From Analyst to Orchestrator: The Emerging Cognitive Role of Humans in Agentic Security Operations

Daniel Boughton[1], and Iain Reid[1,*]

[1] *School of Criminology and Criminal Justice, University of Portsmouth, Portsmouth, United Kingdom*

Abstract

Security Operations Centres (SOCs) are being pitched as the next frontier for AI-driven transformation, but vendor narratives routinely collapse three different concepts into one: conversational large language model (LLM) assistants, conventional automation, and agentic AI. The conflation matters. A practitioner account illustrates how chat interfaces can insert a translation layer between experienced analysts and their telemetry, whereas agentic systems change what can be delegated. When bounded, specialised, and embedded in the workflows analysts already use, they can take on substantive parts of the analyst's workload. The question for security operations is therefore not whether AI will replace the SOC analyst, but how the cognitive work of the SOC is being redistributed between humans and a portfolio of bounded agents. Drawing on socio-technical systems theory, human factors research, and the lived practitioner experience of one of the authors working in a mature SOC, we propose an Analyst-Agent Orchestration Framework that distinguishes manual investigation, conversational assistance, task-specific agents, coordinated agent use, and high-autonomy operation. These are configurations of work, rather than stages of maturity. The argument is that the analyst's role is shifting from investigator to orchestrator. Trust calibration, evidence-first design, and meaningful governance are now the central problems in the design of agentic security operations.

Keywords

Agentic AI; Security Operations Centre; SOC; human-AI collaboration; socio-technical systems; cybersecurity; automation; analyst cognition.

1. Introduction

Security Operations Centres are work systems in which people, processes, technology, and organisational priorities interact under adversarial pressure to detect and respond to threats. The literature on SOCs has long acknowledged this, with Vielberth et al. [1] noting that SOCs combine people, processes, technologies, governance, and compliance, and observing that the academic literature has struggled to settle on a consistent definition of the SOC and its boundaries [1, 2].

The 12th International Conference on Socio-Technical Perspectives in IS (STPIS'26), September 16–17, 2026, L'Aquila, Italy.

* Corresponding author.

daniel.boughton@myport.ac.uk (D. Boughton); iain.reid@port.ac.uk (I. Reid)

0009-0003-1574-8572 (D. Boughton); 0000-0003-4072-7557 (I. Reid)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-4325/paper6.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-10

The pressures on SOCs are well documented, from alert volume and analyst burnout to inconsistent tooling and incomplete context. Measuring analyst performance is difficult too [3, 4, 5, 6]. Those pressures create a market for AI-driven solutions, and vendors are not slow to supply them. Faster triage, reduced workload, automated response, improved detection: these are the marketing claims, made for tools such as Microsoft Security Copilot [7]. The practical value of any particular tool, though, depends less on model capability than on workflow fit, analyst trust, organisational context, and how responsibility is divided between humans and machines.

This position paper argues that the discussion of AI in security operations needs a sharper socio-technical distinction between these three related but different ideas. Conversational large language model (LLM) tools respond to prompts and generate text. Conventional automation executes predefined workflows. Agentic AI plans and selects tools to work toward a goal within a defined environment. It performs actions and gathers evidence, iterating as it goes [8, 9]. The distinction is operationally significant. The practitioner account below illustrates the friction that conversational tools can introduce into an established investigative workflow, and the different demands raised by task-specific delegation.

The position we take here is that agentic AI is not going to remove the need for deep security expertise in mature SOCs. It is going to change how that expertise is exercised. Analysts are becoming orchestrators of specialised agents, deciding what to delegate and under what constraints. They must then evaluate the evidence and judge whether machine-generated conclusions are safe to act on, without losing situational awareness (SA). The shift changes the cognitive role of the analyst, raising real questions about training and skill retention. Trust remains a concern, as do accountability and governance.

The aim of this position paper is to examine how agentic delegation redistributes cognitive work and responsibility in mature SOCs. The question we address is: how should human-agent work be organised so that delegation supports operational performance while preserving analyst judgement, situational awareness, and accountable control? The contribution is a framework of analyst-agent work configurations, together with design principles and a research agenda for testing them. Training, skill retention, trust, and governance are treated as parts of this work system.

