

From Copilots to Autonomous Agents: Rethinking Cybersecurity Operations in the Age of Agentic AI

¹ Mohamed Abourafia

DBA Researcher, eng.molay@gmail.com ¹

² Mr. Prasanna Chothwe

Cloud Engineer, ppchothwe@gmail.com ²

Abstract

Cybersecurity operations are moving from rule-based automation and analyst-led workflows toward AI-assisted and increasingly autonomous systems. AI copilots can support alert analysis, threat intelligence, investigation, and reporting, whereas agentic AI can add goal-directed reasoning, planning, tool use, coordination, and authorized action. This transition creates opportunities for more continuous and adaptive security operations, but it also introduces concerns involving decision authority, agent coordination, secure execution, accountability, and human oversight. This conceptual paper synthesizes recent literature on security operations centers (SOCs), generative AI, autonomous agents, autonomous cyber defense, multi-agent systems, and human–AI collaboration. Based on this synthesis, it develops an integrated framework organized around continuous threat sensing, agentic reasoning, multi-agent coordination, autonomous response, and human governance. The framework is represented as a continuous sense–detect–reason–plan–coordinate–act–monitor–learn cycle. Seven conceptual propositions connect agentic capabilities with operational responsiveness, detection effectiveness, investigation quality, coordination, response efficiency, governed autonomy, and cyber-defense resilience. The paper contributes a concise conceptual basis for understanding the shift from AI-assisted SOC work to coordinated, continuously monitored, and human-governed autonomous cyber defense.

Keywords: Agentic AI; Cybersecurity Operations; Security Operations Center; AI Copilots; Autonomous Agents; Cyber Defense; Human–AI Collaboration

1. Introduction

1.1 Background of the Study

Security Operations Centers (SOCs) provide centralized capabilities for monitoring, detecting, investigating, and responding to cyber incidents. Traditional SOCs depend on rules, signatures, analyst judgment, and multiple security tools. These arrangements can create alert overload, fragmented evidence, repetitive triage, and delayed response. LLM-enabled systems increasingly support SOC functions by summarizing alerts, correlating evidence, and assisting analysts (Habibzadeh et al., 2026; Srinivas et al., 2025).

Agentic AI introduces a further shift. Autonomous agents can perceive information, reason over multi-step tasks, plan actions, use tools, and pursue defined goals with limited intervention (Wang et al., 2024). In cybersecurity, this creates the possibility of moving from assistance toward delegated investigation and response. However, autonomous execution also introduces risks involving incorrect reasoning, excessive permissions, prompt injection, tool

compromise, and weak accountability (Deng et al., 2025; Ferrag et al., 2026).

Recent autonomous-defense research therefore emphasizes reliable operation rather than technical capability alone. Vyas et al. (2025) highlight the need to consider realistic deployment conditions, while Kshetri (2025) describes agentic AI as a potential mechanism for more adaptive cyber defense. The key issue is how autonomy can be integrated into SOC workflows while maintaining authorization, traceability, and human control. The operational difference between a copilot and an agent is primarily one of delegation and control. A copilot normally responds to an analyst request and presents information or recommendations. An agent can initiate a task, sequence actions, interact with tools, and continue toward an objective. This distinction makes governance, risk-sensitive authorization, and continuous monitoring central to agentic cybersecurity operations.

1.2 Problem Statement

Conventional SOC processes remain constrained by alert overload, fragmented

tools, analyst workload, and dependence on predefined workflows. Copilots reduce some cognitive burden, but they generally remain advisory and require analysts to interpret outputs and execute consequential actions. A research challenge therefore remains: how can cybersecurity operations move from isolated AI assistance toward coordinated, semi-autonomous or autonomous workflows without creating unacceptable operational and governance risks? These gaps support the development of an integrated **Agentic Cybersecurity Operations Framework** combining autonomous capabilities with multi-agent coordination and human governance.

1.3 Research Aim

The study aims to develop a conceptual framework explaining how agentic AI can transform cybersecurity operations from AI-assisted support toward coordinated, governed, and increasingly autonomous cyber defense.

1.4 Research Objectives

RO1: Examine the evolution from AI copilots to agentic AI in cybersecurity operations.

RO2: Identify the key capabilities required for effective agentic cybersecurity operations.

RO3: Examine multi-agent coordination across threat detection, investigation, and response.

RO4: Develop a conceptual framework integrating agentic capabilities with cybersecurity operations.

RO5: Examine the role of human oversight and governance in autonomous cyber defense.

1.5 Research Questions

RQ1: How does agentic AI change cybersecurity operations compared with AI copilot-based support?

RQ2: What key capabilities are required for effective agentic cybersecurity operations?

RQ3: How can specialized agents coordinate across threat detection, investigation, and response?

RQ4: How should human oversight and governance operate within autonomous cybersecurity operations?

RQ5: How can agentic AI capabilities be integrated into a coherent cybersecurity operations framework?

1.6 Scope and Significance of the study

The paper concerns defensive operations in organizational SOCs, including triage, investigation, containment, and recovery. Agentic AI is defined here as a goal-directed system that can select and sequence tasks, interact with authorized tools, and continue within an assigned scope. Autonomy is treated as a task-specific allocation of decision rights. A SOC may delegate evidence enrichment extensively while requiring approval for changes affecting production systems. Consequently, an organization need not adopt the same autonomy level across all functions. This framing matters because adoption decisions often emphasize model capability while leaving incident ownership and execution rights implicit. The proposed

framework offers researchers a set of testable relationships and gives managers a basis for aligning agent roles with existing security processes. It does not assume that greater autonomy always produces better security outcomes or that a language model provides a dependable estimate of its own uncertainty.

