Symbiotic Human–AI Collaboration For Augmented Cybersecurity Operations

Reda Yaich[1], Alexandre Balondrade[2], Antoine Sicard[3], Christelle Fouquiau[2], Guillaume Giraud[3], Kahina Amokrane-Ferka[1], Emmanuel Arbaretier[2]

[1] IRT SystemX, [2] Airbus Protect, [3] Réseau de Transport d'Électricité (RTE)

Abstract

Security Operations Centres (SOCs) face mounting cognitive and operational demands as cyber threats increase in scale and complexity. This paper proposes a human-AI collaboration framework to augment SOC effectiveness through cognitive profiling and agentic coordination. We map 29 core SOC functions across three cognitive dimensions, thinking mode, attention level, and coordination context, revealing a concentration of tasks in cognitively saturated zones requiring slow thinking, high attention, or collective decision-making. To address these challenges, we introduce a multi-agent architecture grounded in the Belief–Desire–Intention (BDI) model and structured by an extended VOWEL+U framework that embeds human oversight into agentic ecosystems. We define four AI agent roles, Assistant, Auto-Pilot, Companion, and Operator, aligned with operational autonomy levels to support function-specific delegation. Building on this, we propose a new SOC function: Agent Collaboration and Oversight (F30), reflecting the emerging need for human supervision and configuration of agentic behaviour. Together, these contributions outline a path toward symbiotic human-AI SOCs, which can shift cognitive load, enhance decision quality, and ensure accountable, adaptive cyberdefence.

1 Introduction

The increasing complexity of modern cyber threats and the expanding scope of digital infrastructures have placed unprecedented demands on Security Operations Centres (SOCs). Analysts today must interpret vast and heterogeneous data streams, respond under high temporal and cognitive pressure, and coordinate actions across diverse technical and business domains. While artificial intelligence (AI) has been introduced to assist with detection and response, many deployments remain narrowly focused, poorly aligned with human reasoning, or insufficiently integrated into the broader operational workflow. The result is a growing mismatch between the cognitive capacity of SOC personnel and the complexity of their operational environment. Analysts face what can be described as a cyber cognitive overload, driven by high alert volume, evolving attack vectors, and insufficient decision support. At the same time, emerging AI technologies, particularly those based on large language models, multi-agent systems, and symbolic architectures, offer the potential for deeper collaboration and cognitive augmentation.

This paper argues that the next generation of SOC architectures must move beyond automation toward symbiosis: a design paradigm in which human analysts and AI agents operate as interdependent cognitive partners. Achieving this requires (i) a principled understanding of SOC cognitive functions, (ii) a functional taxonomy of security operations, grounded in human capabilities, (iii) a structured agent architecture that aligns with operational thinking, attention, and collaboration patterns, and (iv) a coordination framework that treats agents and humans as participants in a shared mental model. To achieve this, We present a comprehensive framework for augmenting SOC operations through symbiotic human–agent collaboration. First, we introduce a structured taxonomy of SOC essential functions, organised across three operational domains, *observability, steerability*, and *evolvability*, each capturing a distinct facet of the analyst's responsibilities. This taxonomy forms the foundation for mapping the specific cognitive demands placed on human operators. Building on this taxonomy, we apply established models of cognition, namely, Kahneman's dual-process theory (Kahneman 2011) and Endsley's attention theory(Endsley 2017), to classify each function according to its dominant mode of thinking (fast, slow, or no thinking), its attentional requirement (high, low, or no attention), and its coordination structure (individual, collective, or collaborative). This mapping provides a basis for identifying which tasks are most in need of augmentation, automation, or redesign. To address these cognitive demands, we define four distinct types of AI agents, Assistant, Auto-Pilot, Companion, and Operator, each aligned with specific roles within a five-level autonomy model. These roles draw inspiration from aviation and autonomous driving and are designed to match the varying degrees of human supervision and trust required in SOC settings.

We further integrate the Belief–Desire–Intention (BDI) (Georgeff et al. 1970) agent model to support transparent, context-aware reasoning and to enable shared mental models between humans and machines. This enhances the explainability and adaptability of agent behaviour within collaborative workflows. To ensure scalable and coherent coordination between multiple agents and human operators, we