2. SOCs as Socio-Technical Work Systems

A socio-technical view treats organisational performance as a function of the fit between social and technical components. The tradition goes back to early work on the relationship between work organisation and technical systems [10]. Mumford's [11] account of this tradition emphasises its humanistic commitment to improving working life alongside technical change. Later socio-technical systems engineering literature has stressed that human, social, organisational, and technical factors have to be considered together at the design stage [12]. Sommerville's [13] account of a national digital learning environment illustrates the practical importance of stakeholder involvement, usability, and governance in software design. The view is unusually relevant to SOCs because security operations involve dynamic interaction between skilled practitioners, security tools, adversarial behaviour, organisational risk appetite, and formal response processes.

In a modern enterprise SOC, the technical layer might include a Security Information and Event Management (SIEM) system such as Microsoft Sentinel, Endpoint Detection and Response (EDR) tooling such as Defender for Endpoint, Security Orchestration, Automation and Response (SOAR) playbooks, threat intelligence feeds, case management, identity and cloud telemetry (operational data from those systems), vulnerability data, and a growing number of AI assistants [7, 14]. The social layer covers analysts at different tiers, incident responders, engineers, managers, risk owners, legal counsel, system owners, and business stakeholders. The output of the SOC emerges from the interaction of all of these. A technically powerful platform can fail if it does not fit the way analysts actually work, undermines trust, or adds cognitive load. A technically modest tool can be valuable if it removes friction from a well-understood task.

The framing "Can AI do SOC work?" is therefore close to meaningless. SOC work is a portfolio: monitoring, triage (initial alert assessment), enrichment (adding context), investigation, reporting, containment (limiting an incident's spread), escalation, hunting, detection tuning, vulnerability coordination, compliance support, post-incident learning, and a long tail of administrative work that rarely gets dignified in the literature. Agyepong et al. [3] note that analyst performance is hard to measure largely because analyst functions are so varied; the metrics that exist tend to capture volume rather than judgement. If that is right, then the more useful question is which parts of the portfolio agentic AI can plausibly take on, and the answer is unlikely to be the same in two SOCs of similar size. A single agent that “does SOC work" is a vendor abstraction. What is operationally plausible is a collection of narrower agents, each governed and validated separately, with the caveat that integrating several of them is itself a non-trivial piece of work.

A socio-technical framing therefore focuses on the combined performance of the human-agent work system. Automation research asks how delegation changes the work left to people [15], situational awareness concerns the understanding needed to act in a changing environment [16], and trust research examines appropriate reliance [17]. These perspectives guide the practitioner account, framework, and design principles that follow.

3. From Conversational LLMs to Task-Specific Agents: A Practitioner Experience

The first wave of generative AI adoption in many organisations centred on conversational interfaces [14]. In SOCs, these tools are often pitched as analyst copilots: systems that answer questions, summarise alerts, explain logs, draft reports, or suggest next steps. There are tasks for which this is genuinely useful, particularly documentation, summarisation, training, and knowledge retrieval. There are also tasks for which it is poorly aligned with how experienced analysts work.

A short account drawn from the first author’s practitioner experience illustrates the misfit. In 2024, the internal SOC of a global software and payments company took part in a hands-on private preview of a conversational security assistant in its production environment. The SOC had established investigative workflows and conventional, non-AI automation. The first author was directly involved in the decision to participate, evaluating the tool, attending vendor-led sessions, and discussing and collating the team’s experience of using it. The purpose was to assess whether the organisation would purchase the service once generally available.

The reception among experienced engineers was sceptical, and the reasons were practical rather than aesthetic. The engineers already had direct access to the telemetry, the queries, and the case management. They were already capable of pivoting from an alert to the underlying queries, from those queries to an endpoint process tree, and from there to identity and network signals in the SIEM. In the first author's assessment, the conversational tool inserted a step into that flow rather than removing one. It asked them to type a description of what they wanted to know, wait for a narrative answer, and then validate that narrative against the evidence they could already see. Where the analyst's work was already grounded in evidence, the chat interface introduced a layer of indirection that was at best redundant and at worst misleading. The concern was the additional prompting and validation required in an established workflow. These are observations from use and discussion, rather than measured estimates of time or cognitive effort.