The study is significant because it connects agentic AI with established organizational perspectives and provides a structured basis for understanding controlled autonomy, coordination, governance, and future empirical testing.

2. Literature Review

2.1 Cybersecurity Operations and SOCs

SOCs coordinate continuous monitoring, alert triage, threat analysis, incident response, and security reporting. Alert fatigue and fragmented tooling can reduce analyst effectiveness (Chhetri et al., 2024). Recent studies show that LLM-based systems can support multiple SOC activities, including triage, investigation, threat intelligence, and reporting (Habibzadeh et al., 2026; Srinivas et al., 2025). LLM-enabled systems are

increasingly positioned across several SOC functions rather than a single detection task. Habibzadeh et al. (2026) map LLM applications to NIST CSF functions and use MITRE ATT&CK to examine detection capabilities, while Srinivas et al. (2025) review applications spanning triage, threat intelligence, incident response, reporting, asset discovery, and vulnerability management. These findings support an integrated rather than isolated view of AI-enabled SOC operations. SOC effectiveness depends on the quality and timing of information moving between telemetry sources, analysts, detection systems, and orchestration platforms. This makes cybersecurity operations a useful setting for agentic AI because agents can potentially connect sensing, reasoning, coordination, and action within a common workflow.

2.2 AI Copilots and Generative AI

AI copilots primarily augment analysts. They can summarize telemetry, translate technical evidence, generate queries, retrieve threat context, and support incident analysis (Srinivas et al., 2025). Human–AI teaming research

indicates that such systems can reduce repetitive cognitive work while retaining analyst judgment (Chhetri et al., 2024; Cohen et al., 2025).

Copilot outputs do not remove the need for verification. Generative systems can produce incorrect explanations, follow malicious instructions, or expose sensitive information. Li and Fung (2025) and Ferrag et al. (2026) therefore emphasize security controls around LLMs and agents. Copilots are consequently best understood as an intermediate stage between manual operations and greater delegation. Copilots are particularly useful where the analyst must interpret heterogeneous evidence rather than execute a fixed rule. They can summarize commands and logs, explain unfamiliar indicators, draft queries, compare threat information, and produce incident notes.

The value is therefore partly cognitive: the system compresses information and helps the analyst move from raw telemetry to a usable security narrative. Singh et al. (2026) provide empirical evidence that analysts use LLMs as on-demand aids for sensemaking and context-

building while preserving analyst authority over high-stakes decisions.

2.3 Agentic AI and Autonomous Cyber Defense

Agentic AI systems combine perception, reasoning, planning, memory or context, tool interaction, and action execution. Wang et al. (2024) describe autonomous agents as systems capable of pursuing goals through multiple steps, while cybersecurity research applies these capabilities to threat analysis and defense (Kshetri, 2025; Vyas et al., 2025). Agentic systems also introduce a larger security attack surface. Risks include prompt injection, unsafe tool use, compromised agents, excessive

permissions, and cascading errors (Deng et al., 2025; Ferrag et al., 2026; Tang et al., 2026).

Autonomous cyber defense must therefore combine capability with constrained execution and monitoring. Autonomous cyber defense research provides an important distinction between technical automation and goal-directed agency. An autonomous defense agent is expected to perceive an environment, choose among actions, and adapt to changing conditions rather than simply trigger a fixed playbook. Vyas et al. (2025) show that realistic autonomous defense requires consideration of agent capabilities, environments, learning, explainability, and deployment constraints.

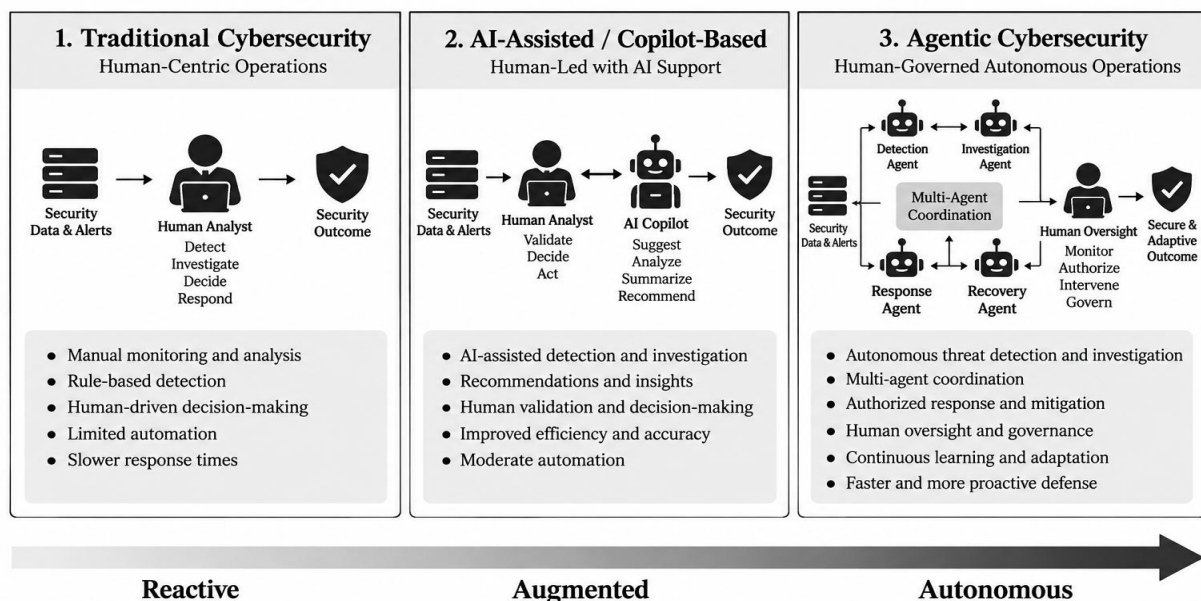


Figure 1. Evolution from Cybersecurity Copilots to Autonomous Cybersecurity Agents

