaptive generative AI meets SOAR technologies," 2022.

[5] M. Saqib, S. Malhotra, D. Mehta, J. Jangid, F. Yashu and S. Dixit, "Optimizing spot instance reliability and security using cloud-native data and tools," J. Inf. Syst. Eng. Manage., vol. 10, no. 14s, pp. 720–731, 2025. [Online]. Available: https://doi.org/10.52783/jisem.v10i14s.2387

[6] S. Dixit and J. Jangid, "Asynchronous SCIM profile for security event tokens," J. Comput. Anal. Appl., vol. 33, no. 6, pp. 1357–1371, 2024. [Online]. Available: https://eudoxuspress.com/index.php/pub/article/view/1935

[7] V. S. Chakravarthi and S. R. Koteshwar, "IOT SOC architecture definition," in System on Chip (SOC) Architecture: A Practical Approach, Cham: Springer, 2023, pp. 91–104.

[8] Md. A. I. Mallick and R. Nath, "Navigating the cyber security landscape: A comprehensive review of cyber-attacks, emerging trends, and recent developments," World Sci. News, vol. 190, no. 1, pp. 1–69, 2024.

[9] T. Ajayi, et al., "An open-source framework for autonomous SoC design with analog block generation," in Proc. 2020 IFIP/IEEE 28th Int. Conf. Very Large Scale Integration (VLSI-SOC), IEEE, 2020.

[10] A. Warzyński, P. Bienias, and G. Kołaczek, "Application and evaluation of selected machine learning algorithms in anomaly detection module for SOC," in Dev. Artif. Intell. Technol. Comput. Robot.: Proc. 14th Int. FLINS Conf., 2020.

[11] S. Khaliq, Z. U. A. Tariq, and A. Masood, "Role of user and entity behavior analytics in detecting insider attacks," in Proc. 2020 Int. Conf. Cyber Warfare Security (ICCWS), IEEE, 2020.

[12] F. Ahmed, "Cloud security posture management (CSPM): Automating security policy enforcement in cloud environments," ESP Int. J. Adv. Comput. Technol. (ESP-IJACT), vol. 1, no. 3, pp. 157–166, 2023.

[13] A. W. Mir and R. K. Ramachandran, "Implementation of security orchestration, automation and response (SOAR) in smart grid-based SCADA systems," in Proc. 6th Int. Conf. Intell. Comput. Appl. (ICICA), Springer, 2021.

[14] F. Ahmed, "Cybersecurity policy frameworks for AI in government: Balancing national security and privacy concerns," Int. J. Multidiscip. Sci. Manage., vol. 1, no. 4, pp. 43–53, 2024.

[15] L. Mathur, P. P. Liang, and L.-P. Morency, "Advancing social intelligence in AI agents: Technical challenges and open questions," arXiv preprint, arXiv:2404.11023, 2024.

[16] O. Oniagbi, A. Hakkala, and I. Hasanov, Evaluation of LLM Agents for the SOC Tier 1 Analyst Triage Process, Master's thesis, Univ. Turku Dept. Comput., 2024. [Online]. Available: https://www.utupub.fi/bitstream/handle/10024/178601/Oniagbi%20Openime%20Thesis.pdf

[17] A. Yaseen, "Accelerating the SOC: Achieve greater efficiency with AI-driven automation," Int. J. Responsible Artif. Intell., vol. 12, no. 1, pp. 1–19, 2022.

[18] A. Basak, S. Bhunia, and S. Ray, "A flexible architecture for systematic implementation of SoC security policies," in Proc. 2015 IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD), IEEE, 2015.

[19] P. Fraccaro, et al., "Deploying an artificial intelligence application to detect flood from Sentinel 1 data," in Proc. AAAI Conf. Artif. Intell., vol. 36, no. 11, 2022.

[20] S. P. Veluru and M. K. Manchala, "Using LLMs as incident prevention copilots in cloud infrastructure," Int. J. AI, BigData, Comput. Manage. Stud., vol. 5, no. 4, pp. 51–60, 2024.