This account is not a formal case study in the methodological sense [18], and is not offered with the research design or the claims to generalisability that such a study would require. It is presented instead as practitioner experience: a sensitising example drawn directly from the first author's work in enterprise security operations. The problem with some conversational AI tools in SOCs is not only hallucination, although hallucination matters and is well-documented [19, 20]. The problem is also cognitive and procedural. A tool can be accurate enough across most cases and still be unwelcome if it disrupts expert flow, forces analysts into unnatural interaction patterns, or returns prose where the analyst needs queries, pivots, or actions.

Agentic AI changes the proposition, where rather than answering questions, agents are built to perform bounded tasks, for example: retrieve relevant logs, enrich indicators, correlate entities across signals, query threat intelligence, generate timelines, compare observed activity against known Tactics, Techniques and Procedures (TTPs), prepare a draft case summary, or execute a constrained containment workflow under approval. Recent surveys of LLM-based autonomous agents describe architectures that combine planning, memory, tool use, environmental perception, and action [8, 9]. Sapkota et al. [21] draw a sharper conceptual line between AI agents as modular task-specific systems and agentic AI as coordinated multi-agent autonomy. In cybersecurity specifically, Kshetri [22] argues that agentic AI is likely to automate critical SOC tasks such as threat detection, decision support, and incident response, while also introducing new risks that require existing security frameworks to be reassessed.

The distinction matters in operational terms. A conversational LLM asks the analyst to bring work into a chat window. A well-designed agent is brought into the analyst's existing workflow. The first feels like asking a generalist for an opinion on something the analyst can already see. The second is more like delegating a defined subtask to a specialist that produces evidence rather than narrative.

A later example comes from trials of triage agents in early 2026, in the same SOC's production environment. The team evaluated pre-configured agents and also worked on building its own. Early iterations often sat on the fence: they struggled to reach a conclusion about whether an alert reflected benign activity or a true positive. When they did reach a conclusion, there was a tendency to label activity suspicious rather than benign. These are observations from operational trials, not results from a controlled comparison, and do not establish an error rate. The practical concern was whether the agent produced a verdict that the analyst could use.

The design implication is that a triage agent should distinguish a supported conclusion from an unresolved case. A cautious label is not a substitute for identifying what evidence is missing. Building such an agent involves more than prompt design. It calls for specifying which queries should precede a verdict, making uncertainty explicit when key telemetry is missing, escalating unresolved cases for human review, budgeting queries, and restricting remediation permissions. These are proposed requirements for the workflow, rather than a claim that every control was enforced in the trials. They are consistent with guidance on least privilege, scope limitation, and human oversight [23, 24].

4. Position: The Analyst as Orchestrator

The position we wish to argue is that agentic AI is not going to remove the need for skilled SOC analysts in mature operations, however it will change the shape of their work. The analyst's role is shifting from performing each investigative step directly toward orchestrating a mixture of manual analysis, automated workflows, and specialised agents.

Delegation changes both the skills analysts exercise and the opportunities through which those skills develop. Some routine enrichment and basic triage work will be taken on by agents, and that is largely a good thing. The analyst remains responsible for framing the problem in the first place and judging whether the agent's path through it is appropriate. They must interpret uncertainty and recognise when context is missing before deciding whether a proposed action is proportionate. This aligns with a long-standing finding in automation research: automation does not simply remove human work. It changes the nature of that work and frequently raises the importance of human oversight during abnormal or ambiguous situations [15]. The more capable the automation, the more the design effort needs to focus on appropriate human control, trust, and intervention.

Several roles fall out of this, and it is worth pulling them apart because they draw on different skills. The analyst frames the problem, which is to say they decide what is actually being investigated; reflecting on the practitioner experience this is the step most often skipped under pressure, and it is also the step most resistant to agentic support, because the framing depends on context the agent does not have. The analyst then dispatches: deciding whether the appropriate next action is enrichment, malware triage, identity investigation, cloud correlation, phishing analysis, or endpoint containment, and choosing the tool or agent for it. They adjudicate evidence, which is the role that most needs protecting, because it is the one that gets compressed first when alert volume climbs. And beneath all of that they hold situational awareness in something like Endsley's [16] sense, though that framework needs handling with some care in this setting.