2.4 Multi-Agent Coordination, Human–AI Collaboration and Governance

Multi-agent architectures distribute work among specialized agents such as detection, investigation, threat intelligence, response, and recovery agents. Security orchestration research suggests that agentic systems can support more integrated automation across SOC workflows (Ismail et al., 2025). Specialization may improve task focus, but it also creates dependencies that require explicit coordination and shared context. Governance is a cross-cutting property rather than a final approval step. NIST AI RMF 1.0 places governance across the AI lifecycle, while NIST CSF 2.0 provides a broader cybersecurity risk-management structure (NIST, 2023; Pascoe et al., 2024). For agentic SOCs, governance should define permissions, escalation thresholds, logging, accountability, and intervention points.

Multi-agent coordination can improve specialization, but specialization also creates dependencies. A detection agent may identify a suspicious event without understanding

business impact; an investigation agent may require threat intelligence; and a response agent may need explicit authorization before changing a live system. Coordination therefore requires shared context, task ownership, communication protocols, and clear conditions for handing work from one agent to another.

2.5 Research Gap

The literature review identifies five key gaps in current research on AI-enabled cybersecurity operations:

- 1. Copilot-to-Agent Gap:** Existing research focuses strongly on AI-assisted cybersecurity, while the transition from copilots to autonomous agents remains underdeveloped.
- 2. Operational Integration Gap:** Detection, investigation, response, and recovery are often examined separately rather than as a continuous agentic workflow.
- 3. Multi-Agent Coordination Gap:** Limited attention has been given to coordination among specialized

cybersecurity agents performing interconnected security tasks.

4. Human Oversight Gap: Clear mechanisms for **authorization, accountability, escalation, and human intervention** in autonomous operations remain insufficiently defined.

5. Continuous Cyber Defense Gap: Existing approaches provide limited integration of **continuous sensing, reasoning, action, monitoring, and learning** within a unified cybersecurity model.

These gaps support the development of an integrated **Agentic Cybersecurity Operations Framework** combining autonomous capabilities with multi-agent coordination and human governance.

3. Theoretical Foundation

3.1 Sociotechnical Systems Perspective

The sociotechnical perspective treats performance as an outcome of interactions among technical and social components.

Cherns (1976) argues that effective systems require joint consideration of technology, work design, and human roles. This is relevant because agentic AI changes how SOC tasks are allocated between analysts and automated systems.

This perspective is particularly relevant to the copilot-to-agent transition because autonomy changes work allocation. Analysts may spend less time on repetitive investigation and more time on validation, exception handling, threat hunting, and governance. The resulting system should therefore be evaluated at the level of the combined human–AI workflow rather than at the level of the model alone. The sociotechnical perspective also helps explain why technical performance alone is insufficient. An agent may achieve high detection accuracy but still produce poor operational outcomes if analysts do not understand its recommendations, if responsibilities are ambiguous, or if the response workflow does not match organizational procedures. The framework therefore treats human roles, technology,

workflow, and governance as mutually dependent components.

3.2 Dynamic Capabilities Perspective

Dynamic capabilities explain how organizations sense changes, seize opportunities, and reconfigure resources in changing environments (Teece et al., 1997; Teece, 2007). Agentic cybersecurity can be viewed through this lens because continuous sensing, rapid response, and learning may strengthen organizational adaptation to evolving threats.

Within the proposed framework, continuous sensing corresponds to environmental awareness, agentic planning supports rapid response, and continuous learning supports reconfiguration. The perspective therefore provides a theoretical rationale for treating agentic cybersecurity as an adaptive organizational capability rather than a one-time technology implementation. Teece et al. (1997) and Teece (2007) distinguish dynamic capabilities from static resources by emphasizing the ability to sense changes, seize opportunities, and reconfigure resources.

Cybersecurity operations closely reflect this logic because attackers and defensive conditions change continuously. A system that performs well only under previously observed conditions may become ineffective as tactics, assets, vulnerabilities, or organizational priorities change.

3.3 Human–AI Collaboration Perspective

Human–AI collaboration focuses on combining complementary strengths rather than treating AI as a replacement for human expertise. Hybrid-intelligence research emphasizes complementary human and artificial capabilities (Dellermann et al., 2019), while SOC evidence shows that analysts continue to rely on AI for sensemaking and task support (Singh et al., 2026). Recent SOC evidence suggests that collaboration should remain context-sensitive. Singh et al. (2026) report that analysts use LLMs heavily for sensemaking while retaining authority for high-stakes decisions. This supports a framework in which autonomy can vary by task and risk level instead of being treated as a binary property. Human–AI collaboration also

requires explicit role boundaries. Dellermann et al. (2019) describe hybrid intelligence as the combination of human and artificial capabilities to achieve outcomes that neither side can reliably produce alone. In cybersecurity, the complementary pattern is visible when agents handle scale and repetitive analysis while analysts provide contextual knowledge, risk interpretation, and responsibility for consequential decisions.

