Agentic AI in Security Operations

Technical Deep Dive | *Approx. five-minute read* | *Event correlation, enrichment, prioritisation and guided response.*

CiBRAI thought leadership series

The SOC has become a relay race

Modern security operations were built from best-of-breed tools. That made sense at the time. A SIEM collected logs, SOAR executed playbooks, CTI added reputation, EDR showed endpoint truth, NDR watched the wire, CSPM watched cloud posture, and case management tried to make the whole thing look tidy. The result is not a lack of information. It is a surplus of disconnected information. Analysts spend too much time moving context between screens, translating one platform's vocabulary into another, and deciding whether five medium-severity alerts are actually one serious incident wearing a false moustache.

The interesting promise of agentic AI is not that it adds another clever panel to the dashboard. We have enough panels. The promise is that it changes the working unit of the SOC from an alert to an investigation. Google's current architecture guidance for agentic security operations describes a multi-agent system that can orchestrate triage and investigation across SIEM, CTI, CSPM and EDR workflows, while preserving human approval for sensitive actions [1]. That is the important shift. The analyst is no longer asked to manually stitch together fragments from six tools before they can even begin to think.

Design principle: agentic AI should reduce the analyst's context switching, not become a new source of context switching.

What an agent actually does

An agent in a SOC should not be treated as a magic analyst in a box. It is better to think of it as a governed reasoning layer. The model is only one part of that layer. Around it sit tool adapters, identity controls, policy boundaries, prompt and retrieval controls, memory, evaluation, audit and a human decision gate. Without those pieces, the agent becomes an enthusiastic intern with production access. Entertaining, perhaps, but not a control objective.

The first job is correlation. Instead of treating a suspicious PowerShell command, an impossible-travel login and an unusual OAuth consent as three unrelated tickets, the agent builds an incident graph. It connects identity, endpoint, network, cloud, vulnerability and business context. Google's Triage and Investigation Agent documentation is revealing here: the system evaluates incoming alerts, executes an investigation plan, and returns structured findings with an explanation of its reasoning [2]. That is much closer to how a good analyst works than how a legacy playbook works.

Correlation, enrichment and prioritisation

Enrichment is where agentic AI becomes more useful than old automation. A fixed playbook follows a predefined path. An agent can decide that one case needs domain registration history, another needs process ancestry, and a third needs cloud control-plane logs. That judgement still has to be bounded. The agent should cite the source of each enrichment step, record the tool call, and separate retrieved evidence from model interpretation. In a mature SOC, provenance is not paperwork. It is how trust survives at speed.

Prioritisation then becomes a defensible decision rather than a louder severity label. Good triage weighs confidence, business impact, exploitability, blast radius and response cost. A low-confidence alert against a crown-jewel identity may outrank a high-confidence event on a disposable lab machine. That is the difference between alert management and operational risk management. The agent's job is to make that trade-off visible, not to hide it behind a mysterious score.

The commercial signals are now hard to ignore. Google said its Triage and Investigation agent processed more than 5 million alerts in the last year and reduced a typical 30-minute manual analysis to about 60 seconds with Gemini [3]. Microsoft positions Security Copilot agents around repetitive, high-volume work that plugs into existing workflows and learns from feedback while keeping teams in control [4]. Separate Microsoft live-operations research found a 30.13 percent reduction in mean time to resolution after three months, while carefully noting that observational data cannot prove causality [5]. Different platforms, same message: speed matters, but only when context and control travel with it.