Endsley's [16] three-level model (perception, comprehension, projection) addresses dynamic decision making, including military aviation and adversarial situations. Its application to SOCs concerns a heterogeneous environment of identities, endpoints, cloud workloads, network paths, and third-party integrations, much of which is opaque even to the people responsible for it. And the timescales are uneven: some incidents demand cockpit-speed responses, others unfold over weeks. Endsley's framework is used here as a vocabulary for what analysts need to retain when agents undertake parts of an investigation, rather than as a model validated by this account.

Agentic workflows add a further difficulty. The Open Worldwide Application Security Project (OWASP) identifies indirect prompt injection as a way for hostile instructions in external content to influence a model's behaviour [24]. The corresponding risk in a SOC is that manipulated evidence or instructions could pass through agent handoffs and emerge as an apparently coherent account. Analysts therefore need access to source evidence, missing or contradictory observations, and the distinction between what was observed and what the agent inferred. This connects situational awareness to the evidence-first principle in Section 7.

There is one further role that deserves separate treatment because it carries more weight than the others. The analyst is the risk governor. They decide whether an agent should act, whether the action is reversible, whether business impact has been considered, and whether human approval is required. This matters most for high-impact responses such as account disablement, host isolation, firewall changes, privilege revocation, or automated takedown activity. Agentic tools that recommend such actions are useful. Agentic tools that take them without explicit approval are a different category of system entirely, and one that needs to be governed accordingly.

Taken together, these roles suggest that agentic AI does not eliminate expertise. It moves expertise upstream. The analyst becomes less like a passive consumer of AI output and more like an incident commander supervising a portfolio of bounded autonomous capabilities. How these responsibilities are exercised depends on the scope of delegation and the points at which humans retain control.

5. The Analyst-Agent Orchestration Framework

To make the shift concrete, an Analyst-Agent Orchestration Framework is proposed. It distinguishes configurations of work by what is delegated and where human review occurs, and identifies the corresponding human cognitive role.

Table 1
Configurations of Analyst-Agent Work

Configuration	Delegation and review boundary	Human role	Principal risk
Manual investigation	Analyst directly queries tools, reads logs, correlates evidence, and decides next steps.	Investigator	Slow response, fatigue, inconsistency.
Conversational assistance	Analyst asks an LLM to summarise, explain, draft, or suggest, and decides the next investigative step.	Questioner and validator	Workflow interruption, hallucination, overreliance.
Task-specific agent	An agent completes a defined task, potentially using several tools. The analyst reviews the outcome before progression beyond that task.	Delegator and evidence reviewer	Narrow automation error, hidden assumptions.
Coordinated agent use	A system manages handoffs between specialised agents across investigative stages, within agreed checkpoints and escalation conditions.	Orchestrator and supervisor	Coordination failure, error propagation, fragmented accountability.
High-autonomy operation	Agents initiate and continue authorised work under standing policy. Human review is triggered by exceptions or specified thresholds.	Overseer or exception handler	Excessive agency, loss of situational awareness, unsafe action.

These configurations describe arrangements of work rather than stages of maturity. They apply to a task or workflow, not to an entire SOC, and several may coexist within an investigation. High autonomy may apply to a single agent or a coordinated workflow. The appropriate arrangement depends on task criticality, reversibility, the evidence available, the operational context, and the potential business impact of getting the response wrong. High-impact decisions require meaningful human control regardless of how technically capable the agent is. Conventional automation can support any of these arrangements by executing predefined parts of the workflow.

The distinction between task-specific and coordinated agent use concerns control over progression between investigative stages, rather than the number of tool calls. A triage agent can retrieve logs, query threat intelligence, and correlate entities while returning one bounded outcome for review. Coordinated use delegates the handoffs between specialised agents across stages, subject to agreed checkpoints and escalation conditions. Human review may therefore occur during the workflow as well as at its conclusion.

The framework also draws a useful line between assistance and orchestration. With conversational assistance, the analyst is using an AI tool. With task-specific or coordinated delegation, the analyst is managing agents as operational resources, which is a different kind of work. It requires understanding not only the security problem but the agent's capabilities, its limitations, its permissions, and its failure modes. That is a cognitive demand that is not yet well reflected in SOC role definitions, training programmes, or performance metrics.

6. Risks of Agentic AI in Security Operations

The move from analyst to orchestrator introduces risks. Some are familiar from earlier automation research. Others are specific to LLM-based and agentic systems. Four are worth treating in detail, with a fifth that sits underneath the others.

