Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

From Ephemeral Chats to Enduring Capabilities: Feedback Loops in SOC AI Implementation

Neal Anand Iyer
Independent Researcher, USA

ARTICLE INFO

Received: 05 Nov 2025

Revised: 20 Dec 2025

Accepted: 29 Dec 2025

ABSTRACT

AI assistants are increasingly being used in Security Operations Centers (SOCs) as the means of triage and investigation of alerts; however, the insights obtained in these engagements are frequently lost once the case is closed. The article introduces an architecture that will change transient AI interactions into long-term organizational capabilities based on four formal feedback loops: Detection Refinement, Playbook Synthesis, Tooling Gap Identification, and Knowledge Curation. The loops are systematic mechanisms of learning patterns of AI-assisted investigations and transforming them into better detection rules, standardized playbooks, infrastructure enhancements, and institutional knowledge. The architecture is based on OpenTelemetry instrumentation and consistent with models such as the NIST AI Risk Management Framework and Cybersecurity Framework, and includes the relevant governance controls and risk mitigations to guarantee security, privacy, and compliance with regulations. These systematic feedback mechanisms would enable organizations to ensure every AI-assisted investigation is added to long-lasting security capabilities in such a way that a continuous improvement cycle is established, building coverage of detection, productivity of the analysts, and operational knowledge with proper human oversight and governance controls.

Keywords: Security Operations, Artificial Intelligence, Feedback Loops, Knowledge Management, Detection Engineering

1. Introduction

The use of AI assistants and agents in security Operations Centers (SOC) in the process of alert triage and investigation is becoming more widespread, transforming the manner in which analysts are exposed to security information and conduct investigations. These AI functions accelerate the process of evidence gathering, enhance the correlation of threats, and facilitate the recording of insights. But the major issue remains: the insightful knowledge generated during AI-augmented investigations tends to vanish once the case is closed. Prompts, tool calls, and analyst choices infrequently feed into future detections, playbooks, or platform enhancements.

The ephemeral character of such AI engagements is a systemic inefficiency of how security agencies are using new technologies. When SOC analysts interact with AI assistants, they create worthwhile intellectual capital—learning effective investigation patterns, recognizing tooling gaps, and creating tacit insights regarding threat behaviors. Absent right instrumentation and feedback loops, this capital goes into closed cases instead of growing into organizational capabilities. The NIST AI Risk Management Framework (AI RMF 1.0) identifies this issue, citing that organizations often do not have processes in place to systematically gather and reuse knowledge created during AI-facilitated operations [1]. Likewise, the joint CISA and UK NCSC guidelines stress that numerous security teams cannot put in place systematic processes that convert AI-facilitated investigation trends into durable capability enhancements [2].

This work examines an architecture and design pattern set that shifts fleeting AI interactions into lasting organizational strength via cemented feedback loops. The envisioned method collects AI usage telemetry using OpenTelemetry standards and directs this information through four complementing

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

cycles: (1) detection tuning, extracting repeated investigation patterns into detection rules; (2) playbook creation, breaking repeated workflows into standardized procedures; (3) tooling gap exposure, revealing absent utilities and access needs; and (4) knowledge preservation, maintaining high-quality findings for future investigations.

These feedback loops operate within a governance framework that is aligned with well-established standards of that field, like the NIST AI RMF and Cybersecurity Framework 2.0. The architecture ensures that the enhancement of capabilities is always aligned with the organizational goals of security and governance through the implementation of adequate privacy, data reduction, and human intervention controls. The outcome is a system that translates every AI-aided inquiry into quantifiable improvements in detection range, analyst productivity, and organizational intelligence—building security capabilities that become more powerful with every use.

The following is the telemetry foundation, design patterns, validation framework, and implementation roadmap for this method. Organizations using these patterns will achieve quantifiable gains in mean-time-to-detection (MTTD), mean-time-to-respond (MTTR), analyst productivity, and knowledge transfer—making AI an enduring investigation tool instead of a temporary one.