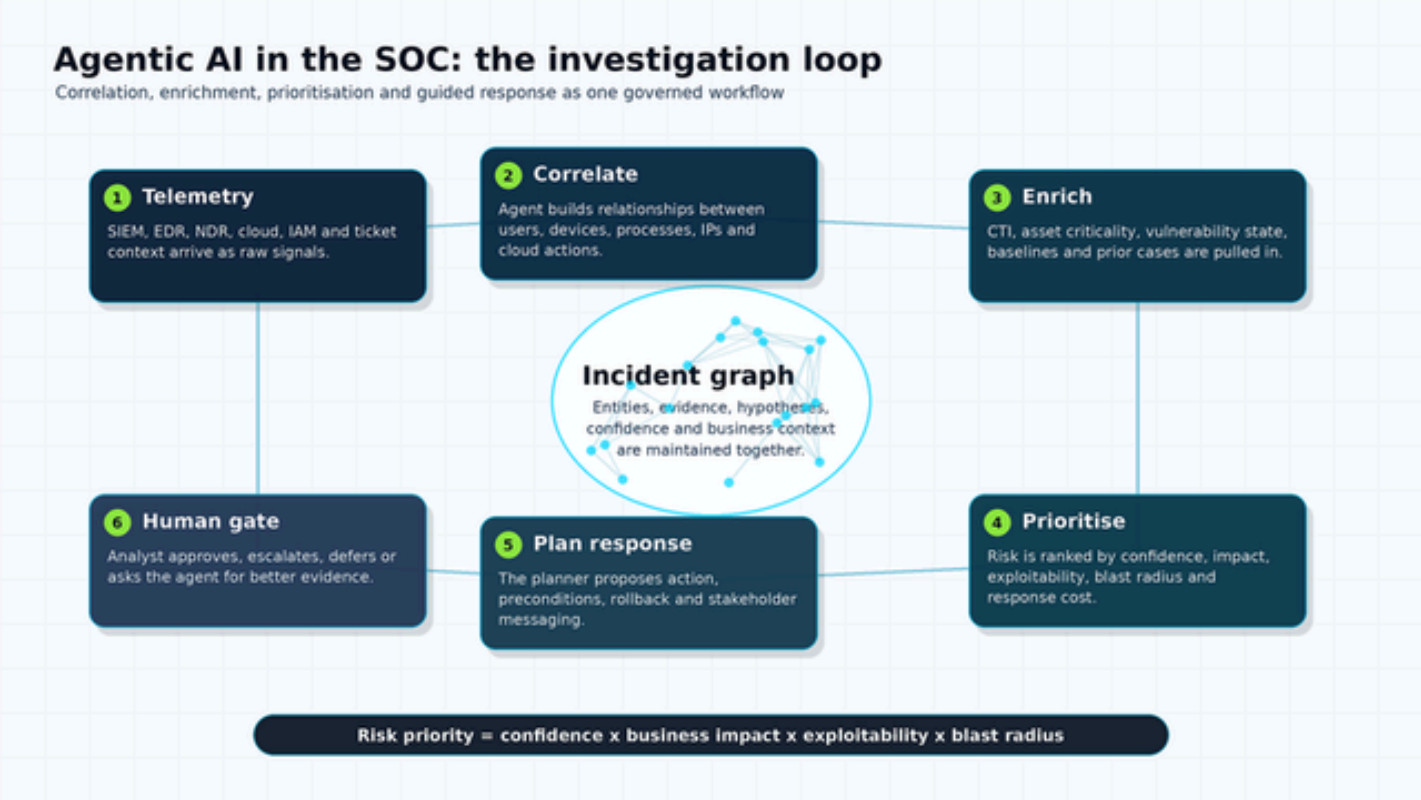

Figure 1. The agentic SOC investigation loop, from telemetry to human-approved response.

Guided response is the real design problem

Response is where agentic AI either becomes operationally valuable or operationally terrifying. A useful agent does not simply say, block the IP and reset the password. It proposes a response package: the evidence, the affected entities, the preconditions, the likely blast radius, the rollback plan, the business impact and the communications note. It should know which actions are auto-approved, which require analyst sign-off, and which should be escalated to incident command. Machine speed is only helpful when the machine is travelling inside the lane markings.