6.1. Automation complacency and overreliance

Trust in automation is not a purely technical matter. Lee and See [17] argue that trust shapes reliance particularly in complex systems where users cannot fully understand what the automation is doing. In a SOC, that creates the risk that analysts accept agent outputs because the system appears authoritative, especially under time pressure and high alert volume. OWASP's guidance on LLM security highlights overreliance and excessive agency as specific risks in LLM-enabled systems, particularly where outputs are not critically assessed or where the system has been granted permissions wider than the task requires [24]. The mechanism is not really carelessness, although it can look like it in a post-incident review. It is closer to a slow recalibration of effort. Alert volume, time pressure, and interface design combine to make weak validation feel reasonable, and reasonable weak validation is still weak validation. Polished summaries with confident recommendations can encourage acceptance without sufficient checking [17]. The research question is which interface designs can interrupt that drift in SOC workflows.

6.2. Loss of situational awareness

Agentic systems can gather and process evidence faster than humans, but they can also hide the investigative path. An agent that reports "likely credential compromise" without showing

which logs were queried, which assumptions were made, which alternatives were considered, and what evidence was conspicuously absent gives the analyst a conclusion without giving them comprehension. That matters because SOC work is dynamic. Analysts need to understand not only what happened, but what might happen next. A tool that produces answers without preserving the analyst's understanding can improve apparent efficiency while weakening operational resilience.

6.3. Skill degradation

Bainbridge's [15] observation that automation creates an irony, leaving the human responsible for abnormal situations while removing the opportunities to practise the skills required to handle them, has direct application here. If agents handle routine triage, enrichment, and investigation, junior analysts get fewer opportunities to learn how to do those things manually. Over time, the organisation can become dependent on agentic systems while losing the human capacity to detect when those systems are failing or to operate without them. Bednar and Welch [25] raise the related question of how people will acquire the task knowledge needed to exercise discretion when intelligent agents perform the work. Agentic workflows should therefore preserve learning: exposing evidence trails, supporting "show me how you got there" explanations, requiring confidence classifications, and periodically taking analysts off agent-assisted workflows to keep manual skills current.

6.4. Excessive agency and unsafe action

Agentic AI becomes consequential when it can act. Tool use, plugins, permissions, and platform integrations let agents modify tickets, query sensitive datasets, disable accounts, isolate hosts, initiate scans, or change configurations. OWASP describes excessive agency as a vulnerability arising from too much functionality, too many permissions, or too much autonomy in LLM-based systems [24]. In a SOC, the equivalent concern is that an agent takes damaging action because of a misleading prompt, a compromised input, a hallucinated chain of reasoning, or a misconfigured permission boundary. The design implication is straightforward: agent operation in SOCs should be built on least privilege, approval gates for high-impact action, reversible operations where possible, comprehensive audit logging, and a clear separation between recommendation and execution.

6.5. The accountability gap underneath

When an analyst acts manually, accountability is usually traceable through tickets, logs, approvals, and procedures. When an agent contributes to an investigation, accountability becomes distributed across the analyst, the SOC manager, the vendor, the model provider, and whoever configured the tool. The National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF) presses organisations to govern, map, measure, and manage AI risks throughout the lifecycle [26]. Applied to SOCs, that means deciding in advance where agentic AI may be used, what autonomy is permitted at each point, who approves high-impact actions, how agent performance is measured, and how agent failures get reviewed. Retrofitting any of that after deployment is harder than getting it right at the start. Two bodies of work are directly relevant here. The explainable AI (XAI) literature provides vocabulary and methods for making automated reasoning legible to the humans held accountable for it [27],

and recent national guidance on agentic AI stresses that organisations should establish ownership, oversight, and the ability to intervene before deployment rather than after [23].

7. Designing for Human-Agent Security Operations

If the analyst is becoming an orchestrator, SOC design has to support orchestration directly rather than inserting AI into existing tools and trusting the workflow to adjust around it. The following design principles address the risks identified in Section 6.