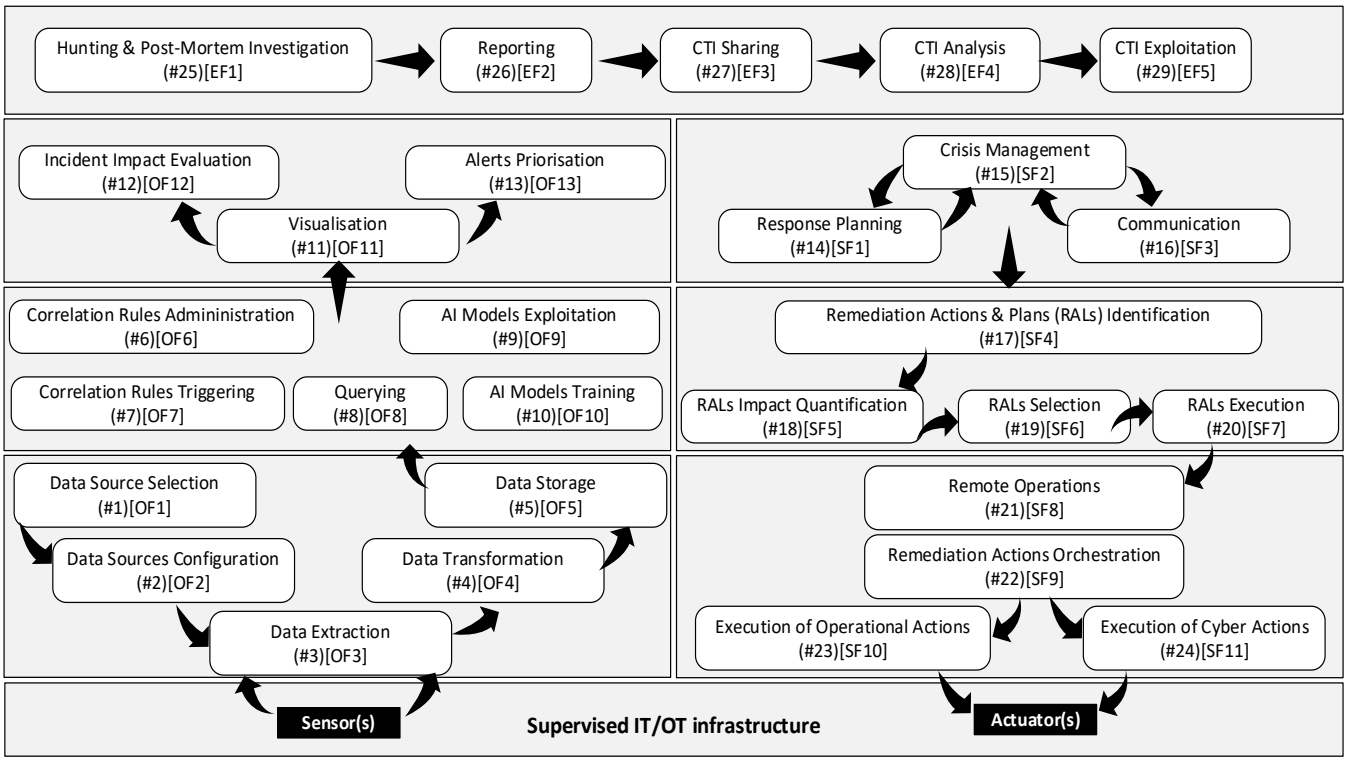

Figure 1: Reference Model of SOC Functions across Observability, Steerability and Evolvability operational domains

extend the established VOWELS (da Silva and Demazeau 2002) multi-agent system framework by introducing a new 'U' dimension for User. The resulting VOWELS+U model captures the structural, environmental, and social components of agentic coordination while embedding human actors as first-class participants in the reasoning loop.

Finally, we formalise a new SOC function, Agent Collaboration and Oversight, capturing the emerging responsibility of analysts to supervise, align, and co-evolve the behaviour of intelligent agents. This function exists both at the operational level (through real-time co-reasoning) and at the architectural level (through agent system design and configuration), thus completing the proposed taxonomy and reframing the SOC as a hybrid cognitive system.

2 The Challenge of Security Operations in Modern SOCs

Modern SOCs face a confluence of challenges that strain both technological infrastructure and human cognition. To understand how human-AI collaboration can meaningfully augment operations, we must first examine the nature of decision-making under pressure and the functional landscape of SOC activities.

2.1 Decision-Making Under Complexity and Pressure

Security Operations Centres (SOCs) operate under intense pressure, characterised by continuous monitoring, high alert volumes, and the need for rapid, high-stakes decision-making. Analysts must detect, assess, and respond to evolving cyber threats with incomplete information, often in real time. This high-cognitive-load environment is further compounded by what we refer to as the seven V's of cybersecurity complexity: Volume (e.g., thousands of alerts per day), Velocity (real-time decision pressure), Variety (diverse IT and OT data sources), Veracity (frequent inaccuracies and false positives), Value (difficulty identifying what matters), Variability (rapid shifts in attacker techniques), and Visualisation (fragmented, cognitively demanding user interfaces). Each factor amplifies the cognitive burden on operators, leading to risks of alert fatigue, tunnel vision, and decision paralysis.

Adding to this cognitive saturation is the increasing unpredictability of the threat landscape. Drawing inspiration from the Johari Window model, we characterise SOC decision contexts along four axes: Known Knowns (routine alerts with known responses), Known Unknowns (recognised but unsolved threat categories), Unknown Knowns (latent knowledge within the organisation but inaccessible in real time), and Unknown Unknowns (entirely novel or adversarially obfuscated attacks). Analysts must navigate these uncertain threat spaces while managing time constraints, incident escalation pathways, and coordination across technical and business domains.

This convergence of cognitive overload and informational uncertainty places fundamental limitations on the human operator. It also exposes the misalignment between current au-

tomation approaches, which often substitute for labour without respecting human mental models, and the real needs of collaborative, adaptive cybersecurity work.