2. The Core Challenge

When SOC analysts use AI for investigation, they create immense value with their interactions—uncovering efficient investigation patterns, spotting tooling gaps, and building tacit knowledge. This value goes to waste in closed cases if the organization lacks proper instrumentation and feedback loops.

This is a significant inefficiency in the use of advanced AI technologies by security organizations. Cybersecurity research indicates that security teams regularly go through repetitive investigative activities that can be improved through structured improvement cycles. The MITRE ATT&CK framework, which captures adversary tactics and techniques through observation of actual events, offers a shared language for characterizing threat activity but demands ongoing integration of emerging patterns that are discovered through investigations [3]. Yet few SOCs possess the infrastructure to systematically retain these investigation insights and turn them into improved detection capabilities.

The issue is especially severe with generative AI assistants, which generate considerable investigative value that is lost as soon as cases are closed. When analysts interact with these systems to enhance alerts, correlate behaviors with threat techniques, or produce investigation narratives, they tacitly put their expertise into prompts and evaluation choices. OpenTelemetry's Generative AI semantic conventions highlight the need to capture such interactions by means of structured telemetry that logs prompts, completions, tool usage, and contextual metadata [4]. In the absence of such instrumentation, organizations lose important chances to distill individual analyst skills into collective competencies that augment overall security posture.

This ongoing disparity between individual investigations and organizational capability is the result of various interacting factors. First, many AI assistants used in security scenarios function as discrete tools instead of being part of a learning system. Second, SOC standard metrics are centered on case-level performance measures (time-to-close, escalation rate) instead of knowledge capture and reuse. Third, security teams usually don't have the telemetry infrastructure required to recognize repeated patterns across cases. All these elements combine to form a system where critical intelligence is left stranded in chat logs and case notes instead of becoming detection rules, playbooks, or knowledge assets.

The price of this disconnection is high. Security teams keep resolving the same issues over and over, develop duplicate solution patterns, and do not systematically close tooling gaps found during investigations. New analysts are subjected to long onboarding processes even though the organization itself has already addressed comparable challenges in the past. Detection engineering is isolated from

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

investigation patterns, developing a consistent reactive stance instead of an ongoing improving capability. A solution to this issue necessitates a foundational change to the manner security organizations instrument, capture, and convert AI-aided investigations into lasting capabilities.

Challenge Category	Description	Impact	Potential Solution
Ephemeral Knowledge	AI-assisted investigation insights disappear after case closure	Repeated problem-solving, lost expertise	Structured telemetry capture
Siloed Tools	AI assistants function as discrete tools rather than learning systems	Disconnected capabilities, limited growth	Integrated feedback loops
Case-Centric Metrics	Focus on time-to-close rather than knowledge capture	Value prioritization misalignment	Knowledge reuse measurements
Missing Infrastructure	Lack of telemetry for pattern recognition	Unable to identify recurring workflows	OpenTelemetry implementation
Reactive Posture	Detection engineering is isolated from investigation patterns	Persistent security gaps, slower response	Detection refinement loop
Analyst Onboarding	New analysts undergo lengthy training despite existing solutions	Operational inefficiency, knowledge gaps	Knowledge curation mechanisms

Table 1: The Hidden Costs of Ephemeral AI: Lost Value in Security Operations [3, 4]

3. A Four-Loop Architecture for Sustainable Value

The architecture outlined collects AI-usage telemetry and converts it into four separate feedback loops:

- Detection Refinement Loop: Converting repeating investigation patterns into detection rules
- Playbook Synthesis Loop: Standardizing duplicate analyst processes into repeatable procedures
- Tooling Gap Identification Loop: Exposing the lack of utilities and access requests
- Knowledge Curation Loop: Saving high-quality findings for reference in future inquiries

These loops function within the architecture for governance that aligns with NIST AI Risk Management Framework (RMF), NIST Cybersecurity Framework 2.0, and CISA and the UK's NCSC guidelines for secure AI development.