Bounded specialisation is the first. Agents should be built for narrow, well-defined tasks, not for broad claims of SOC autonomy. Phishing evidence collection is a defensible scope. The narrower the scope, the easier it is to test the agent, to govern its use, and to explain its output to a sceptical engineer at 4am, which is the audience that matters. Vendor claims that an agent "handles SOC work" deserve the same scepticism one would apply to a vendor claiming to handle "IT", and for the same reasons. Scope and permission limits should be enforced by the surrounding system, rather than left to the agent's instructions [24].

Evidence-first output is the second, and it is the one most often missing in the products seen. A useful SOC agent shows what it queried, what it found, what it did not find, what assumptions it made, and how confident it is. Narrative conclusions without that trail are worse than useless: they look authoritative and cannot be checked at the speed analysts have to work. Designing for evidence-first output also has a second-order benefit, which is that the analyst can see where the agent looked and adjust their own approach accordingly.

The third is meaningful human approval for high-impact action. Recommendation is fine. Execution that affects user access, production availability, customer data, or legal exposure is not fine without explicit approval, except where the action is genuinely low-risk, reversible, and covered by a policy that was approved in advance rather than retrofitted around the agent's actual behaviour. The interface for these approval gates matters as much as the policy. An approval flow that defaults to "accept" under time pressure is, in practical terms, not a gate. Sadok, Alter and Bednar [28] found a disconnect between formal security policies and everyday practices in small and medium-sized enterprises. Applied here, that finding suggests approval gates should be assessed by how analysts actually use them, including whether they have the time and evidence to challenge a recommendation.

The fourth principle is skill retention. Agentic workflows need to be designed with skill retention in mind, which is harder than it sounds; reducing repetitive burden is easy, but doing so without hollowing out the expertise required for exceptional cases takes deliberate effort. Training modes, manual practice exercises, and visible side-by-side comparison between agent and manual investigation are candidate ways of doing this, whose effectiveness needs testing.

The fifth is auditability. Every agent action should be logged, attributable, and reviewable, tied to a policy approved in advance rather than reconstructed after a failure, and the policies that matter most are the ones that constrain what the agent can do without human sign-off, not the ones that document what it has already done.

Consistent with the socio-technical emphasis on working life and stakeholder involvement [11, 13], analysts and accountable operational owners should help define these arrangements and review them against actual working practice.

8. A Research Agenda

The research agenda follows from these design principles. Their value needs to be tested against the risks in Section 6, rather than inferred from the apparent fluency or autonomy of the agents.

The most tractable is empirical and task-level. Which SOC tasks are genuinely suited to agentic delegation, and which only appear to be? The literature has a habit of treating the SOC as a single unit of analysis, which is convenient but obscures the variation that practitioners care about. A more useful approach is to characterise tasks along several dimensions, complexity, evidence requirements, reversibility, business impact, the degree of human judgement involved, and then ask, task by task, whether agentic AI improves the combined human-agent system or only the metrics that are easiest to collect. This would test bounded specialisation and whether the configurations in Table 1 can be applied consistently to real tasks.

The second concerns situational awareness. Does agentic AI support or erode the analyst's perception, comprehension, and projection of an unfolding incident? This is a candidate for direct empirical work, comparing manual, conversational, and agentic workflows on matched incident scenarios with presentation order counterbalanced. Tests should examine whether visible evidence trails help analysts reconstruct an investigation, recognise missing evidence, and anticipate what might happen next.

The third concerns trust calibration. The goal is neither distrust nor blind acceptance, but appropriate reliance. Which interface cues, confidence reports, evidence presentations, and feedback mechanisms produce calibrated trust over time? Amershi et al. [29] offer a useful starting set of human-AI interaction guidelines for testing the evidence-first principle in Section 7. An experiment could include correct and deliberately incorrect agent conclusions, with and without access to source evidence. Outcomes should include appropriate acceptance and rejection, error detection, decision time, and workload, rather than self-reported trust alone.

A fourth set of questions concerns skills. What new skills do orchestrating analysts actually need, and how should they be trained? The job description is changing faster than the training is. Existing analyst metrics are already contested [3], and agentic AI complicates the picture further. If an analyst supervises ten agents, what does good performance look like, and what fraction of it can be measured without a researcher in the room? The skill-retention principle needs longitudinal evaluation of whether analysts retain unaided investigative capability while working with agents.