This is also where human expertise becomes more important, not less. Analysts should spend less time collecting screenshots and more time applying judgement. They should ask whether the proposed containment will break a production workflow, whether a user's behaviour has a legitimate explanation, or whether the incident should be handled quietly because it may involve insider risk. The agent gathers evidence and drafts the plan. The human owns the call.

Architecture matters

Under the bonnet, an agentic SOC needs an operating layer rather than another point product. Telemetry needs to be normalised into entities and relationships. Runbooks need to be available as living knowledge, not as dusty PDFs. Tools need strict scopes and identities. Memory needs to capture what happened in the case without leaking sensitive data into the wrong model. The response plane needs reversible actions, approval gates and audit. This is the unglamorous engineering that turns an impressive demo into a trustworthy operating model.

Model choice also matters. The right LLM for a malware reverse-engineering task may not be the right model for a sensitive HR-related insider-threat investigation. Some work belongs on-premises because of data sensitivity, latency or sovereignty. Some work benefits from the latest cloud frontier model because the capability gap is material. The platform should route work by sensitivity, capability, cost and policy. A single-model strategy is neat on a procurement slide, but security operations are rarely that tidy in real life.

NIST's Cyber AI Profile is helpful here because it frames AI as both something to secure and something that can enhance cyber defence, while also recognising AI-enabled attacks [6]. That is the right mental model. The agent is not just a tool inside the SOC. It becomes an actor inside the SOC, which means it needs identity, permissions, monitoring, evaluation and incident response procedures of its own.