The architecture solves the inherent problem of fleeting AI value by creating deliberate processes for acquiring, examining, and putting into action the insights produced by AI-facilitated investigations. Underpinning it is a rich telemetry layer that records all the interactions with AI, from prompts to completions, tool invocations, and analyst actions. This telemetry provides the substrate for patterns to be identified and transformed throughout the four loops.

The Detection Refinement Loop is the most straightforward way to enable greater security capabilities. By inspecting repeated patterns of investigation—the repeated sequences of enrichment operations that detect specific types of threats—the system can suggest detection rules that allow similar threats to be automatically identified. NIST AI Risk Management Framework highlights the need for such feedback mechanisms, highlighting that AI systems would include processes that "turn operational experience into quantifiable system improvements" for improving security posture over time [1]. This cycle ties detection engineering to true investigation patterns, forming a closer link between detection capabilities and security operations.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Playbook Synthesis Loop helps solve the problem of operational consistency by highlighting recurring workflows for multiple analysts and investigations. When several analysts use similar procedures to investigate specific alert types, these patterns are opportunities for standardization. NIST Cybersecurity Framework 2.0 expressly calls for the establishment of "systematic response processes" under the Respond (RS) function, reinforcing how "response planning processes and procedures are executed and maintained" to provide uniform treatment of security incidents [5]. By converting these patterns into playbooks with the right human oversight gates, organizations can enhance consistency, speed up investigations, and decrease the workload on analysts.

The Tooling Gap Identification Loop targets infrastructure constraints that hinder productive investigations. AI assistants typically undertake actions that do not work because of a lack of utilities or inadequate access rights—trying to parse log formats without the right parsers or inquire about systems without required entitlements. By pooling these unsuccessful actions and frequency across investigations, companies can systematically determine and rank tooling gaps by their operational impact. This strategy turns frustration over stalled actions into an organized capability-improvement plan consistent with the NIST AI RMF's focus on "continuous assessment of system capabilities against operational requirements" [1].

The Knowledge Curation Loop also retains the useful knowledge created through investigation for later use. Where AI assistants generate especially useful analyses or explanations, these are knowledge artifacts that must continue to exist beyond the immediate investigation. By recording these findings with their context and sources, organizations build an institutional knowledge base that facilitates future research and analyst ramp-up. This method aids the NIST Cybersecurity Framework's Recovery (RC) function of stressing the need to "incorporate lessons learned into revised response strategies" within ongoing improvement [5].

The four loops collectively constitute a cycle of self-improvement. Enhancements in detection decrease analyst burden on known threats, freeing capacity for more thorough investigation of new threats. Standardization of playbooks enhances consistency and provides transparent workflows for automation. Enhancements to tooling eliminate drag from investigations, speeding up results and lessening analyst angst. Preserving knowledge speeds up follow-on investigations and onboarding of analysts, forming a cycle of capability increase.

Architecture includes proper governance controls to allow these loops to function within formal security and AI management contexts. All suggested improvements are properly reviewed and verified, including human approval points for security-critical modifications. The system produces extensive audit histories of pattern detection and transformation, establishing transparency and responsibility throughout improvement. This governance methodology parallels the NIST AI Risk Management Framework's "Govern" function, which calls for "establishing ongoing oversight processes for AI systems throughout their lifecycle" [1] and the Cybersecurity Framework's Governance (GV) function, which centers on "developing and implementing the appropriate structures, policies, and processes" to address cybersecurity risk [5].

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Feedback Loop	Input Source	Primary Output	Key Benefit	Governance Alignment
Detection Refinement	Repeated investigation patterns	Detection rules	Automated threat identification	NIST AI RMF - "Operational experience into system improvements"
Playbook Synthesis	Recurring analyst workflows	Standardized procedures	Operational consistency	NIST CSF 2.0 - "Systematic response processes"
Tooling Gap Identification	Failed/blocked agent actions	Infrastructure improvement roadmap	Reduced investigation friction	NIST AI RMF - "Continuous capability assessment"
Knowledge Curation	Validated AI analyses	Institutional knowledge base	Accelerated analyst onboarding	NIST CSF 2.0 - "Lessons learned integration"