3.4 Autonomous Decision-Making and Integrated Perspective

Autonomous decision-making concerns the ability of systems to evaluate situations, select actions, and execute tasks with limited intervention. In cybersecurity, autonomy should be treated as a continuum: low-risk tasks may be automated, medium-risk actions may require confirmation, and high-risk actions should remain under stronger human control. The integrated perspective therefore links autonomy with organizational adaptation and human–technology design. The integrated theoretical perspective links this continuum to organizational adaptation and human–

technology design. It explains why the proposed framework includes both autonomous capabilities and governance controls: technical autonomy provides operational capacity, while sociotechnical and dynamic-capability perspectives determine how that capacity is embedded, adapted, and supervised. Autonomous decision-making should therefore be understood as a continuum. Low-risk tasks may be fully automated, medium-risk tasks may require confirmation, and high-impact actions may require direct human authorization. This risk-based interpretation is consistent with the governance emphasis in NIST AI RMF and NIST CSF 2.0 and provides a practical basis for assigning different autonomy levels to different SOC activities.

4. Research Methodology

4.1 Research Design

This study adopts a conceptual integrative literature review to synthesize research on agentic AI and cybersecurity operations. The corpus contains 21 entries spanning

foundational organizational theory, SOC and agent research, security surveys, governance frameworks, and a technical threat knowledge resource. No primary data were collected. The study integrates technical, organizational, and governance perspectives to develop a conceptual framework and propositions. The framework and propositions are analytical outputs, and their performance implications remain untested.

4.2 Literature Search Strategy

The literature search focused on **SOC operations, AI copilots, generative AI, LLM-based agents, autonomous cyber defense, multi-agent systems, human–AI collaboration, and AI governance**. Recent research was prioritized for emerging agentic-AI applications, while foundational studies were retained for established theoretical perspectives.

4.3 Inclusion and Exclusion Criteria

Sources were included when they directly addressed AI-enabled cybersecurity

operations, autonomous or agentic systems, SOCs, multi-agent security, human–AI collaboration, or cybersecurity governance. Duplicates and sources without substantive relevance to these areas were excluded.

4.4 Thematic Synthesis and Framework Development

The literature was synthesized around five recurring themes: threat sensing and detection, agentic reasoning, multi-agent coordination, autonomous response, and human governance and learning. These themes were mapped to operational capabilities and expected outcomes, forming the basis of the proposed framework and seven conceptual propositions. The framework therefore identifies data quality, tool integrity, task complexity, asset criticality, and analyst expertise as boundary conditions. Transparent reporting also requires distinguishing recommendations developed in this paper from findings established in the supplied studies. No operational benefit is reported as a demonstrated result.

Table 1. Comparison of Traditional, AI-Assisted, and Agentic Cybersecurity Operations

Dimension	Traditional Cybersecurity	AI-Assisted / Copilot-Based	Agentic Cybersecurity
Primary role	Human-led operations	AI-supported analyst	AI agents with human oversight
Threat detection	Rules and analyst monitoring	AI-assisted detection	Continuous agentic sensing
Investigation	Manual analysis	AI-supported analysis	Agentic reasoning and investigation
Decision-making	Primarily human	Human-led with recommendations	Controlled autonomous decisions
Response	Manual or predefined automation	AI-recommended actions	Authorized autonomous execution
Coordination	Analyst and tool coordination	Human-directed integration	Multi-agent and tool coordination
Human role	Direct execution	Supervision and validation	Oversight, authorization, escalation
Adaptability	Limited	Moderate	Continuous monitoring and adaptation
Governance	Conventional controls	AI validation	Authorization, monitoring, accountability

The literature also suggests that future frameworks should distinguish capability from authorization. An agent may possess the technical ability to execute an action without being permitted to do so in a particular context. This distinction is central to responsible autonomous cyber defense and is incorporated into the proposed governance layer. The gaps are interconnected. The copilot-to-agent gap creates the need to define delegated action; the operational gap requires an end-to-end workflow; coordination requires shared context and task ownership; governance defines acceptable autonomy; and continuous defense requires controlled learning. The proposed framework addresses these gaps as parts of one operational system.

