CSO

by **Josh Fruhlinger**
Contributing Writer

Agentic AI in IT security: Where expectations meet reality

Feature

Sep 29, 2025 • 14 mins

Artificial Intelligence | Security

Agentic AI is already changing how security operations centers function, handling repeatable tasks and freeing analysts for higher-level investigations. But trust, pricing, and oversight remain critical issues.

Credit: Summit Art Creations/ Shutterstock

Agentic AI [https://www.csoonline.com/article/3574697/beyond-chatgpt-the-rise-of-agentic-ai-and-its-implications-for-security.html?] has quickly shifted from lab demos to real-world security operations centers (SOC) [https://www.csoonline.com/article/3840447/security-operations-centers-are-fundamental-to-cybersecurity-heres-how-to-build-one.html] deployments. Unlike traditional automation scripts, autonomous software agents are designed to act on signals and execute security workflows intelligently, correlating logs, enriching alerts, and even take first-line containment actions.

For some security leaders, the value of agentic AI in the SOC is obvious: freeing analysts from endless triage and scaling response capacity in the face of overwhelming alert volume. For others, the risks of opaque decision-making, integration complexity, and spiraling costs loom large.

[Related: Agentic AI – Ongoing news and insights [https://www.computerworld.com/article/3843138/agentic-ai-ongoing-coverage-of-its-impact-on-the-enterprise.html]]

To get a clear view of where the technology stands today, we spoke with security executives, product leaders, and researchers who are piloting, deploying, or advising on agentic AI for cybersecurity. Their perspectives highlight what agents do well — and where they stumble — as well as the organizational changes, pricing experiments, and governance models that will shape whether agentic AI becomes a staple of IT security or a short-lived trend.

What agentic AI is (and isn't) good at

Agentic AI [https://www.computerworld.com/article/3617392/what-are-ai-agents-and-why-are-they-now-so-pervasive.html] has carved out a niche performing tasks typically handled by tier one security analysts [https://www.csoonline.com/article/569239/soc-analyst-job-description-salary-and-certification.html]. Instead of simply flagging behavior to be reviewed, agent-based systems "handle first-level tasks, like triaging alerts, correlating signals across tools, and in some cases even taking steps to contain a threat, like isolating an endpoint, allowing analysts to focus on other strategic and more important tasks," says Jonathan Garini [http://linkedin.com/in/garini], CEO and enterprise AI strategist at fifthelement.ai.

Vinod Goje [https://www.linkedin.com/in/vinod-goje/], a data-driven solutions and applied AI expert, notes that in an SOC environment, AI agents operate "much like digital tier-one analysts, sifting through data, gathering contextual information, and even producing detailed reports on their activities." Goje points to practical uses of AI agents in malware examination, script deobfuscation, and coordinating tools.

Itay Glick [https://www.linkedin.com/in/itayglickcyber/], VP of products at OPSWAT, adds that agents "are good at the 'first 15 minutes' with pulling context, checking threat intel, summarizing logs, and proposing actions for review." They also help with exposure management by prioritizing vulnerabilities and with hygiene tasks like spotting stale accounts.

Dipto Chakravarty [https://www.linkedin.com/in/diptochakravarty/], chief product and technology officer at Black Duck, notes that AI agents reduce alert fatigue by clustering alert patterns and correlating them with threat intel feeds, while natural language processing (NLP)-driven tools summarize alerts at scale.

A common theme among many who have used AI agents is that they can free human analysts from "the repeatable grind," as Garini says, so they can concentrate on higher-level exploration and threat hunting. Agent-enabled teams also see "quicker response, more streamlined team structures, and greater resilience in handling the overwhelming number of alerts," according to Goje.

Still, limits remain. Glick warns that without clean data or clear playbooks, agents can chase noise or invent steps. Chakravarty points to problems with false positives and overfitting, while **Prashant Jagwani** [https://www.linkedin.com/in/prashant-jagwani-2675959/?], SVP and global head of cybersecurity services at Mphasis, notes that ambiguous signals or multilayered context can still confound even the best-trained agents. For now, most organizations deploy them to augment rather than replace human analysts.

Integration approaches: Add-on vs. standalone

The first decision regarding AI agents is whether to layer them onto existing platforms or to implement standalone frameworks. The add-on model treats agents as extensions to **security information and event management (SIEM)** [https://www.csoonline.com/article/524286/what-is-siem-security-information-and-event-management-explained.html], **security orchestration, automation and response (SOAR)** [https://www.csoonline.com/article/571201/soar-the-smart-response-to-rising-security-threats.html], or other security tools, providing quick wins with minimal disruption. Standalone frameworks, by contrast, act as independent orchestration layers, offering more flexibility but also requiring heavier governance, integration, and change management.

Ask a question

Ask

Fifthelement.ai's Garini says these systems "are only as good as their interfaces." Out-of-the-box add-ins tend to be most effective when built directly on top of SIEM or SOAR platforms, he says, while standalone frameworks "often require a larger lift for orchestration and governance."