Table 2: SOC AI Feedback Loop Architecture: Transforming Ephemeral Insights into Durable Capabilities [1, 5]

4. Design Pattern 1: Usage→Detection Loop

When analysts continually repeat the same cycle of enrichment and investigation steps for comparable alerts, these patterns are detection improvement opportunities. The Usage→Detection loop:

- Extracts repeated assistant/agent interaction patterns (e.g., routine enrichment sequences for phishing alerts)
- Suggests detection or correlation rules from these patterns
- Aligns to pertinent MITRE ATT&CK techniques
- Monitors false-positive rates prior to and following implementation

This trend transitions the organization from being purely reactive to progressively proactive security stances through the integration of analyst skills into detection logic.

The Usage→Detection loop maps repeated investigation patterns into stronger detection abilities through deterministic analysis of AI assistant use. Through the determination of when multiple analysts utilize similar investigation paths for specific alert types, the system can suggest detection improvements that automate the recognition of similar threats. This approach aligns with MITRE D3FEND's knowledge graph architecture, which provides "a framework for reasoning about cybersecurity countermeasures" by connecting defensive techniques to the offensive techniques they counter, enabling organizations to systematically improve their detection coverage based on observed attack patterns [6].

The implementation process begins with telemetry analysis to identify investigation sequences that recur across multiple cases with similar outcomes. The system subsequently picks out the principal indicators and correlation patterns from such sequences and converts them into suggested detection rules or correlation logic. Each of the proposals involves mapping to applicable MITRE ATT&CK techniques for adequate categorization and coverage evaluation. The FIRST CSIRT Services Framework places particular stress on the necessity of such formalized methods of detection improvement, citing that effective security services call for "continuous refinement of detection capabilities based on operational insights" within the "Monitoring and Detection" service field [7].

Before deployment, suggested detection improvements are validated against historical data to determine potential rates of false positives. Post-deployment, actual performance is monitored against these forecasts, establishing an ongoing validation loop. This evidence-based practice ensures that detection improvement yields quantifiable security value with reduced alert fatigue. By converting

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

analyst skill into detection logic, the company transitions from reactive-only security operations to a more proactive stance with constantly enhanced detection ability.

Stage	Process Steps	Outcomes	Metrics	Framework Alignment
Pattern Extraction	Analyze telemetry for recurring analyst actions	Identification of common investigation sequences	Pattern frequency across analysts	MITRE D3FEND - Knowledge Graph
Rule Development	Convert investigation patterns to detection logic	Automated detection capabilities	Coverage of previously manual patterns	MITRE ATT&CK - Technique Mapping
Validation Testing	Evaluate against historical data	False positive prediction	Precision/recall metrics	Evidence-based detection
Production Monitoring	Track performance post-implementation	Continuous refinement cycle	False positive reduction rate	FIRST CSIRT - Detection Services
Capability Enhancement	Document and extend detection coverage	Expanded proactive detection	Reduced manual investigation time	Operational threat visibility

Table 3: Translating Analyst Intelligence into Automated Detection [6, 7]

5. Design Pattern 2: Usage→Playbook Loop

Where several analysts perform similar multi-step investigation recipes, these are opportunities to standardize. The Usage→Playbook loop:

- Infers repeated investigation workflows among analysts
- Composes these into SOC platform-compatible or CACAO-compatible playbooks
- Embeds human-in-the-loop gates for state-altering actions
- Tracks the decrease in Tier-1 analyst toil after deployment

This pattern enhances consistency at the cost of proper human oversight for key decisions.

The Usage→Playbook loop converts iterative investigation workflows into routine procedures that increase consistency and speed. Through the evaluation of AI assistant interactions within many analysts and cases, the system determines shared workflow patterns that are candidates for standardization. This process is consistent with some of the latest research into security orchestration, which indicates that systematic examination of operational patterns can uncover "standardization opportunities that balance automation with appropriate human judgment" within security workflows, especially when using AI-augmented investigation telemetry [8].