5. Proposed Agentic Cybersecurity Operations Framework

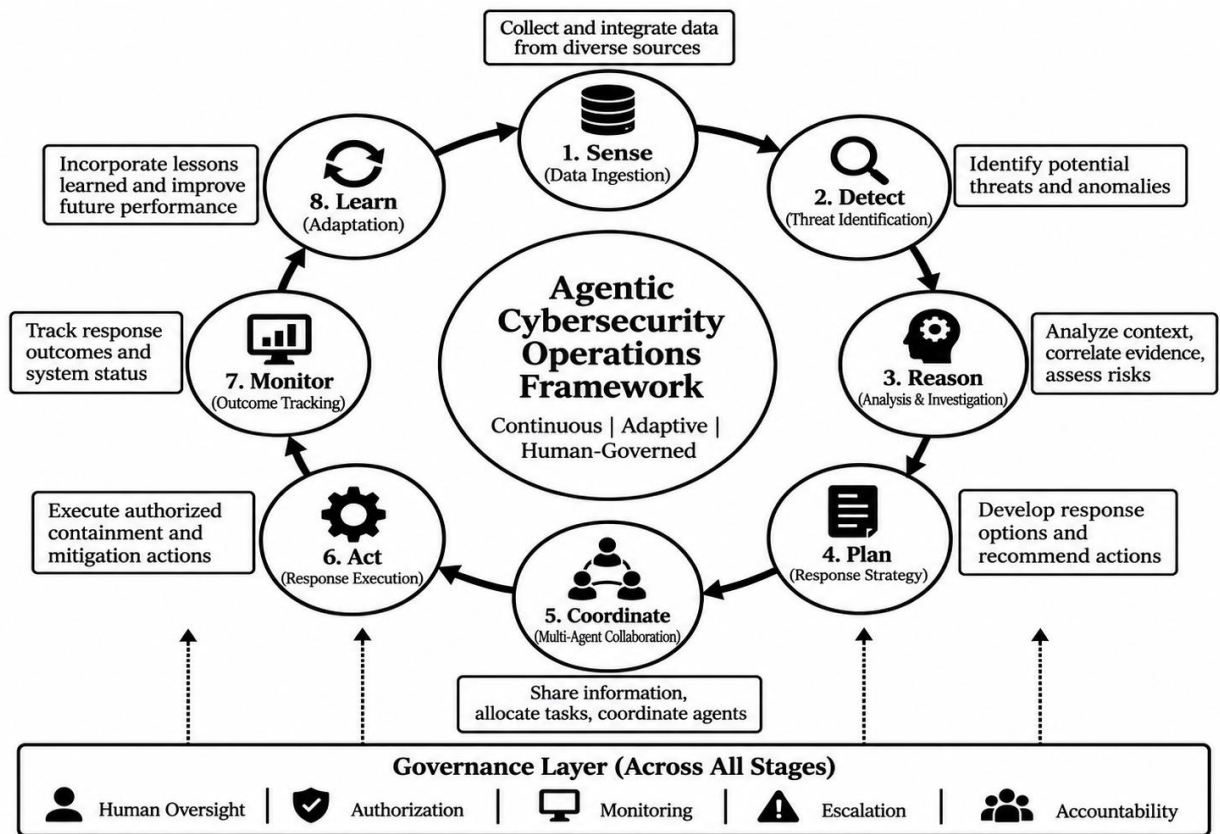


Figure 2. Agentic Cybersecurity Operations Decision Loop

Figure 2 integrates agentic capabilities within a governed incident cycle. The sequence comprises sensing, detection, reasoning, planning, coordination, execution, monitoring, and learning, grouped into four operational stages for clarity. The cycle can pause, return to evidence collection, or escalate to an analyst. Continuous operation refers to monitored recurrence of these activities, not uninterrupted autonomous execution. Shared evidence and controlled tool services connect

the stages, while governance defines permissible action throughout.

5.1 Continuous Threat Sensing and Contextual Awareness

The first capability is continuous sensing of security telemetry and contextual information. Agents can ingest logs, alerts, endpoint and network signals, vulnerabilities, and threat intelligence. The purpose is to maintain a current security state rather than evaluate isolated alerts. Operationally, the sensing layer can feed normalized events to downstream

agents and maintain a continuously updated security state. This can reduce repeated manual collection and provide a common information base for investigation and response agents. The quality of downstream reasoning remains dependent on data quality, provenance, and access controls. Sensing should combine technical telemetry with contextual information. Security events become more meaningful when correlated with asset criticality, vulnerability status, user or service identity, threat intelligence, and business impact. The proposed framework therefore treats contextual awareness as part of sensing rather than as a later analyst-only activity.

5.2 Agentic Reasoning and Investigation

Agentic reasoning moves beyond single-alert classification toward evidence correlation, investigation planning, and threat assessment. Agents can combine internal and external context, identify relationships across events, and construct investigation paths. Because reasoning can influence later actions, traceability is important. Because reasoning may influence subsequent actions, traceability

is important. Investigation agents should record the evidence used, tools invoked, assumptions made, and confidence or uncertainty associated with the result. Such records support human review and make it easier to detect when an agent has been misled by incomplete or manipulated information. Reasoning is the point at which an agent moves from recognizing a signal to constructing an explanation of what may be happening. The agent can correlate related events, enrich indicators, compare evidence with known threat patterns, and identify the next investigative action. MITRE ATT&CK can provide a common vocabulary for relating observed activity to adversary tactics and techniques, while NIST CSF can provide a broader risk-management context.

5.3 Multi-Agent Coordination and Autonomous Response

Specialized agents can divide cybersecurity responsibilities across detection, investigation, threat intelligence, response, and recovery. Coordination allows tasks to be passed between agents while maintaining shared

context. Ismail et al. (2025) support the use of but autonomous execution must remain agentic approaches for security orchestration, bounded by authorization and risk controls.

Table 2. Agent Types and Cybersecurity Responsibilities

Agent Type	Primary Responsibility	Key Activities
Detection Agent	Threat identification	Monitor alerts; detect anomalies; prioritize suspicious activity
Investigation Agent	Threat analysis	Correlate events; analyze evidence; investigate incidents
Threat Intelligence Agent	Threat context	Enrich indicators; retrieve intelligence; provide context
Response Agent	Threat mitigation	Execute authorized containment and mitigation actions
Recovery Agent	System recovery	Support restoration; assess outcomes; identify adaptation needs

Autonomous responses should be risk-sensitive. Reversible actions with limited impact may be delegated more broadly, whereas actions affecting critical systems, identities, or production availability may require human approval. This principle turns autonomy into a controlled operating policy rather than a blanket permission to act. Coordination can be organized around explicit task ownership. A detection agent can hand a prioritized event to an investigation agent; the investigation agent can request threat context; the response agent can prepare a mitigation

plan; and the recovery agent can verify restoration and capture lessons. This division reduces the need for one general-purpose agent to perform every task and makes responsibilities easier to monitor.

5.4 Human Governance, Continuous Learning and Framework Integration

Human governance forms the control layer across all stages. High-risk actions require explicit authorization or escalation, while lower-risk actions may be delegated within predefined limits. Governance should include access controls, action thresholds, audit logs,

rollback mechanisms, and clear accountability. feedback rather than uncontrolled self-Continuous learning should use validated modification.