2.2 SOC Functions' Taxonomy

To move beyond simplistic automation and toward meaningful augmentation, we must first understand what SOC analysts actually do, and how they do it. We introduce a taxonomy of 29 core SOC functions, grouped into three interdependent operational domains that capture the lifecycle of cybersecurity operations: Observability, Steerability, and Evolvability. Observability encompasses tasks that transform raw telemetry into situational awareness: selecting data sources, configuring log extraction, performing correlation and anomaly detection, querying event databases, evaluating impact, and prioritising alerts. Steerability refers to decision-making and action-taking functions, including planning and triggering remediation, orchestrating cyber or operational commands, managing remote interventions, and escalating to crisis management if required. Evolvability covers learning and adaptation activities such as conducting post-mortems, writing reports, sharing or consuming Cyber Threat Intelligence (CTI), and updating detection rules or system configurations to avoid recurrence.

Each function is identified with a unique label (e.g., OF1–OF13 for observability, SF1–SF11 for steerability, EF1–EF5 for evolvability), and reflects a cognitively distinct unit of work. These functions were derived from a detailed decomposition of current SOC practices across sectors (e.g., critical infrastructure, Aerospace, Automotive, Healthcare, and defence) and were validated through expert review and empirical literature on security operations.

In summary, SOC augmentation demands more than simple AI-Powered technological tooling. It requires a shift toward symbiotic architectures that relieve attentional burdens, align with human reasoning patterns, and enable scalable, transparent coordination across both technical and organisational lines. The next section introduces such a framework, grounded in multi-agent design principles and structured around complementarity with human operators.

3 Cognitive Analysis of SOC Activities

The 29 functions defined in our SOC taxonomy represent more than a set of operational tasks, they define a cognitive landscape within which human analysts must continuously perceive, interpret, decide, and act. To understand where human-AI collaboration is most needed and how agentic augmentation should be deployed, we analyse these functions along three interdependent dimensions: thinking mode, attentional requirement, and coordination structure. These dimensions are grounded in well-established models of human cognition and team performance and together provide a principled lens for designing symbiotic human-agent systems.

3.1 Thinking Modes: Fast Slow, and No Thinking

Drawing on Kahneman's dual-process theory (Kahneman 2011), we classify each SOC function by its dominant thinking mode. System 1 (fast thinking) involves intuitive, automatic responses based on pattern recognition and prior experience. It supports rapid triage, recognition of familiar attack signatures, and routine decisions that require little conscious effort. In contrast, System 2 (slow thinking) refers to deliberate, effortful reasoning that is activated in novel, ambiguous, or high-stakes scenarios. It underpins analytical tasks such as multi-source correlation, impact assessment, and cross-domain coordination. Finally, no thinking describes fully automated or reflexive tasks that require no conscious cognitive effort from the operator. These are akin to human autonomic functions like breathing or heartbeat, essential, continuous, and performed without deliberation. In the SOC, this includes tasks such as routine data ingestion, low-risk log forwarding, or system-enforced policy checks.

Our mapping shows that most SOC functions fall into the slow-thinking or hybrid zones. Tasks such as selecting remediation plans, assessing incident impact, or analysing CTI demand high cognitive involvement. Even tasks that appear routine often escalate into reflective reasoning due to signal ambiguity or high-stakes consequences. This prevalence of System 2 cognition exposes a key pain point: operators are forced to reason deeply across a wide array of tasks, many of which are ripe for delegation to agents, provided they are transparent, explainable, and context-sensitive.

3.2 Attentional Requirements: Managing the Finite Resource

While thinking mode determines how operators reason, attention determines where and for how long they focus. Drawing from attention theory (Endsley 2017), we classify each function into one of three categories: High Attention: Tasks that require sustained, focused attention (e.g., impact analysis, live coordination, CTI assessment). Low Attention: Tasks that can be interleaved or backgrounded (e.g., alert triage dashboards, rule maintenance). No Attention: Tasks that require no operator engagement, and are ideally fully automated (e.g., data retention policy enforcement, scheduled log extraction). Our analysis reveals a troubling trend: the majority of high-value SOC functions are also high-attention tasks. This includes complex detection queries, threat modelling, incident communication, and CTI exploitation. In contrast, only a handful of tasks—often low-impact or infrastructural—are currently no-attention or truly automatable. This imbalance creates attentional congestion, where critical functions compete for finite cognitive bandwidth, increasing the risk of errors and delays. Moreover, attention is not a purely individual phenomenon in the SOC. In many cases, effective attention must be shared across team members, especially during incident response. Disparities in mental models, domain expertise, or tool visibility often lead to misaligned perception and fragmented situational awareness, particularly when cyber analysts and operational stakeholders must act together.

3.3 Capturing Coordination Complexity: From individual Tasks to Cross-Domain Collaboration

While some SOC functions—such as simple alert dismissal or rule triggering—can be performed independently by a single analyst, many critical tasks demand richer forms of coordination. To capture these layered interactions, we distinguish three coordination contexts: (i) Individual tasks can be completed by one analyst with local authority (e.g., executing a remediation command), (ii) Collective tasks require alignment among multiple SOC actors (e.g., L1-L3 escalations or peer review of impact assessments), and (iii) Collaborative tasks span cybersecurity and business/operational domains (e.g., communicating risk to plant engineers or coordinating crisis response with safety teams).

