Edward Wu, CEO, Dropzone AI
April 24, 2025

Share

Coaching AI agents: Why your next security hire might be an algorithm

Security teams are drowning in alerts. The sheer volume of threats, suspicious activity, and false positives makes it nearly impossible for analysts to investigate everything effectively. Enter agentic AI, capable of completing hundreds of tasks simultaneously without tiring.

Organizations increasingly turn to agentic AI to handle repetitive security tasks, such as alert triage, allowing human analysts to focus on the most critical threats. But while agentic AI may be fast, it isn't infallible. It doesn't inherently understand an organization's unique risk landscape or security priorities.

Like any new hire, an AI agent needs guidance to be effective. It must be adapted, monitored, and refined to align with security policies and operational workflows.

The shift in security operations isn't about replacing human analysts but augmenting them. An AI agent functions as an extra set of hands, sifting through data and identifying potential threats. However, left unchecked, AI can reinforce incorrect assumptions, misinterpret data, or generate misleading conclusions.

AI as a junior analyst: Capable, but needs training

AI agents for cybersecurity aren't like the traditional rule-based systems of the past. Unlike playbook-based automation, which follows a fixed set of instructions, agentic AI is dynamic—it learns, adapts, and refines its approach over time.

Instead of blindly executing predefined rules, it makes decisions based on logical reasoning, analyzing patterns across vast datasets to detect anomalies that might signal a threat. But this flexibility is also a double-edged sword. Without clear direction, AI can misinterpret signals, overlook critical context, or reinforce incorrect assumptions.

In many ways, an AI agent functions like a very smart junior security analyst. But, like a newly hired human analyst who doesn't understand the business context on Day 1, AI lacks inherent awareness of an organization's risk appetite, critical assets, or internal workflows out of the box. It may flag insignificant anomalies while missing genuinely dangerous threats if not given the right context.

Like junior staff, agentic AI needs onboarding, mentorship, and periodic feedback to ensure its decisions align with real-world priorities. Security teams must treat AI not as a fire-and-forget tool but as a developing analyst, one that becomes more effective over time.

Without that guidance, AI without business context is just another unpredictable variable in an already complex security landscape. With that guidance, AI will grow to become a trusted teammate that never forgets details and consistently applies organizational policies, practices, and preferences.

Building an agentic AI onboarding process

Like any new team member, AI agents need onboarding before operating at maximum efficacy. Without proper onboarding, they risk misclassifying threats, generating excessive false positives, or failing to recognize subtle attack patterns. That's why more mature agentic AI systems will ask for access to internal documentation, historical incident logs, or chat histories so the system can study them and adapt to the organization.

Historical security incidents, environmental details, and incident response playbooks serve as training material, helping it recognize threats within an organization's unique security landscape. Alternatively, these details can help the agentic system recognize benign activity. For example, once the system knows what are allowed VPN services or which users are authorized to conduct security testing, it will know to mark some alerts related to those services or activities as benign.

Context is key. AI agents can be easily trained on business-specific risk factors: what assets are most critical, which users have elevated privileges, and what behaviors should be considered safe versus suspect. It should also be configured to follow established investigative workflows, escalation procedures, and reporting structures to ensure its decisions align with security teams' operations. When properly adapted and configured, agentic AI doesn't just process alerts; it makes informed, context-aware decisions that enhance security without adding unnecessary noise.

Coaching AI for continuous improvement

Adapting AI isn't a one-time event, it's an ongoing process. Like any team member, agentic AI deployments improve through experience, feedback, and continuous refinement.

The first step is maintaining human-in-the-loop oversight. Like any responsible manager, security analysts must regularly review AI-generated reports, verify key findings, and refine conclusions when necessary. AI should not operate as a black box, its reasoning must be transparent, allowing human analysts to understand how decisions were reached.

Understand that mistakes will happen. False positives and negatives are inevitable, but how security teams handle them determines whether AI becomes smarter or stagnates. Analysts must intervene, correct AI's reasoning, and feed those insights into the system.

Most agentic AI systems accept overwrites and corrections to refine their reasoning. Over time, this iterative process sharpens AI's ability to reduce noise while improving the system's accuracy in identifying real threats.

When AI is treated as a learning system rather than an infallible oracle, it becomes more than just a tool, it becomes a partner in cybersecurity. The teams that invest in adapting their AI will reduce alert fatigue and build a security operation that continuously evolves to meet new threats.

AI and human collaboration: The future of SOC work

The future of security operations isn't about choosing between AI and human analysts. It's about leveraging both to build a stronger, more scalable SOC. Analysts will transition into supervisory and strategic roles as AI takes on the heavy lifting of investigating every alert, querying logs to answer questions, and putting everything together in reports.

Instead of being bogged down by repetitive alert triage, security professionals will focus on overseeing AI agents and projects that require human judgement and finesse. This shift will improve efficiency and allow security teams to be more proactive, spending more time on threat modeling, attack surface management, and long-term risk reduction.

To embrace this AI-augmented future, organizations must learn to manage and fine-tune agentic AI systems. Analysts must develop new skills in AI oversight, using natural language prompts to adjust AI behaviors, and investigative auditing, ensuring that AI remains a reliable asset rather than a liability.

Future cybersecurity job descriptions will reflect this evolution, demanding expertise in threat detection and the ability to supervise and refine AI-powered security workflows, just like a SOC manager. Those who adapt to this new reality will create a security operation that is not just faster and more efficient but also more resilient.

More about

Artificial intelligence cybersecurity Dropzone AI opinion SOC

Share

+ Featured news

Rack Ruby vulnerability could reveal secrets to attackers (CVE-2025-27610)

Critical Commvault RCE vulnerability fixed, PoC available (CVE-2025-34028)

Understanding 2024 cyber attack trends

eBay CISO on managing long-term cybersecurity planning and ROI

Resources

Enzoic AD Lite Password Audit Report

Report: Fortune 500 employee-linked account exposure

eBook: What does it take to be a full-fledged virtual CISO?

Don't miss

Rack Ruby vulnerability could reveal secrets to attackers (CVE-2025-27610)

Top must-visit companies at RSAC 2025

Critical Commvault RCE vulnerability fixed, PoC available (CVE-2025-34028)

Skyhawk Security brings preemptive cloud app defense to RSAC 2025

Understanding 2024 cyber attack trends

CYBERSECURITY NEWS