Table 3. Agentic Cybersecurity Capabilities and Expected Operational Outcomes

Agentic Capability	Operational Function	Expected Outcome
Continuous threat sensing	Monitor security signals and context	Faster and broader threat identification
Agentic reasoning	Analyze events and investigate threats	Improved investigation quality
Multi-agent coordination	Share information and coordinate agents	Integrated security operations
Autonomous response	Execute authorized containment and mitigation	Faster incident response
Human governance	Provide oversight, authorization, escalation, accountability	Controlled and responsible autonomy
Continuous learning	Learn from outcomes and adapt strategies	Improved cyber-defense adaptability

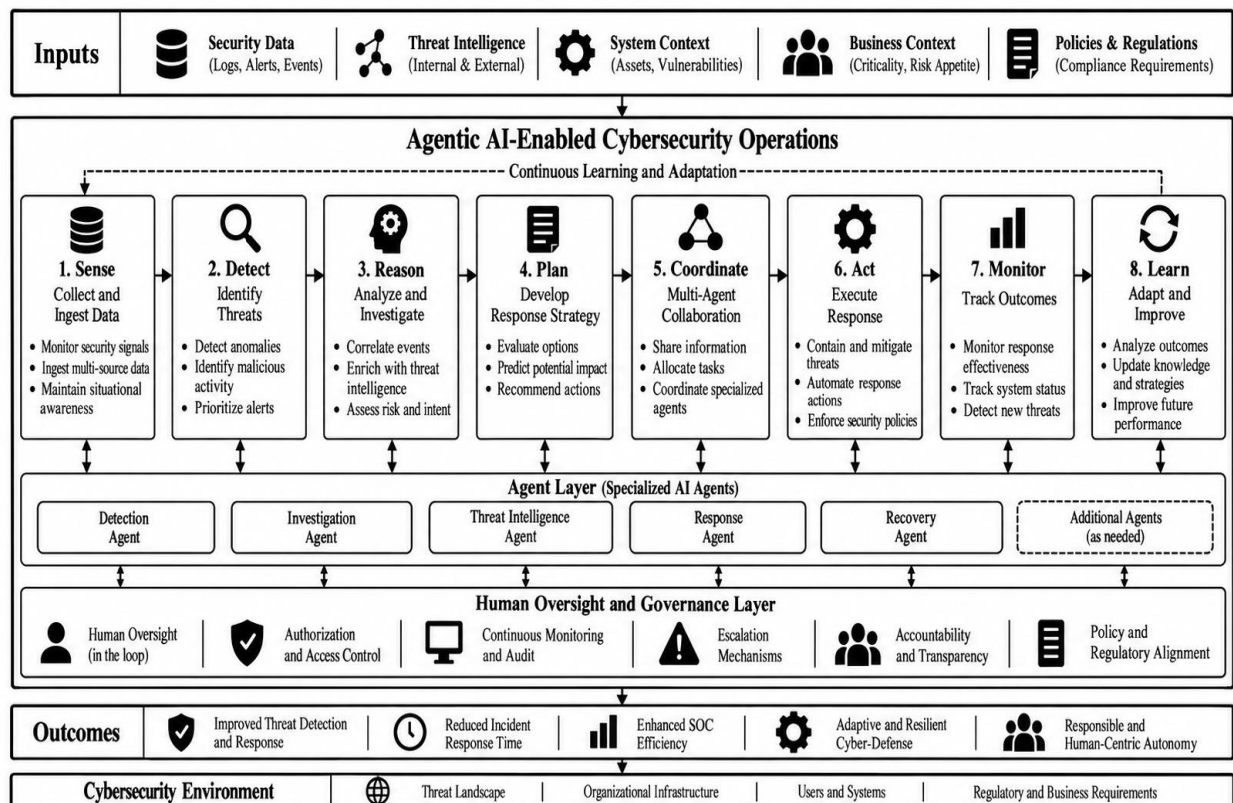


Figure 3. Conceptual Framework for Agentic AI-Enabled Cybersecurity Operations

Continuous learning should also be distinguished from uncontrolled self-modification. The framework assumes that learning is governed through validated feedback, controlled updates, and monitoring. This distinction is important because agentic systems can interact with changing environments and untrusted inputs; learning processes themselves therefore require security and change-management controls. Governance mechanisms should operate before, during, and after an agent action. Before execution, access rights and thresholds define what is permitted. During execution, monitoring and intervention provide control. After execution, logging, evaluation, and feedback support accountability and learning. This creates a closed operational loop from sensing to improvement.

6. Conceptual Propositions

The proposed framework establishes relationships between agentic cybersecurity capabilities, governance mechanisms, and operational outcomes. Based on the identified

literature themes and theoretical perspectives, seven propositions are developed:

P1: Stronger agentic AI capabilities are positively associated with cybersecurity operational responsiveness.

P2: Stronger continuous threat-sensing capabilities are positively associated with cybersecurity threat-detection effectiveness.

P3: Stronger agentic reasoning and investigation capabilities are positively associated with cybersecurity investigation quality and response effectiveness.