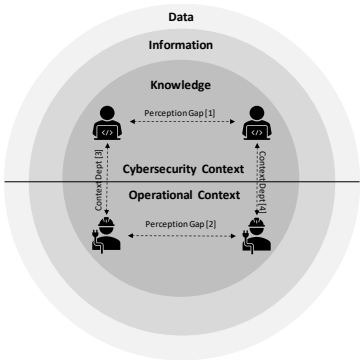

Figure 2: Illustration of Perception Gaps & Context Dept in Collective and Collaborative SOC activities

The figure above (Figure 2), illustrates why cross-domain collaboration is especially challenging. In the cybersecurity context (upper half), two SOC analysts draw on the same raw data but form different interpretations, creating a horizontal Perception Gap [1] between their mental models. In the operational context (lower half), business or safety stakeholders likewise develop a distinct understanding of the same data, yielding a second Perception Gap [2] within that team. The vertical arrows labeled Context Depth (cyber $\rightarrow$ ops) and (ops $\rightarrow$ cyber) capture the fact that SOC and operational teams operate with different sets of contextual information (e.g., technical threat indicators versus business-process constraints). Because each side possesses only a partial context, they must translate and enrich one another's understanding to align on a shared situational picture. These dual perception gaps and asymmetric context depths highlight why simple hand-offs or static dashboards often fail. Agents designed for SOC augmentation must therefore not only automate low-level functions but also actively mediate between disparate knowledge domains, providing shared representations, bilingual explanations, and context-aware summaries that bridge both horizontal and vertical divides.

3.4 From Complexity to Clarity: Mapping Cognitive Load Across SOC Functions

To better understand where cognitive burden accumulates in Security Operations Centre (SOC) workflows, we classified the 29 SOC functions presented before along three interrelated dimensions: thinking mode (Fast, Slow, No thinking), attention level (No, Low, High), and coordination context (Individual, Collective, Collaborative). This tridimensional taxonomy enables a structured diagnosis of human cognitive load and reveals where specific functions are more amenable to automation, augmentation, or human preservation (cf. Figure 3).

The *Individual Context* heatmap shows that some tasks in this category, such as data extraction (OF3), correlation rule triggering (OF7), or Execution of Cyber Actions (SF11)—cluster in the low cognitive demand zones (fast or no thinking, low or no attention). These are strong candidates for automation, where human effort adds little incremental value. However, a few individual tasks—like dashboard visualisation (OF10) and querying (OF9)—still sit within high-attention zones, suggesting they would benefit from interface-level augmentation rather than full autonomy.

In contrast, the *Collective context* heatmap is densely populated in the slow-thinking, high-attention quadrant, representing cognitively intensive activities performed by SOC teams. These include Correlation Rules Administration (OF6), Response Planning (SF1), Remediation Action and Plans Selection (SF6), and Remediation Actions Orchestration (SF9). Their complexity and sensitivity demand careful, deliberative coordination and highlight where tools should support shared reasoning and reduce ambiguity rather than attempting replacement.

The *Collaborative context* heatmap—capturing tasks that involve coordination with business or operational stakeholders, shows the broadest distribution across the grid. Functions like Crisis Management (SF2), communication (SF3), and RALs (Remediation Actions and pLans) Identification and Impaxct quantification (SF4, SF5) span multiple cognitive and attentional categories, reflecting the socio-technical complexity of decision-making across domains. These functions cannot be isolated to automation pipelines; they require systems that support mutual understanding, shared mental models, and cross-domain explanation. When considered together, these heatmaps illustrate a pronounced asymmetry: very few SOC functions occupy the "ideal" quadrant of fast thinking, low attention, and individual execution, the conditions most favourable for automation. Instead, the cognitive landscape of SOCs remains dominated by high-attention and efforts, interdependent decision-making, underscoring the burden placed on analysts and teams to bridge technological and organisational gaps and depts. This multidimensional mapping not only provides a diagnostic lens for cognitive saturation but also forms the foundation for strategic delegation in Human–AI collaboration. It helps identify (i) which tasks are ripe for agent delegation or intelligent automation; (ii) where human cognitive effort must be preserved or enhanced; and (iii) which coordination bottlenecks most impair agility and decision quality.

In summary, SOC augmentation demands more than simple AI-Powered technological tooling. It requires a shift toward symbiotic architectures that relieve attentional burdens, align with human reasoning patterns, and enable scalable, transparent coordination across both technical and organisational lines. The next section introduces such a framework, grounded in multi-agent design principles and structured around complementarity with human operators.

4 A Symbiotic Human–Agent Collaboration for Cybersecurity operations

Drawing from the cognitive characteristics of Security Operations Centre (SOC) work, the cognitive burdens imposed by modern cyber threats, and the architectural affordances of agent-based systems, we propose a symbiotic framework for Human–AI collaboration in cybersecurity operations. This framework recognises that effective augmentation requires more than automation; it requires the design of collaborative agentic systems that can align with, adapt to, and extend human cognitive capabilities.

At the heart of our proposed framework lies a triadic alignment that enables effective human–AI collaboration in SOC environments. First, cognitive alignment ensures that agents support human reasoning by modelling both thinking modes (fast, slow, none) and attention levels (high, low, none), thereby matching the psychological demands of the task at hand. Second, agentic alignment structures the AI system around specialised agent roles, each corresponding to distinct levels of autonomy and complementarity with human capabilities. Finally, coordinative alignment is achieved through a VOWEL+U multi-agent architecture (J. Da Silva and Demazeau 2002), which formalises how agents interact, organise, and interface with human operators across team and organisational boundaries. These three layers of alignment form a cohesive socio-cognitive defence model in which the division of labour between human and machine is not fixed, but dynamic, explainable, and symbiotic.