Amit Weigman [https://www.linkedin.com/in/amitweigman/], cybersecurity and AI expert at Checkpoint, points to current industry practice, noting that "Microsoft's Security Copilot ... is helping analysts auto-triage alerts and cut through the noise. CrowdStrike is **doing something similar** [https://www.csoonline.com/article/4057472/crowdstrike-bets-big-on-agentic-ai-with-new-offerings-after-290m-onum-buy.html], and Google's got Gemini-powered agents that can actually investigate alerts end-to-end. That's where we're mostly at right now: bolt-ons and extensions."
Weigman notes that one reason bolt-ons are popular is that replacing or deeply integrating a new SOC platform is a big undertaking: "It can take months of deployment, retraining, and process changes. And all the while, the team is still fighting off live threats."

Fergal Glynn [https://mindgard.ai/authors/fergal-glynn], chief marketing officer and AI Security Advocate of Mindgard, frames the choice as a tradeoff between speed and flexibility. "Add-ins may work well for quick adoption, but they are less dynamic," he says. "Standalone systems give better control, but they may require more setup and maintenance."

OPSWAT's Glick agrees, describing a "rule of thumb" where add-ins fit if most of the data lives in existing SIEM/SOAR pipelines, while a dedicated agent layer works better and "helps cut swivel-chairing" when you need to deal with data that's scattered across IT, OT, **cloud** [https://www.infoworld.com/article/2238873/what-is-cloud-computing.html], and **software-as-a-service (SaaS)** [https://www.infoworld.com/article/2256637/what-is-saas-software-as-a-service-defined.html].
Mphasis's Jagwani notes that most enterprises start with add-ins because they can be layered onto existing investments and provide a controlled test environment. Standalone frameworks, he says, usually come later, when organizations are ready to centralize across hybrid or multicloud estates.

"One lesson we have drawn from client engagements," Jagwani says, "is that many SOCs underestimate integration complexity. It is not only about APIs connecting systems. It is also about aligning the agent's decision logic with existing playbooks and risk tolerances.

Add-in approaches provide a gentler path to that alignment, while standalone orchestration is often a second-phase maturity step."

Governance and organizational change

Agentic AI adoption rarely happens overnight. As Checkpoint's Weigman puts it, "Most security teams aren't swapping out their whole SOC for some shiny new AI system, and one can understand that: It's expensive, and it demands time and human effort, which at the end of the day could appear be too disruptive and costly."
Instead, leaders look for ways to incrementally layer new capabilities without jeopardizing ongoing operations, which makes pilots a common first step.

"My first tip for organizations looking into agentic AI: Start with a smaller use case on a pilot basis, such as phishing response or credential abuse, before scaling up to broader detection and response," says fifthelement.ai's Garini. Targeting contained scenarios helps teams test value and reliability before making wider changes.

Once agents are in place, governance must evolve. OPSWAT's Glick notes that teams don't throw out existing frameworks so much as adapt them: "Existing change-control and segregation-of-duties rules get mapped into the agent flow — e.g., two-person sign-off for destructive actions, risk tiers that decide what's auto vs. ask vs. escalate, and sandboxes to test playbooks before rollout." He adds that agents are now **included in red team testing [https://www.csoonline.com/article/4029862/how-ai-red-teams-find-hidden-flaws-before-attackers-do.html]** through prompt injections and jailbreak attempts. "The ladder stays the same," he says. "It just gets made explicit in the agent's world."
Mphasis's Jagwani sees a similar pattern. Governance and risk controls are extended through "human in the loop" approvals rather than rewritten from scratch. Replacing regulatory frameworks, he argues, won't be practical until AI reaches a more advanced level of general intelligence.

Trust, oversight, and human collaboration

The promise of agentic AI is autonomy, but that quality is also a real barrier to adoption: Many organizations are reluctant to let agents operate freely in production environments.

"An agent designed to carry out a sequence of actions in response to a threat could inadvertently create new risks if misused or deployed inappropriately," says Goje. "For instance, there's potential for unregulated scripts or newly discovered vulnerabilities." As a result, most organizations are unwilling to permit fully autonomous operation without strong safeguards, he says.
Checkpoint's Weigman frames the issue as one of transparency. "AI still feels like a bit of a black box," he says. "With human analysts, mistakes aren't necessarily better, but

managers know the error range and can put a dollar value on it. With AI, it's more like, 'We don't know what we don't know,' and that makes people understandably nervous."

To overcome this challenge, experts emphasize building visibility and accountability into AI workflows. OPSWAT's Glick argues that "you need an audit trail for everything," from prompts and tool calls to outputs and approvals.

Kyle Kurdziolek [https://www.linkedin.com/in/kyle-kurdziolek-175923124/], VP of security at BigID, adds that documentation is essential: "Anything that's regulated needs to be documented, validated, and ultimately auditable. While it's important to see 'what' it did, we also have to construct the rationale as to 'why' it took specific actions." Mphasis's Jagwani emphasizes that regulators in financial services expect "explainability in a form that is auditable. This means outputs cannot simply be black-box recommendations. Teams are beginning to implement layered audit trails, where an agent's decision can be decomposed into inputs, confidence scores, and escalation logic."