A final question is one of governance. What governance models are actually defensible for agentic cyber defence at enterprise scale? Frameworks such as NIST AI RMF, the OWASP LLM Top 10, and the European Union Agency for Cybersecurity (ENISA) guidance on AI cybersecurity challenges [30] supply much of the raw material, but stitching them into a coherent operating model for a SOC is a research and practice problem that is still wide open. For meaningful approval and auditability, the tests are practical: do the controls prevent unauthorised high-impact action, and can a reviewer reconstruct the evidence, agent actions, policies, and human decisions behind an outcome?

The framework remains a conceptual proposal informed by literature and situated practitioner experience. It does not establish comparative performance, error rates, or the

effectiveness of the proposed controls. Its applicability to less-resourced SOCs, analysts with different levels of experience, and different investigative tasks requires empirical examination.

9. Conclusion

Agentic AI is a meaningful development in security operations, but its value is not best understood through replacement narratives. In the practitioner setting described here, experienced analysts had limited patience for generic conversational interfaces that ask them to prompt for information they can already obtain directly. The more durable use of agentic AI is as a set of bounded, specialised, auditable capabilities that perform defined operational tasks under human direction.

The cognitive role of the analyst is changing as a result. Analysts are not disappearing. They are becoming orchestrators, framing investigations and deciding what to delegate. Responsibility for judging the evidence remains with them, as does the need to maintain situational awareness and govern risk on behalf of the business. The shift creates real opportunities to reduce repetitive workload and to improve consistency. It also creates real risks. The framework identifies alternative configurations of this work; the design principles and research agenda set out how their risks and value might be examined. Whether the discipline can move faster than that is, in honesty, an open question. The incentives in security tooling have not historically pointed in that direction, and there is no obvious reason to expect agentic AI to be different.

A socio-technical view is essential because the future of the SOC is not going to be settled by model capability. It is going to be settled by the fit between agents, analysts, workflows, governance, organisational culture, and adversarial behaviour. The work for researchers and practitioners now is to design that fit deliberately, before the systems being deployed today become the legacy that the next generation of analysts must navigate.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly (Grammarly.com) to check grammar and spelling. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. Vielberth, F. Böhm, I. Fichtinger, G. Pernul, Security operations center: A systematic study and open challenges, IEEE Access 8 (2020) 227756–227779. doi:10.1109/ACCESS.2020.3045514.

[2] F. B. Kokulu, A. Soneji, T. Bao, Y. Shoshitaishvili, Z. Zhao, A. Doupé, G.-J. Ahn, Matched and mismatched SOCs: A qualitative study on security operations center issues, in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19, ACM, 2019, pp. 1955–1970. doi:10.1145/3319535.3354239.

[3] E. Agyepong, Y. Cherdantseva, P. Reinecke, P. Burnap, Towards a framework for measuring the performance of a security operations center analyst, in: 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), IEEE, 2020, pp. 1–8. doi:10.1109/CyberSecurity49315.2020.9138872.