4.1 Defining Agentic Roles for the Cognitive and Operational Augmentation of SOCs

The strategic integration of AI agents into SOCs demands a careful consideration of how and when decision-making authority should shift from human analysts to machine counterparts. A valuable reference point for conceptualising this shift lies in the progressive autonomy models established in high-reliability domains such as automotive driving, aerospace, and robotics, where human–machine task allocation is structured along defined levels of control and delegation. These domains typically define autonomy along a six-level spectrum (Levels 0 to 5), ranging from full human control to full system autonomy. At Level 0, all decisions and actions are performed manually by humans. Level 1 introduces assistive automation, where the system provides suggestions or alerts, but the human retains full decision authority. Level 2 allows partial execution, with the system carrying out predefined tasks under human supervision. Level 3 corresponds to conditional autonomy, where the system can execute task sequences independently within known conditions but requires human intervention when uncertainties arise. At Level 4, the system achieves high autonomy, operating independently in most scenarios with minimal oversight. Finally, Level 5 represents full autonomy, where the system performs all functions across all conditions without any human input or supervision.

In the cybersecurity domain, such levels are conceptually useful but practically uneven in their applicability. SOC environments deal with incomplete data, evolving threats, and contextual ambiguity—conditions that often defy clean automation thresholds. As such, we adopt a more functionally grounded taxonomy, collapsing the six-level spectrum into four practical levels of autonomy, preceded by Level 0, where no AI is involved and all functions are handled manually by human operators. This simplification serves two key purposes. First, it recognises that not all SOC activities require fine-grained control separation; most functions can be supported by either assistive, automated, collaborative, or delegative agents. Second, it acknowledges that the level of autonomy varies across SOC functions, and even across phases of the same task. For instance, in intrusion detection and prevention, many organisations already use Level 3 systems, such as Intrusion Prevention Systems (IPS) that can automatically isolate compromised endpoints under predefined conditions. Similarly, SOAR (Security Orchestration, Automation and Response) platforms may triage low-priority alerts or execute pre-scripted remediation actions without human intervention. Yet this autonomy is far from universal. For critical or ambiguous incidents, such as assessing the business impact of an attack or coordinating a crisis response, humans remain irreplaceable due to their ability to interpret uncertainty, reconcile conflicting inputs, and weigh trade-offs. Much of the current human workload in SOCs stems from the absence or immaturity of autonomous support in functions that otherwise could be shared or delegated. This heterogeneity necessitates a hybrid model, in which agents vary in autonomy depending on task characteristics, and human operators act as both analysts and orchestrators of the AI systems they oversee.

To make this model operational, we define four core agentic roles, each corresponding to a distinct level of practical autonomy. Assistant Agents (Level 1–2): These agents support the operator with recommendations, contextual information, or summarisation, but never act independently. They are suited to cognitive or attentional relief in low-risk, operator-led tasks. Auto-Pilot Agents (Level 2–3): These agents execute predefined actions under stable conditions, typically in high-volume, low-criticality workflows. While they reduce operational overhead, they operate strictly within scripted boundaries. Companion Agents (Level 3–4): These agents reason alongside the operator, handling tasks that require deliberation, interpretation, and multi-domain coordination. They are collaborative partners, not just executors. Operator Agents (Level 4+): These are high-autonomy agents capable of making decisions and executing actions under constrained delegation. They are applicable in real-time containment, autonomous hunting, or policy enforcement scenarios, where timely and trustable autonomy is critical. These agentic roles serve as practical in-

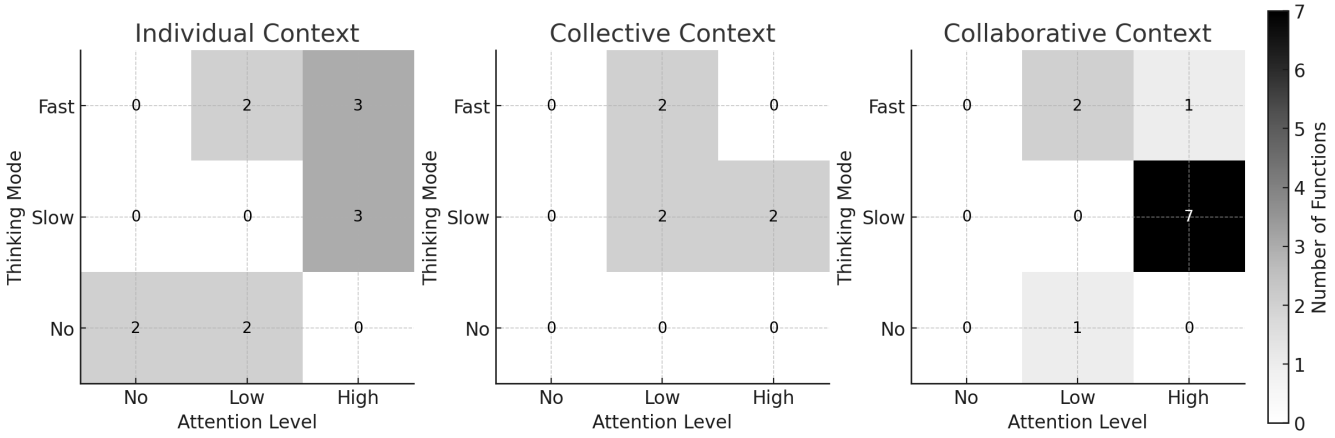

Figure 3: Cognitive load across SOC functions, by Thinking Mode (Fast/Slow/No) and Attention Level (No/Low/High), separated by Operational Context (Individual | Collective | Collaborative).