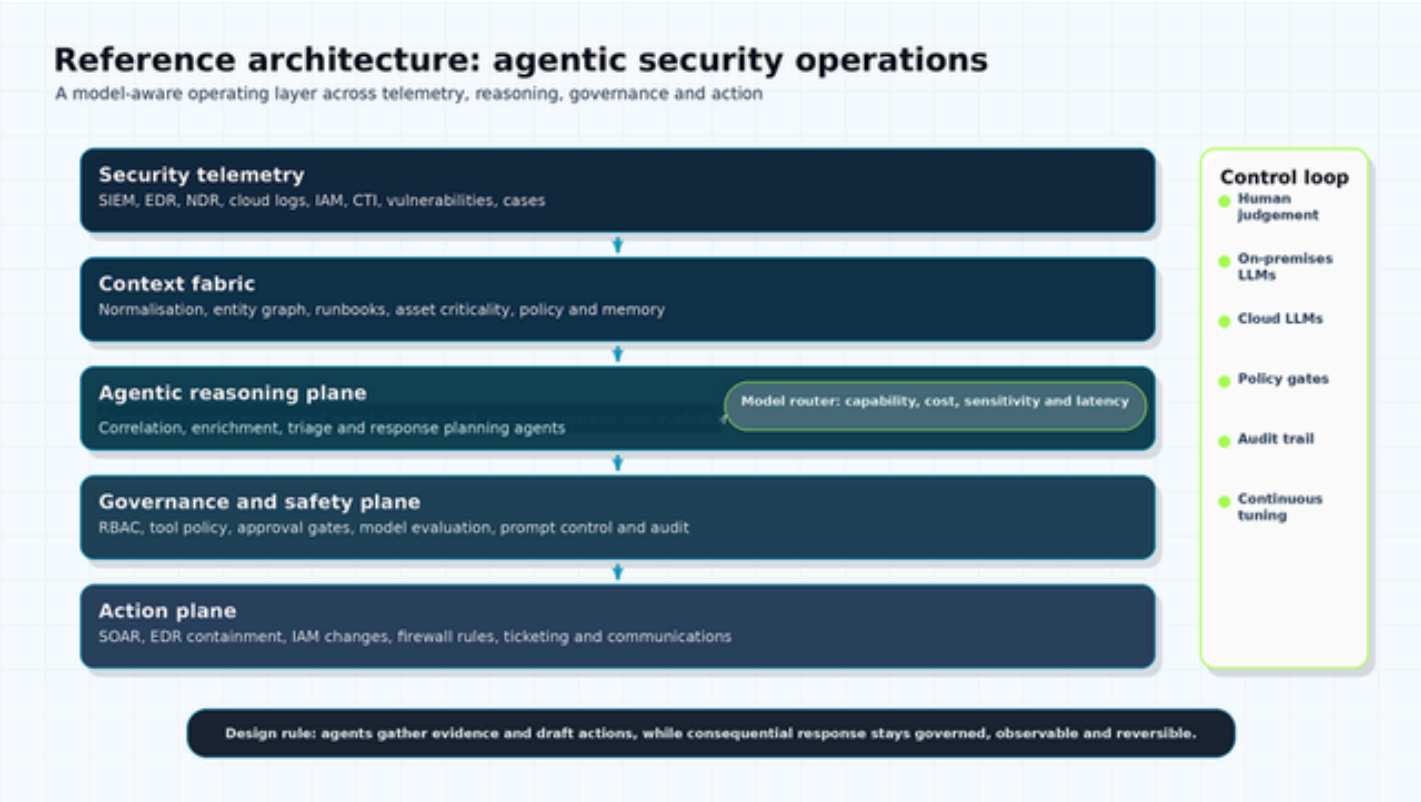

Figure 2. A reference architecture for agentic security operations, with governance as an explicit control plane.

The Mythos moment

Project Glasswing and Claude Mythos Preview make this feel less theoretical. Anthropic describes Project Glasswing as an initiative to secure critical software with early access to frontier AI, backed by launch partners including major cloud, software and security vendors [7]. Whether one is excited or nervous about that direction, the signal is clear: the frontier is moving quickly. Frontier models are no longer being discussed only as office productivity tools. They are being positioned as defensive cyber capability.

The defender's problem is that offensive pressure does not wait for SOC architecture to catch up. AI-augmented discovery, exploit reasoning and campaign planning compress timelines. Defenders therefore need to compress investigation, decision and response without losing control. That is the core tension. Agentic AI must be fast enough to matter, but constrained enough to trust.

Where CiBRAI fits

This is the design philosophy behind the CiBRAI Cybersecurity Operating System. The goal should not be to replace every existing security tool overnight. That would be expensive, unrealistic and, frankly, a bit dramatic. The better goal is to create an operating layer where human operators and agentic AI operators share the same case, the same evidence, the same policy constraints and the same response plan. The analyst should not have to ask which tab knows the truth today.

CiBRAI's integration of on-premises and cloud LLMs matters because the model frontier will not stand still. SOC teams should be able to use the latest defensive capability where it is appropriate, while keeping sensitive

investigations under tighter local control where required. The platform should make that choice routine rather than heroic. It should also preserve the boring things that security leaders care about: auditability, repeatability, approval, cost control and measurable outcomes.

The 10x efficiency ambition is not a claim that an agent makes analysts ten times smarter. Analysts are already smart. It is a claim that the operating model can remove enough context switching, manual enrichment, duplicated triage and brittle handoffs to make the team feel radically less constrained. That is what a Cybersecurity Operating System should do. It should make the difficult parts of security operations clearer, faster and safer, while leaving the final judgement where it belongs: with the human operator. And yes, it should probably save a heroic amount of coffee.

Selected references

[1] Google Cloud. Agentic AI use case: Orchestrate security operations workflows.

[2] Google Cloud. Triage and Investigation Agent.

[3] Google Cloud Blog. Next ’26: Redefining security for the AI era.

[4] Microsoft Learn. Security Copilot agents overview.

[5] Bono, Grana, & Xu. Generative AI and SOC Productivity.

[6] NIST. IR 8596 Cybersecurity Framework Profile for AI.

[7] Anthropic. Project Glasswing.

CiBRAI

CYBER INTELLIGENCE - BEHAVIOURAL RESPONSE