The deployment starts with the discovery of workflows that look across multiple analysts with similar results—like typical enrichment orders for particular alert types or typical response methods for given incidents. The system then combines these patterns into formatted playbooks in standard security automation protocols. As emphasized in NIST's Special Publication 800-61r2 Computer Security Incident Handling, incident response playbooks are critical to successful incident response: "Applying the same method to incident handling is critical. Incident response procedures should be documented" to preserve consistency and efficiency [9].

One of the most important features of this pattern is the inclusion of proper human oversight gates. Though standard enrichment steps can be automated completely, state-altering actions—such as isolation decisions or account changes—must receive specific analyst approval. This balances gained efficiency with proper human judgment for impactful decisions. Once deployed, the system gauges

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

impact through measurement of the decrease in Tier-1 analyst time devoted to mundane work, allowing for data-driven evaluation of value for the standardization.

By turning individual skill into an organizational process, the Usage→Playbook loop enhances operational consistency, facilitates faster routine investigations, and saves analyst capacity for hard cases involving deeper expertise.

Process Stage	Key Activities	Outcomes	Metrics
Workflow Identification	Analyze AI-assisted investigations across analysts	Common investigation patterns discovered	Pattern frequency and consistency
Playbook Creation	Transform patterns into structured workflows	SOC platform-compatible or CACAO-compatible playbooks	Standardization coverage
Human Oversight Design	Identify critical decision points	Human approval gates for state-changing actions	Governance compliance
Automation Implementation	Automate non-critical enrichment steps	Reduced manual processing	Time saved per investigation
Performance Tracking	Measure impact on analyst workload	Quantified efficiency improvements	Tier-1 analyst time reduction

Table 4: Standardizing Security Operations Through Playbook Automation [8, 9]

6. Design Pattern 3: Usage→Tooling Loop

AI assistants tend to try things that don't work because there are missing utilities or a lack of access rights. The Usage→Tooling loop:

- Summarizes "failed" or "blocked" agent actions
- Exposes missing utilities (parsers, sandboxes, identity scopes)
- Translates these gaps into development backlog items
- Estimates ROI in terms of potential time saved per case

This trend methodically detects and bridges infrastructure deficits hindering investigation efficiency.

The Usage→Tooling cycle of feedback converts failed AI assistant activity into a formal plan for infrastructure enhancement. By detecting situations where AI agents try but fail to perform actions—lacking parsers, lacking access rights, or lacking utilities—the system recognizes capability deficits affecting investigation efficiency. This strategy is consistent with "Hidden Technical Debt in Machine Learning Systems," which asserts that ML systems rely not only on models but also on the environment surrounding them: "only a small fraction of real-world ML systems is made up of the ML code. The surrounding infrastructure required by it is enormous and convoluted" [10].

The deployment starts with instrumenting AI assistant interactions to record blocked or failed actions and their related context—like trying to parse unsupported log formats, accessing systems with improper credentials, or running analyses without required tools. These failures are tallied across investigations to determine repeating patterns and high-impact gaps. According to research on security orchestration and automation, successful security operations involve ongoing enhancement of the "supporting technology stack that allows security analysts to effectively execute response actions" [11].

After they are identified, such gaps are transformed into development backlog items with explicit value propositions according to their operational effect. Every item contains a calculation of ROI that expresses the potential time saved per impacted case times the rate of occurrence, providing a fact-based prioritization system. With this strategy, infrastructure investments are always aligned with real operational demands and not speculative ones.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

By methodically finding and filling tooling gaps, organizations eliminate friction from investigations, enhance analyst experience, and better enable effective AI assistants. This creates a virtuous cycle where infrastructure enhancements facilitate more productive investigations, which in turn uncover more advanced improvement opportunities.

7. Design Pattern 4: Usage→Knowledge Loop

Good AI responses frequently embody good organizational knowledge. The Usage→Knowledge loop:

- Sustains validated AI outputs with their citations
- Converts these into reusable knowledge artifacts
- Makes them accessible for future cases and analyst onboarding
- Aligns materials to NIST IR and FIRST CSIRT service models

This pattern retains investigation insights beyond single cases, speeding up ongoing operations and new analyst onboarding.