stantiations of the autonomy spectrum tailored to SOC contexts. They do not replace humans but complement them, shifting selected tasks from high to low attention, from slow to fast thinking, and from distributed coordination to localised execution. This enables a structured, cognitively aligned, and role-specialised augmentation of the modern SOC.

4.2 Building Human-AI Shared Situational Awareness with BDI Mental Models

Security operations are not only technical but deeply cognitive and collaborative. As incidents increase in complexity and cross-functional boundaries, effective cybersecurity decision-making increasingly depends on a team's ability to build and maintain shared situational awareness—a common understanding of what is happening, what it means, and what must be done. To support this, we propose that the design of AI agents in SOCs should be guided by an explicit model of mental representation and deliberation—specifically, the Belief–Desire–Intention (BDI) model (Georgeff et al. 1970).

Originally proposed by Michael Bratman in the context of human practical reasoning and widely adopted in agent-based systems, the BDI model offers a structured cognitive architecture that mirrors human decision-making processes. Its three core components, Beliefs, Desires, and Intentions, are well-suited to modelling how SOC analysts process data, set goals, and decide on actions in uncertain and high-pressure environments. Beliefs represent the agent's current model of the world, including what it infers to be true about the system's status, alert history, threat indicators, and operational constraints. In SOCs, these beliefs would include ongoing telemetry, prior incidents, business-critical systems, and contextual threat intelligence. Desires correspond to possible or preferred system states that the agent may wish to bring about. These can be strategic (e.g., maintaining system uptime, ensuring compliance) or tactical (e.g., isolating a suspicious endpoint). Desires serve to encode priorities aligned with both cybersecurity and business objectives. Intentions are the subset of desires to which the agent commits in the current context, based on available beliefs. They guide the agent's planning and action. Importantly, intentions are context-sensitive and must remain consistent with both the agent's beliefs and organisational rules of engagement.

This BDI architecture is particularly well-suited for supporting shared mental models between humans and agents. By explicitly modelling its beliefs and intentions, and exposing these in natural language or visual form to the human operator, an agent can enable transparent collaboration, improving trust and allowing the human to challenge, override, or adopt proposed goals. This supports both individual situational awareness and its extension to team-level shared awareness, which is essential during collaborative threat response or crisis management. Moreover, BDI agents can be designed to update their beliefs dynamically based on environmental inputs (e.g., new log entries, alert status changes, or external CTI feeds), to generate new desires from threat models or escalation thresholds, and to filter and revise intentions through deliberation. This loop mirrors the human cognitive cycle and allows agents to participate not just in automation, but in sensemaking and anticipatory reasoning. In collaborative SOC scenarios, agents built on BDI principles can also support cross-role coordination. For example, one agent may model the desires of a SOC operator (e.g., mitigating a risk), while another encodes business continuity priorities (e.g., maintaining system availability). Negotiation protocols or arbitration policies can then resolve conflicting intentions, ensuring that actions align with both technical risk and organisational impact. In sum, embedding BDI models into SOC agents provides a robust cognitive scaffold for both local autonomy and inter-agent coordination, while preserving the explainability and malleability needed for human oversight. As SOCs evolve toward multi-agent, multi-role environments, BDI-based mental models will be key to realising symbiotic human–AI cognition, grounded in shared awareness, adaptable reasoning, and operational trust.

While the BDI model provides a high-level architectural scaffold for reasoning and decision-making, it does not prescribe specific implementations of belief formation, desire generation, or intention planning. This modularity makes BDI particularly well-suited for integration with state-of-the-art AI techniques. For instance, beliefs—the agent's representation of world state—can be derived from machine learning classifiers, anomaly detection algorithms, or log-parsing models trained to recognise patterns of compromise. Desires may be shaped by outputs from risk scoring systems, rule-based policy engines, or business priority encoders. Intentions, in turn, can be formed or updated using planning models or large language models (LLMs) capable of scenario simulation, natural language reasoning, and playbook synthesis. In this hybrid approach, LLMs function as deliberative modules within the BDI loop—providing narrative explanations, exploring alternative hypotheses, or negotiating intent under ambiguity, while structured ML components ensure robustness and statistical grounding. This synergy allows agents to operate across symbolic and sub-symbolic domains, enhancing both autonomy and transparency in real-time human–machine collaboration.

4.3 Agentic Coordination with VOWEL+

To enable scalable, explainable, and resilient augmentation of SOC workflows, the integration of AI agents must go beyond the definition of individual roles. A robust coordination architecture is required to govern how agents interact with each other, with the environment, and with human operators. To this end, we extend the VOWEL model, originally proposed by Demazeau (da Silva and Demazeau 2002), by introducing a sixth dimension focused explicitly on user interaction: VOWEL+U. In its original form, the VOWEL model defines four interlocking components, Agents, Environments, Interactions, and Organizations, each modular and independently configurable. Agents are defined by their internal architecture (e.g., BDI or reactive), communication capabilities, and decision-making logic. Environments encapsulate the data sources, system dynamics, and real-time constraints within which agents operate. Interactions define the protocols of communication and coordination, and Organizations formalize roles, reporting hierarchies, policies, and rules.

