NEWS RELEASE

Cisco Reimagines Security for the Agentic Workforce

2026-03-23

With end-to-end security across AI actions, Cisco is helping organizations confidently deploy AI agents at scale

News Summary:

- Cisco extends Zero Trust Access to agents with agent discovery in Cisco Identity Intelligence, agentic Identity and Access Management (IAM) in Duo, and model context protocol (MCP) policy enforcement and adaptive risk protection in Secure Access security service edge (SSE).
- AI Defense: Explorer Edition democratizes AI safety and security by providing developers with self-serve tools to test model and application resilience against attacks and embed robust guardrails into agents before they are deployed.
- Cisco introduces DefenseClaw, an open source secure agent framework that automates security and inventory, with plans to integrate with NVIDIA OpenShell as the sandbox to eliminate manual steps and accelerate secure agent deployment.
- New Splunk AI innovations transform security operations by automating response workflows, enabling teams to outpace sophisticated adversaries at machine speed.

SAN FRANCISCO, March 23, 2026 /PRNewswire/ -- **RSA CONFERENCE 2026 --** Cisco (NASDAQ: CSCO) today announced significant security innovations designed for the agentic AI ecosystem, where software no longer just answers questions—it acts. At RSA Conference 2026, Cisco is introducing solutions to address AI security issues and remove a top barrier to agent adoption. By establishing trusted identities, enforcing strict Zero Trust Access controls, hardening agents before deployment, enforcing guardrails at runtime, and giving security operations center (SOC) teams the tools to stop threats at machine speed, Cisco is building security into the foundation of the emerging AI economy.

"AI agents aren't just making existing work faster; they're a new workforce of co-workers that dramatically expand what organizations can accomplish," said **Jeetu Patel, President and Chief Product Officer at Cisco**. "Projects shelved for lack of resources are now within reach. The only limit is imagination, and security teams are the key to unlocking this opportunity by making the agentic

workforce safe enough to trust."

In a recent Cisco survey of major enterprise customers, 85% reported experimenting with AI agents, but just 5% had moved agentic technology into production.

To unleash the vast potential of AI agents, Cisco is addressing three key pillars to securing the agentic workforce. First: Protecting the world from agents, ensuring they can only act as intended. Second: Protecting agents from the world, ensuring they can't be manipulated or corrupted. Third: Detecting and responding to AI incidents at machine speed and scale.

Protect the world from agents: Establish trust before agents go to work

Like new employees, AI agents need onboarding to establish their identity, understand their function, and map them to an accountable human manager. Yet today, most enterprises are unaware of which agents are running, let alone who is responsible if something goes wrong. Existing SSE tools weren't built to enforce time-bound access for agentic workload identities, nor can they understand context behind agent requests.

According to the **2025 Cisco Talos Year in Review** release today, attackers overwhelmingly targeted a subset of components that directly authenticate users, enforce access decisions, or broker trust between systems. Adversaries' focus on identity will only accelerate with the rise of agentic workloads.

To address these challenges, today **Cisco is extending Zero Trust Access to AI agents**, holding them accountable to a human employee and securing agentic actions. New Duo IAM capabilities integrate with novel MCP policy enforcement and intent-aware monitoring in Cisco Secure Access to enforce strict access control, uniquely helping organizations gain full visibility and governance over their agentic workforce. These capabilities include:

- **Agent Identity Management:** Customers can register agents in Duo IAM and map them to accountable human owners, ensuring every agent has a verified identity and enabling traceability of actions.
- **Agent and Tool Visibility:** Cisco Identity Intelligence discovers agentic and non-human identities to help organizations understand existing AI usage.
- **Strict Access Control:** Agents are assigned fine-grained permissions only for the specific tasks they perform or resources they need for a short duration, with all tool traffic routed through an MCP gateway to eliminate blind spots.

"Organizations are eager to embrace AI, but they need to do so without creating security coverage gaps. Cisco's Zero Trust Access for AI agents gives visibility into agentic identities and restricts access to exactly what's needed," said **Jeremy Nelson, CISO North America, Insight.** "We're excited to bring these capabilities to customers to secure their data while scaling their AI initiatives."

"In this dynamic agentic tech environment, strict access control for AI agents is critical but challenging to enforce consistently with legacy tools designed for human users. This creates uneven enforcement and blind spots, leading to gaps that agents in an agentic world will inevitably exploit," said **Fernando Montenegro, Vice President & Practice Lead, Cybersecurity & Resilience, Futurum**. "Cisco's platform approach is well-positioned to address these challenges by modernizing tooling to ensure consistent, adaptive security for AI agents."

Protect agents from the world: AI Defense safeguards the agentic workforce

As businesses race to deploy AI agents across increasingly complex and distributed environments, Cisco is expanding AI Defense with powerful new tools that help organizations test, trust, and secure their AI agents and the interactions between them.

Traditional scanning tools cannot simulate the real-world threats agents encounter, which are marked by longer conversations and access to tools and resources.

To empower more organizations to meet this challenge head-on, Cisco is democratizing the industry-leading capabilities of AI Defense by launching **Cisco AI Defense: Explorer Edition**. This new self-service solution is built on the same core AI Defense Validation engine trusted by Global 2000 customers. After signing up, users can begin red teaming the AI models and applications that will be deployed into agentic workflows to uncover susceptibility to attacks and measure risk posture before deployment. This toolkit enables AI developers, AppSec teams, and security researchers to build and secure AI agents.