The Usage→Knowledge loop converts vetted AI-authored content into lasting organizational assets. By flagging high-quality AI responses that have been validated by analysts—like cogent threat explanations, sound investigative methods, or helpful correlation patterns—the system retains valuable insights above specific cases. This practice is in line with research on situation awareness for dynamic systems that draws attention to the fact that effective operational teams need not only individual competence but "shared mental models and distributed knowledge" that can be tapped into across the organization. The study points out that in multi-faceted security environments, "knowledge preservation mechanisms are essential for maintaining operational continuity" even as personnel changes [12].

Implementation commences by capturing AI-generated material that analysts have specifically validated or improved upon in the course of investigations. Such information is processed to preserve factual accuracy, correct attribution of sources, and suitable classification of sensitivity levels. As noted in the NICE Cybersecurity Workforce Framework, successful security teams need methodical methods of "knowledge management and skills development" that maintain institutional knowledge and make it available for workforce development. The framework underlines that "cybersecurity work calls for specialized knowledge that must be structured and made accessible" to enhance current operations as well as workforce development [13].

The resultant knowledge artifacts are converted into retrievable items with consistent metadata tagging mapped to approved standards such as NIST IR procedures and FIRST CSIRT service categories. The uniform tagging facilitates context-aware recovery in subsequent investigations and establishes learning paths for new analysts joining the team as structured entities. The system observes the usage patterns in order to identify the valuable knowledge assets and gaps in knowledge that require additional documentation.

Maintaining investigative results in an institutional knowledge format, as opposed to an individual knowledge format, enables organizations to reduce the dependence of specific analysts, accelerate the onboarding of new employees, and maintain operational consistency following a change in personnel. This creates an institutional capital of knowledge, which grows with each inquiry.

8. Risk Management and Ethical Considerations

The architecture addresses several major areas of risk:

- Privacy protection through enforced masking, minimum retention policies, and access controls
- Feedback poisoning prevention by avoiding direct re-training from raw chat logs
- Prompt-injection safeguards with red-team testing aligned to OWASP LLM Top 10

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

- Governance alignment with ISO/IEC 23894:2023 for purpose limitation and data retention rationale

While the four feedback loops generate large amounts of operational value, they also present particular risks that need to be systematically addressed. Applying these loops in an overall risk management strategy helps guarantee capability enhancements do not undermine security, privacy, or governance demands. Architecture includes a number of key safeguards consistent with evolving standards for secure implementation of AI.

Protecting privacy is an essential concern when collecting AI interaction telemetry. Security audits often work with sensitive information, such as personally identifiable information, authentication credentials, and trade-secret business information. The OWASP Top 10 for Large Language Model Applications cites "LLM01: Prompt Injection" and "LLM07: Insecure Output Handling" as urgent risk categories that may lead to data leakage and sensitive data disclosure. The architecture highlights that organizations need to enforce "appropriate data sanitization, well-defined data boundaries, and output validation" to avoid unauthorized access to confidential information via AI interactions [14]. The framework deals with this via enforced sensitive content masking, minimal retention policies retaining only necessary interaction patterns, and rigorous access controls based on least-privilege principles.

Prevention of feedback poisoning ensures that improvement loops are made robust against adversarial manipulation. If an attacker can manipulate the responses of the AI assistant and these are directly added to security playbooks or rules for detection, they might be able to compromise security controls. This threat is consistent with ISO/IEC 23894:2023, which requires that AI risk management should comprise "controls against adversarial attacks and manipulations" and clarifies that "AI systems employed for security-critical operations need more rigorous validation procedures and control mechanisms" so that they are not compromised [15]. The architecture mitigates this through multi-level validation processes, requiring explicit analyst approval for improvements, and maintaining comprehensive audit trails of all modifications.