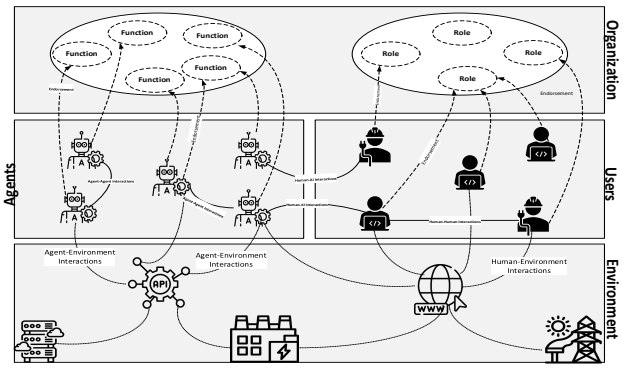

Figure 4: VOWEL+U Conceptual Architecture

Our extension adds a critical fifth dimension: The User. In SOCs, human analysts are not merely observers or consumers of AI output; they are active participants in the system's epistemic and operational cycles. The "U" dimension therefore represents human-centred features, such as transparence, explanation, argumentation, mutual adaptation, that enable trustworthy and meaningful human–agent collaboration. The proposed VOWEL+U architecture allows multiple types of agents (i.e., Assistants, Auto-Pilots, Companions, and Operators) to function within a cohesive multi-agent system. Coordination is achieved not through rigid pipelines but through distributed reasoning and context-aware task allocation. Human oversight is embedded at the organisational and interaction levels, while agent-to-agent communication ensures distributed coverage and responsiveness. As agents grow in autonomy and specialisation, the need for such architecture becomes increasingly critical to ensure performance, resilience, and trust.

4.4 Evolution of SOC Functions: From Operation to Agentic Oversight

As SOCs move toward deeper integration of AI agents, the human role must evolve accordingly. This mirrors the architectural expansion from the traditional VOWEL model, focused on agent, environment, interaction, and organisational design, toward VOWEL+U, which explicitly integrates the human user as a core participant in the multi-agent system. In this paradigm, human analysts are no longer merely recipients of automation outputs but are active collaborators in shaping agent behaviour, validating system decisions, and co-constructing shared cognitive models.

To reflect this shift, we introduce a new SOC function: F30 – Agent Collaboration and Oversight. This function formalises the human–agent interface as an operational capability. It encompasses monitoring agent recommendations, challenging flawed or incomplete inferences, tuning agent parameters based on evolving threat contexts, and intervening in ambiguous or adversarial situations. It also includes high-level tasks such as designing intent models, setting trust thresholds, and calibrating response playbooks across agent teams. Importantly, F30 manifests differently across SOC tiers. At Tier 1, it may involve reviewing and approving agent-proposed triage or containment actions. At Tier 2, it involves adjusting response strategies in concert with agent assessments. At Tier 3 or architectural levels, it includes curating the multi-agent system itself—defining its structure, constraints, and evolution over time. As such, Agent Collaboration and Oversight is not a technical add-on, but a strategic function that ensures human values, situational judgment, and mission alignment remain embedded in automated operations. The integration of F30 signals a broader transformation in SOC practice: from a reactive control room to a symbiotic socio-technical system. It enables agility, resilience, and human-in-the-loop accountability at scale—hallmarks of effective cyberdefence in the AI-augmented era.

5 Related Work

Building effective human–AI collaboration in Security Operations Centres (SOCs) draws on two intertwined streams of research (i) human–AI teaming and complementarity and (ii) use of AI in cybersecurity operations.

Human–AI teaming often disappoints. Pairing with AI isn't a plug-and-play upgrade; success hinges on relative strengths, task type, and a smart division of labor. A meta-analysis of 106 studies finds that, on average, human–AI pairs trail the stronger solo performer, with decision tasks showing outright negative synergy and only creative tasks delivering modest gains (Vaccaro, Almaatouq, and Malone 2024). When AI assumes an embodied, "desk-mate" role rather than a passive tool, team performance can degrade further. (Qin, Lee, and Sajda 2025) show that introducing a human-like AI teammate suppresses human–human communication and disrupts neural synchrony, demonstrating that trust in AI alone fails to ensure effective collaboration under high cognitive load. Yet complementarity remains both a powerful goal and an attainable when AI is considered as an active collaborator. Recent work, formalized Complementary Team Performance (CTP) as the case where a human–AI team outperforms either partner alone (Hemmer et al. 2024). Conceptually, complementarity arises when each partner—human or machine—focuses on tasks that leverage its unique capabilities: humans on nuanced judgment, ethical reasoning, and contextual interpretation; AI on high-volume pattern matching, rapid computation, and consistency . Empirical evidence (Hemmer et al. 2024) confirms this: when tasks are carefully decomposed so that AI handles sub-routines (e.g., low-level triage or data aggregation) while humans perform high-stakes decisions, overall performance can exceed either working alone. (Steyvers et al. 2022) further decomposed CTP into complementarity potential (theoretical gains) and complementarity effect (realized gains), pinpointing information and capability asymmetries as key drivers of synergy (Steyvers et al. 2022). Empirical studies in domains such as radiological diagnosis and financial forecasting (Fragiadakis et al. 2024) show that, without thoughtfully designed interaction mechanisms, human–AI teams rarely exploit their full complementarity. Our work brings these insights into the SOC domain, mapping cybersecurity tasks to cognitive demands and identifying precisely where human–agent teaming can deliver the greatest uplift.