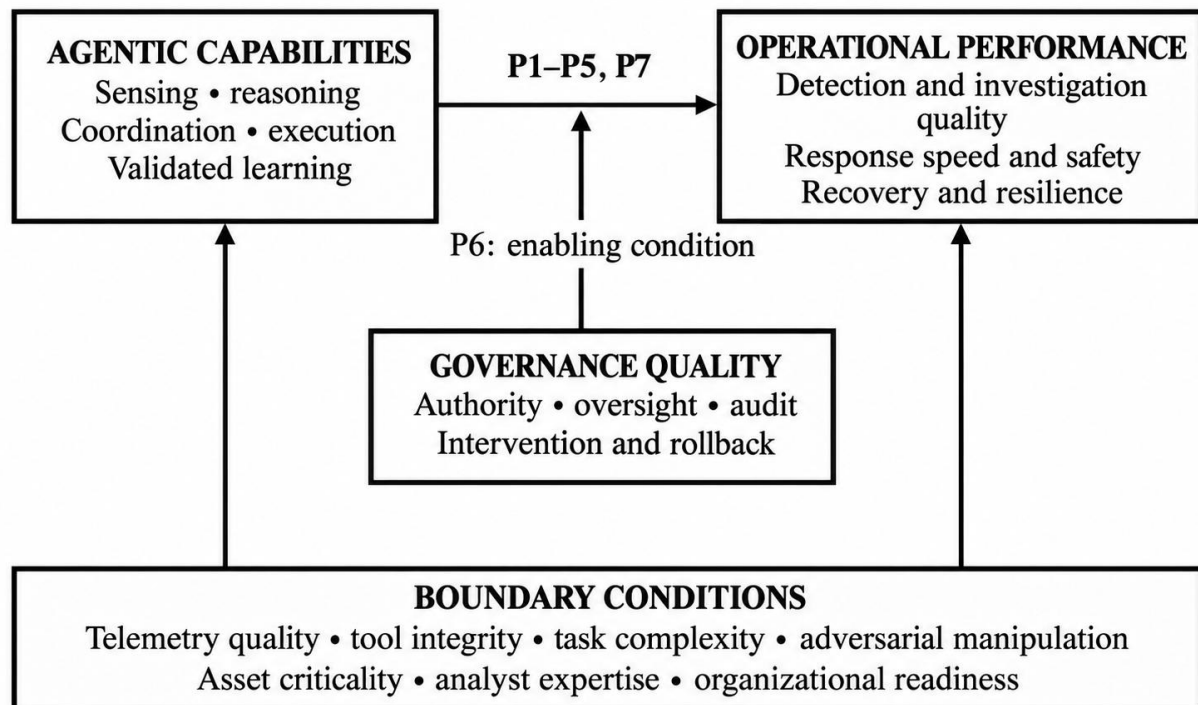
P4: Effective multi-agent coordination is positively associated with integrated and coordinated cybersecurity operations.

P5: Higher levels of controlled autonomous execution are positively associated with cybersecurity response efficiency.

P6: Effective human oversight and governance are positively associated with the reliable and accountable deployment of autonomous cybersecurity agents.

P7: Integrated agentic cybersecurity continuous cyber-defense performance and capabilities are positively associated with organizational resilience.

Figure 4. Conceptual relationships propositions and boundary conditions



Conceptual relationships and proposed boundary conditions; no empirical estimates.

7. Discussion

7.1 Theoretical Contributions

The study contributes to the emerging literature on agentic AI and cybersecurity operations by conceptualizing the transition from AI copilots to autonomous agents as a change in work allocation, coordination, decision authority, and governance, rather than simply an increase in automation. The

framework integrates sociotechnical systems, dynamic capabilities, human–AI collaboration, and autonomous decision-making perspectives to explain how agentic capabilities can be embedded within cybersecurity operations. The novelty of the study lies in connecting the copilot-to-agent transition with an end-to-end cybersecurity operations model. Rather than treating agentic AI as an isolated technological

capability, the framework integrates continuous sensing, agentic reasoning, multi-agent coordination, controlled autonomous response, human governance, and continuous learning within a single operational cycle. This extends existing AI-enabled SOC research by conceptualizing autonomy as a governed organizational capability rather than simply increased automation. The framework also provides a theoretical bridge between technical agent capabilities and organizational outcomes. Sociotechnical theory explains the changing relationship between analysts and AI systems, dynamic capabilities explain continuous adaptation, human–AI collaboration explains complementary role allocation, and autonomous decision-making explains controlled delegation. Together, these perspectives provide a basis for understanding how agentic AI can support adaptive and governed cybersecurity operations.

7.2 Managerial and SOC Implications

For SOC managers, agentic AI should be introduced through clearly bounded tasks, explicit permissions, and measurable

outcomes. Agents can support enrichment, triage, investigation, and low-risk response, while high-impact actions should remain subject to appropriate authorization and escalation. Key measures include response time, investigation quality, escalation frequency, and unauthorized-action rates.

7.3 Practical Contributions

The framework provides a compact operating model for integrating agentic AI with existing security tools and SOC processes. It emphasizes continuous monitoring, evidence-based reasoning, coordinated action, risk-sensitive autonomy, and human governance. It can also serve as a design checklist for mapping current workflows to agent capabilities. The model also supports incremental adoption. An organization does not need to move directly from a copilot to full autonomy. Intermediate levels can combine AI recommendations, human confirmation, limited delegated actions, and continuous monitoring before broader authority is granted. The framework can be used as a design checklist for agentic SOC initiatives. Teams

can map existing tools and workflows to the eight operational stages, identify where agents can add value, and then assign governance controls to each autonomous action. This creates a practical bridge between existing SOAR infrastructure and more flexible agentic workflows.

7.4 Research Contributions

The seven propositions provide a basis for future empirical testing. Future studies can compare human-only, copilot-assisted, and agentic workflows using measures such as response time, investigation quality, analyst workload, coordination effectiveness, and error rates.

8. Limitations and Future Research

This study is conceptual and does not empirically validate the proposed framework or propositions. Future research should test the model through SOC case studies, controlled experiments, simulations, and longitudinal studies. Key areas include prompt injection, compromised tools, reasoning errors, agent

coordination failures, and human intervention.