[4] S. C. Sundaramurthy, J. Case, T. Truong, L. Zomlot, M. Hoffmann, A tale of three security operation centers, in: Proceedings of the 2014 ACM Workshop on Security Information Workers, ACM, 2014, pp. 43–50. doi:10.1145/2663887.2663904.
[5] S. C. Sundaramurthy, A. G. Bardas, J. Case, X. Ou, M. Wesch, J. McHugh, S. R. Rajagopalan, A human capital model for mitigating security analyst burnout, in: Eleventh Symposium on Usable Privacy and Security (SOUPS 2015), USENIX Association, 2015, pp. 347–359.
[6] S. Tariq, M. Baruwal Chhetri, S. Nepal, C. Paris, Alert fatigue in security operations centres: Research challenges and opportunities, ACM Computing Surveys 57 (9) (2025) 224. doi:10.1145/3723158.
[7] Microsoft, Microsoft unveils Microsoft Security Copilot agents and new protections for AI, Microsoft Security Blog, 2025. URL: https://www.microsoft.com/en-us/security/blog/2025/03/24/microsoft-unveils-microsoft-security-copilot-agents-and-new-protections-for-ai/.
[8] A. Plaat, M. van Duijn, N. van Stein, M. Preuss, P. van der Putten, K. J. Batenburg, Agentic large language models, a survey, Journal of Artificial Intelligence Research 84 (2025) 29. doi:10.1613/jair.1.18675.
[9] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin, W. X. Zhao, Z. Wei, J. Wen, A survey on large language model based autonomous agents, Frontiers of Computer Science 18 (2024) 186345. doi:10.1007/s11704-024-40231-1.
[10] E. L. Trist, K. W. Bamforth, Some social and psychological consequences of the Longwall Method of coal-getting, Human Relations 4 (1951) 3–38. doi:10.1177/001872675100400101.
[11] E. Mumford, The story of socio-technical design: Reflections on its successes, failures and potential, Information Systems Journal 16 (4) (2006) 317–342. doi:10.1111/j.1365-2575.2006.00221.x.
[12] G. Baxter, I. Sommerville, Socio-technical systems: From design methods to systems engineering, Interacting with Computers 23 (2011) 4–17. doi:10.1016/j.intcom.2010.07.003.
[13] I. Sommerville, Designing software for a million users: It's not about the technology, in: Proceedings of the 1st International Workshop on Socio-Technical Perspective in IS Development (STPIS 2015), CEUR Workshop Proceedings 1374, 2015, pp. 1–4. URL: https://ceur-ws.org/Vol-1374/invitedpaper.pdf.
[14] R. Singh, S. Tariq, F. Jalalvand, M. Baruwal Chhetri, S. Nepal, C. Paris, M. Lochner, LLMs in the SOC: An empirical study of human-AI collaboration in security operations centres, arXiv preprint (2025). doi:10.48550/arXiv.2508.18947.
[15] L. Bainbridge, Ironies of automation, Automatica 19 (6) (1983) 775–779. doi:10.1016/0005-1098(83)90046-8.
[16] M. R. Endsley, Toward a theory of situation awareness in dynamic systems, Human Factors 37 (1) (1995) 32–64. doi:10.1518/001872095779049543.
[17] J. D. Lee, K. A. See, Trust in automation: Designing for appropriate reliance, Human Factors 46 (1) (2004) 50–80. doi:10.1518/hfes.46.1.50_30392.
[18] R. K. Yin, Case Study Research and Applications: Design and Methods, 6th ed., Sage, 2018.
[19] E. M. Bender, A. Koller, Climbing towards NLU: On meaning, form, and understanding in the age of data, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 2020, pp. 5185–5198. doi:10.18653/v1/2020.acl-main.463.

[20] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung, Survey of hallucination in natural language generation, ACM Computing Surveys 55 (12) (2023) 248. doi:10.1145/3571730.
[21] R. Sapkota, K. I. Roumeliotis, M. Karkee, AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges, Information Fusion 126 (2026) 103599. doi:10.1016/j.inffus.2025.103599.
[22] N. Kshetri, Transforming cybersecurity with agentic AI to combat emerging cyber threats, Telecommunications Policy 49 (6) (2025) 102976. doi:10.1016/j.telpol.2025.102976.
[23] National Cyber Security Centre, Careful adoption of agentic AI services, Joint guidance co-authored with CISA, NSA, ASD ACSC, the Canadian Centre for Cyber Security and NCSC-NZ, 2026. URL: https://www.ncsc.gov.uk/blogs/thinking-carefully-before-adopting-agentic-ai.
[24] OWASP Foundation, OWASP Top 10 for Large Language Model Applications, OWASP GenAI Security Project, 2025.
[25] P. M. Bednar, C. Welch, Socio-technical perspectives on smart working: Creating meaningful and sustainable systems, Information Systems Frontiers 22 (2) (2020) 281–298. doi:10.1007/s10796-019-09921-1.
[26] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, 2023. doi:10.6028/NIST.AI.100-1.
[27] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Information Fusion 58 (2020) 82–115. doi:10.1016/j.inffus.2019.12.012.
[28] M. Sadok, S. Alter, P. Bednar, It is not my job: Exploring the disconnect between corporate security policies and actual security practices in SMEs, Information and Computer Security 28 (3) (2020) 467–483. doi:10.1108/ICS-01-2019-0010.
[29] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, E. Horvitz, Guidelines for human-AI interaction, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM, 2019, pp. 1–13. doi:10.1145/3290605.3300233.
[30] European Union Agency for Cybersecurity, Artificial intelligence cybersecurity challenges, ENISA, 2020.