In cybersecurity, the use of AI to augment cyberdefense operators has matured over the last decade. Modern SOCs increasingly leverage AI-Based anomaly-detection models to sift through millions of logs in real time. Likewise, sophisticated SOAR platforms heavily rely on Symbolic AI to automate playbooks for containment and remediation (Kearney et al. 2023). (Baruwal Chhetri et al. 2024) $\mathcal{A}$ framework makes an important step by adapting agent autonomy based on analyst workload, demonstrating a 30% reduction in self-reported fatigue during triage exercises. However, the proposed framework remains narrowly focused on that single workflow and on three defined teaming modes (collaboration, augmentation, automation). It neither explains why or when to switch from one mode to another, nor offers a well defined mechanism for sustaining shared situational awareness. In contrast, our work profiles SOC functions across three cognitive dimensions (thinking mode, attention demand, coordination context), prescribes four finely graded agent roles (Assistant, Auto-Pilot, Companion, Operator) grounded in BDI transparency to enable dynamic mode selection and shared situational awareness across the entire SOC lifecycle.

6 Conclusion & Future Work

This paper has proposed a structured approach to human–AI collaboration in Security Operations Centres, grounded in cognitive theory, task analysis, and multi-agent systems architecture. We introduced a taxonomy of SOC tasks and activities, classified by thinking mode, attention level, and collaboration context. We mapped these to four agentic roles, each corresponding to a specific teaming scope, and showed how they align with levels of autonomy. We extended the VOWEL framework to include human agency (VOWEL+U), creating a foundation for cohesive agent–human collaboration. Our framework enables immediate augmentation of high-friction workflows and longer-term transformation toward adaptive, trustable, and symbiotic cyberdefence teams.

Future work will include prototyping agent capabilities, testing human–agent trust dynamics, and validating the framework in operational SOC environments. As future work, we plan to operationalize and rigorously evaluate our framework through the development of concrete agent prototypes—Assistant, Auto-Pilot, Companion, and Operator—integrated within existing SIEM/SOAR platforms, where we will instrument real-time metrics of cognitive load and uncertainty to drive dynamic autonomy selection.

7 Acknowledgment

This work has been supported by the French government under the "France 2030" program, as part of the SystemX Technological Research Institute R&D Program CYBELIA. The CYBELIA Program is co-funded by Airbus Protect and Réseau de Transport d'Électricité (RTE).

References

Baruwal Chhetri, M.; Tariq, S.; Singh, R.; Jalalvand, F.; Paris, C.; and Nepal, S. 2024. Towards Human-AI Teaming to Mitigate Alert Fatigue in Security Operations Centres. *ACM Transactions on Internet Technology*, 24(3): 1–22.

da Silva, J. L. T.; and Demazeau, Y. 2002. Vowels coordination model. (January 2002): 1129.

Endsley, M. R. 2017. Toward a theory of situation awareness in dynamic systems. *Human Error in Aviation*, 37(March 1995): 217–249.

Fragiadakis, G.; Diou, C.; Kousiouris, G.; and ... 2024. Evaluating Human-AI Collaboration: A Review and Methodological Framework. *arXiv preprint arXiv ...* .

Georgeff, M.; Pell, B.; Pollack, M.; Tambe, M.; and Wooldridge, M. 1970. The Belief-Desire-Intention Model

of Agency. In *Lecture Notes in Computer Science*. ISBN 978-3-540-65713-2.

Hemmer, P.; Schemmer, M.; Kühl, N.; Vössing, M.; and Satzger, G. 2024. Complementarity in Human-AI Collaboration: Concept, Sources, and Evidence. *CoRR*, abs/2404.00029.

Kahneman, D. 2011. *Thinking , Fast and Slow Thinking , Fast and Slow*. New York, NY, US: Farrar, Straus and Giroux. ISBN 9780374275631.

Kearney, P.; Abdelsamea, M.; Schmoor, X.; Shah, F.; and Vickers, I. 2023. Combating Alert Fatigue in the Security Operations Centre. *SSRN Electronic Journal*, 1.

Qin, Y.; Lee, R. T.; and Sajda, P. 2025. Perception of an AI Teammate in an Embodied Control Task Affects Team Performance, Reflected in Human Teammates' Behaviors and Physiological Responses. 1–30.

Steyvers, M.; Tejeda, H.; Kerrigan, G.; and Smyth, P. 2022. Bayesian modeling of human–AI complementarity. *Proceedings of the National Academy of Sciences of the United States of America*, 119(11): 1–7.

Vaccaro, M.; Almaatouq, A.; and Malone, T. 2024. When Are Combinations of Humans and AI Useful? *Nature Human Behaviour*, 8(December).