The framework also abstracts from vendor- and sector-specific conditions. Future research can therefore examine different autonomy levels and governance requirements across sectors such as critical infrastructure, healthcare, finance, and cloud environments.

9. Conclusion

Agentic AI is extending cybersecurity operations from AI-assisted analysis toward systems capable of reasoning, planning, coordination, and authorized action. This study proposes a conceptual framework integrating continuous sensing, agentic reasoning, multi-agent coordination, risk-sensitive autonomy, human governance, and continuous learning. The framework and seven propositions provide a basis for future empirical research. The central contribution is that effective agentic cybersecurity depends not only on autonomy, but also on authorization, accountability, human oversight, and continuous monitoring.

References

1. Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509–533. [https://doi.org/10.1002/\(SICI\)1097-0266\(199708\)18:7<509::AID-SMJ882>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z)
2. Chhetri, M. B., Tariq, S., Singh, R., Jalalvand, F., Paris, C., & Nepal, S. (2024). Towards human-AI teaming to mitigate alert fatigue in Security Operations Centres. *ACM Transactions on Internet Technology*, 24(3), Article 12. <https://doi.org/10.1145/3670009>
3. Cherns, A. (1976). The principles of sociotechnical design. *Human Relations*, 29(8), 783–792. <https://doi.org/10.1177/001872677602900806>
4. Cohen, D., Te'eni, D., Yahav, I., Zagalsky, A., Schwartz, D., Silverman, G., Mann, Y., Elalouf, A., et al. (2025). Human–AI enhancement of cyber threat intelligence. *International Journal of Information Security*, 24, 99. <https://doi.org/10.1007/s10207-025-01004-4>
5. Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering*, 61(5), 637–643. <https://doi.org/10.1007/s12599-019-00595-2>
6. Deng, Z., Guo, Y., Han, C., Ma, W., Xiong, J., Wen, S., & Xiang, Y. (2025). AI agents under threat: A survey of key security challenges and future pathways. *ACM Computing Surveys*, 57(7), Article 182. <https://doi.org/10.1145/3716628>
7. Ferrag, M. A., Lakas, A., Tihanyi, N., & Debbah, M. (2026). Securing LLM agents: From prompt sanitization to autonomous red teaming and beyond. *Internet of Things and Cyber-Physical Systems*, 5, 185–209. <https://doi.org/10.1016/j.iotcps.2026.03.001>
8. Habibzadeh, A., Feyzi, F., & Ebrahimi Atani, R. (2026). Large language models for Security Operations Centers: A comprehensive survey. *Journal of Electrical and Computer Engineering*, 2026, Article 3383674. <https://doi.org/10.1155/jece/3383674>

9. Ismail, Kurnia, R., Brata, Z. A., Nelistiani, G. A., Heo, S., Kim, H., & Kim, H. (2025). Toward robust security orchestration and automated response in Security Operations Centers with a hyper-automation approach using agentic artificial intelligence. *Information*, 16(5), 365. <https://doi.org/10.3390/info16050365>
10. Kim, J., Liu, X., Wang, Z., Qiu, S., Li, B., Guo, W., & Song, D. (2026). The attack and defense landscape of agentic AI: A comprehensive survey. *arXiv preprint arXiv:2603.11088*. <https://doi.org/10.48550/arXiv.2603.11088>
11. Kshetri, N. (2025). Transforming cybersecurity with agentic AI to combat emerging cyber threats. *Telecommunications Policy*, 49(6), 102976. <https://doi.org/10.1016/j.telpol.2025.102976>
12. Li, M. Q., & Fung, B. C. M. (2025). Security concerns for large language models: A survey. *Journal of Information Security and Applications*, 95, 104284. <https://doi.org/10.1016/j.jisa.2025.104284>
13. MITRE. (2026). MITRE ATT&CK Enterprise Matrix. MITRE ATTaCK. <https://attack.mitre.org/matrices/enterprise/>
14. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
15. Pascoe, C., Quinn, S., & Scarfone, K. (2024). The NIST Cybersecurity Framework (CSF) 2.0 (NIST CSWP 29). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.CSWP.29>
16. Singh, R., Tariq, S., Jalalvand, F., Chhetri, M. B., Nepal, S., Paris, C., & Lochner, M. (2026). LLMs in the SOC: An empirical study of human-AI collaboration in Security Operations Centres. *Proceedings of the IEEE Symposium on Security and Privacy*, 4262–4281. <https://doi.org/10.1109/SP63933.2026.00111>

17. Srinivas, S., Kirk, B., Zendejas, J., Espino, M., Boskovich, M., Bari, A., Dajani, K., & Alzahrani, N. (2025). AI-augmented SOC: A survey of LLMs and agents for security automation. *Journal of Cybersecurity and Privacy*, 5(4), 95. <https://doi.org/10.3390/jcp5040095>
18. Tang, Y., Liu, Y., Lan, J., Yan, Z., & Gelenbe, E. (2026). Security of LLM-based agents regarding attacks, defenses, and applications: A comprehensive survey. *Information Fusion*, 127, 103941. <https://doi.org/10.1016/j.inffus.2025.103941>
19. Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350. <https://doi.org/10.1002/smj.640>
20. Vyas, S., Mavroudis, V., & Burnap, P. (2025). Towards the deployment of realistic autonomous cyber network defence: A systematic review. *ACM Computing Surveys*, 58(1), Article 5. <https://doi.org/10.1145/3729213>
21. Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345. <https://doi.org/10.1007/s11704-024-40231-1>