Even with more transparency, human collaboration remains critical. Goje suggests treating agents as "collaborative digital allies," and Weigman notes that most adopters are "keeping humans firmly in the loop for higher-risk actions, so the AI can recommend or triage, but the final call sits with an analyst."

Weigman also says that deploying narrow, specialized agents can aid with visibility. "Instead of one giant opaque AI brain, you've got a collection of specialized agents, each with a narrow scope you can monitor and explain," he says.

Training the next generation

If AI agents take on the work of tier-one agents, how do new SOC team members learn the ropes? Vinod Goje is optimistic.

Tier one analyst work has traditionally been the training ground of security careers. The paradox of agentic AI is that while it relieves humans of repetitive triage, it also risks eroding the very "muscle memory" new analysts used to build by grinding through alerts.

But much of the rote triage, such as filtering out obvious false positives, cutting through duplicate alerts, and escalating routine phishing cases, teaches analysts little more than patience. AI excels at handling these menial tasks, allowing human analysts to focus on more complex challenges.

That transforms tier-one from grunt work into guided training ground: Instead of drowning in noise, new analysts study curated, AI-documented cases and learn by interrogating the agent's rationale. So yes, if left unchecked, agentic AI could create a talent-pipeline gap. But used deliberately, it can actually accelerate skill development.

Pricing, value, and program design

Agentic AI capabilities and governance are important, of course, but one of the biggest drivers for adopting agentic AI in security comes down to economics. Security leaders want to know: How much money and time does this save us? The answer is not always straightforward.

"Pricing remains a friction point," says Fifthelement.ai's Garini. "Vendors are playing with usage-based models, but organizations are finding value when they tie spend to analyst hours saved rather than raw compute or API calls."

Mindgard's Glynn notes the **variability in AI pricing models [https://www.cio.com/article/4046457/vendor-pricing-experiments-leave-cios-ai-costs-in-flux.html]** available today. "A charge can be per subscription, per seat, or per alert. Other vendors may offer usage-based plans, too," he says. "Advanced agent systems are usually costly as they have wider impact and opportunity of savings on analyst workloads."
OPSWAT's Itay Glick has seen teams experiment with "per seat, per task, or hybrid" models, but warns that "hidden expenses like storage, API fees, long prompts, and playbook upkeep add up fast." ROI, he argues, should ultimately show up in metrics like faster detection and response, more cases closed per analyst, and fewer junk alerts.

Black Duck's Chakravarty says teams are "spanning a full spectrum of experimentation," with usage-based and hybrid models still evolving. Organizations need to budget not only for the software, but also for the hybrid infrastructure costs of running large models both on-prem and in the cloud, he says.

Mphasis's Jagwani cautions that simplistic pricing metrics often miss the point: "Hidden costs typically show up in areas like retraining models on domain-specific data or building pipelines for clean, structured telemetry." The best ROI, he says, comes when agents are seen as part of a long-term redesign of processes rather than just another plug-in.
BigID's Kurdziolek believes there isn't one right way to measure ROI. "Every organization is different," he says. "Some organizations look at it from an efficiency perspective: How many true positive or false positive events are we seeing from the agent? How many incidents have been raised by the agent? Some look at it from a resource perspective:

How much time are we saving triaging alerts? How often are we double-checking the agent's output, and are we truly gaining time savings?"

As he sees it, the key question is simple: Are agents saving enough time in triage and investigation to let security teams build greater capacity to secure the enterprise?

That question will likely shape the future of agentic AI in cybersecurity. The technology is maturing fast, but its staying power will depend on whether organizations see it as a sustainable way to reimagine how SOCs operate.

Agentic AI brings risks along with promise

While agentic AI offers new possibilities for automation, the technology also brings inherent risk factors, according to a July report from Gartner entitled, "Emerging Tech: The Future of Agentic AI in Enterprise Applications."

1. **Security and compliance:** Adopting agentic AI without fully understanding it can lead to project failure and threaten security and compliance. Because AI agents can operate with autonomy, they must be "secure by design" — that is, they must be built with security in mind from the ground up. Without safeguards in place, agentic systems could take actions that violate legal and regulatory frameworks.
2. **Integration complexity:** Beyond connecting APIs, the challenge of integrating agentic AI in your workflow is aligning decision-making logic with a company's strategy and risk tolerances. This can be particularly complex due to a lack of standardization and interoperability protocols.
3. **Trust and governance:** The "black box" nature of agentic AI creates major barriers. Without transparency and explainability, it is difficult to audit an agent's decisions. Gartner suggests implementing human-in-the-loop controls and audit trails to ensure that agents operate within safe and ethical boundaries and that human remain accountable for high-stakes decisions.
4. **Evolving skills:** Enterprises risk creating a talent-pipeline gap for new employees who traditionally have learned by performing the low-value tasks now delegated to AI agents. Gartner suggests this can be mitigated if new employees are taught to govern and work with agents rather than performing repetitive tasks themselves.

— Dan Muse