Prompt-injection mitigations defend against efforts to trick the AI assistant into making inappropriately suggestive recommendations or evading security measures. It is designed with regular attacks by the red team on the OWASP LLM Top 10 vulnerabilities, with a particular focus on the immediate cases of injection, which may disrupt the feedback loops. It should be noted that these tests are necessary to make sure that the system is resistant to unexpected abuses and even systematic manipulation.

This is because the governance has been aligned so that every data collection and retention practice is based on the needs of the organization and the industry. The architecture enforces purpose limitation principles of ISO/IEC 23894:2023, documenting explicitly the operational reason for collecting all telemetry and defining suitable retention policies through proven value. This establishes transparency and accountability across the system lifecycle, guaranteeing that mechanisms for improvement are in function within defined governance structures.

Conclusion

The shift of the short-lived interaction between AI assistants to long-term organizational capabilities is one of the opportunities for SOCs. Through the use of organized feedback mechanisms and instrumentation, governance, and validation, security teams can choose to guarantee that every AI-aided investigation results in a sustainable change in detection, responses, tooling, and organizational know-how. Not only does this more fully harness AI investments, but it does so within systems that provide proper control, privacy, and risk management, and transforms what would be otherwise isolated AI interactions into a constantly developing security capability.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

References

[1] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST Special Publication NIST.AI.600-1, 2023. [Online]. Available: https://www.nist.gov/itl/ai-risk-management-framework
[2] NCSC, "Guidelines for Secure AI System Development,". [Online]. Available: https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development
[3] MITRE Corporation, "ATT&CK,". [Online]. Available: https://attack.mitre.org/
[4] OpenTelemetry, "Semantic conventions for Generative AI events,". [Online]. Available: https://opentelemetry.io/docs/specs/semconv/gen-ai/gen-ai-events/
[5] National Institute of Standards and Technology, "Cybersecurity Framework 2.0," 2024. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.29.pdf
[6] Peter E. Kaloroumakis and Michael J. Smith, "Toward a Knowledge Graph of Cybersecurity Countermeasures," MITRE Technical Report. [Online]. Available: https://d3fend.mitre.org/resources/D3FEND.pdf
[7] FIRST.org, "CSIRT Services Framework Version 2.1,". [Online]. Available: https://www.first.org/standards/frameworks/csirts/FIRST_CSIRT_Services_Framework_v2.1.0.pdf
[8] Orestis Tsirakis et al., "Operationalizing Cybersecurity Knowledge: Design, Implementation & Evaluation of a Knowledge Management System for CACAO Playbooks," arXiv:2503.05206v2, 2025. [Online]. Available: https://arxiv.org/html/2503.05206v2
[9] A. Nelson, S. Rekhi, M. Souppaya, and K. Scarfone, "Computer Security Incident Handling Guide," NIST Special Publication (SP) 800-61, Revision 2, 2025. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-61r2.pdf
[10] D. Sculley et al., "Hidden Technical Debt in Machine Learning Systems,". [Online]. Available: https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf
[11] Andy Applebaum et al., "Intelligent, automated red team emulation," ACSAC '16: Proceedings of the 32nd Annual Conference on Computer Security Applications, 2016. [Online]. Available: https://dl.acm.org/doi/10.1145/2991079.2991111
[12] Mica R. Endsley, "Measurement of Situation Awareness in Dynamic Systems," Human Factors, vol. 37, no. 1, pp. 65-84, 1995. [Online]. Available: https://www.researchgate.net/publication/200773058_Measurement_of_Situation_Awareness_in_Dynamic_Systems
[13] National Initiative for Cybersecurity Education, "National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework,". [Online]. Available: https://www.cisa.gov/national-initiative-cybersecurity-education-nice-cybersecurity-workforce-framework
[14] OWASP Foundation, "OWASP Top 10 for Large Language Model Applications,". [Online]. Available: https://owasp.org/www-project-top-10-for-large-language-model-applications/
[15] International Organization for Standardization, "Information technology — Artificial intelligence — Guidance on risk management," ISO/IEC, 2023. [Online]. Available: https://cdn.standards.iteh.ai/samples/77304/cb803ee4e9624430a5db177459158b24/ISO-IEC-23894-2023.pdf