At launch, Cisco AI Defense: Explorer Edition features:

- **Dynamic Agent Red Teaming:** Conduct multi-turn adversarial testing for models and applications that power agentic workflows, with Cisco's bespoke AI red teaming framework.
- **Model and Application Security Testing:** Validate resistance to prompt injection, jailbreaks, and other unsafe outputs.
- **Straightforward Security Reporting:** Get actionable AI security insights, exportable for compliance review.
- **API-First Access:** Tap into CI/CD integration for GitHub Actions, GitLab, Jenkins, and custom pipelines.
- **Team Collaboration:** Invite teammates; upgrade to AI Defense Enterprise for advanced role-based access control (RBAC).

Separately, Cisco is unveiling its **Agent Runtime Software Development Kit (SDK)**, which embeds policy enforcement directly into agent workflows at build time. The Agent Runtime SDK supports major frameworks including AWS Bedrock AgentCore, Google Vertex Agent Builder, Azure AI Foundry, LangChain, and more.

Cisco is also introducing the **LLM Security Leaderboard**, a comprehensive resource for evaluating model risk and susceptibility to adversarial attacks. By providing transparent evaluation signals, this leaderboard contextualizes model performance metrics against evaluations of how models handle malicious prompts, jailbreak attempts, and other manipulation strategies. The tool empowers organizations with a clear, objective understanding of model risk and informs defense-in-depth approaches to AI deployments.

Together, these capabilities let organizations move from pilot to production with confidence: knowing their agents have been tested, benchmarked, and hardened before they ever touch a production system.

Security is a team sport, and Cisco continues to lead with transparency and collaboration. Building on the release of its first open source foundation AI model at last year's RSA Conference, Cisco is today introducing **DefenseClaw** — a secure agent framework designed to eliminate friction between development and security. By integrating a suite of essential open source tools — including Skills Scanner, MCP Scanner, AI BoM, and CodeGuard — DefenseClaw helps ensure that every skill is scanned and sandboxed, every MCP server is verified, and every AI asset is automatically inventoried, enabling developers to deploy secure agents with greater speed and confidence.

DefenseClaw features will directly hook into NVIDIA's OpenShell, extending the ongoing collaboration to provide robust, automated security at the runtime level. By consolidating these capabilities into a single framework, Cisco eliminates the need for manual security steps or separate tool installations, allowing organizations to maintain zero-trust integrity while scaling agentic workforces.

Detect and respond at machine speed: Empowering the agentic SOC

AI technologies are a double-edged sword. As the latest Talos Year in Review report shows, vulnerabilities like React2Shell have seen near instant and automated exploitation, likely fueled by agentic AI being used to build new exploit kits.

The same AI agents posing new security challenges can also be the most powerful tool in a defender's arsenal. Today's SOC analysts are overwhelmed by alert fatigue and fragmented data, spending more time on research than response.

Splunk, part of Cisco's security portfolio, has already moved to embed AI capabilities into key SOC workflows. Today, it is further evolving the SOC from reactive to proactive with:

- **Exposure Analytics:** Now integrated into Splunk Enterprise Security by default, this provides a continuously updated inventory of all assets and users. It delivers real-time risk scoring and relationship mapping, providing total visibility using data that organizations are already ingesting.
- **Detection Studio:** A unified workspace that streamlines the entire detection engineering lifecycle — planning, building, testing, deploying, and monitoring detections. It automatically maps detection coverage against the MITRE ATT&CK framework to identify and close gaps with precision.
- **Federated Search:** A unified search that allows SOC analysts to uncover and correlate data across multiple environments, reducing costs and accelerating investigations.
- **The Agentic SOC Expansion:** Specialized AI agents — including the Detection Builder Agent, Standard Operating Procedures (SOP) Agent, Triage Agent, Malware Threat Reversing Agent, Guided Response Agent and Automation Builder Agent — move beyond data surfacing to active evaluation and execution. By automating security workflows, security tasks shift from a bottleneck to an accelerator, enabling the SOC to move at machine speed and scale.

"The evolution of the security operations center from reactive to proactive is now a necessity in today's threat landscape. By introducing specialized AI agents, Cisco is empowering analysts to move beyond manual triage and prioritize the most important threats quickly," said **Ryan Morris, President, Blackwood**. "This is exactly the innovation required to help security teams stay ahead of constantly increasing and evolving SOC workloads."

Detection Studio and Malware Threat Reversing Agent are generally available. Exposure Analytics, SOP Agent and Federated Search are expected to launch in April and May. Automation Builder Agent and Triage Agent are expected to launch in June. Detection Builder Agent and Guided Response Agent target June 2026 for prerelease testing.

For more information, visit cisco.com/go/security.

Additional Resources:

- Blog: Reimagining Security for the Agentic Workforce
- Blog: Securing Agentic AI: How Cisco Brings Zero Trust to Your New Digital Workforce
- Blog: Introducing Duo Agentic Identity

- Blog: Cisco AI Defense: Explorer Edition Brings Agentic AI Red Teaming to Builders
- Blog: Introducing the Cisco LLM Security Leaderboard: Bringing Transparency to AI Security
- Blog: Cisco Announces DefenseClaw
- Blog: The Evolution of the SOC: Moving from Reactive to Agentic with Enterprise Security at RSAC 2026
- Blog: 2025 Talos Year in Review: Speed, scale, and staying power

About Cisco
Cisco (NASDAQ: CSCO) is the worldwide technology leader that is revolutionizing the way organizations connect and protect in the AI era. For more than 40 years, Cisco has securely connected the world. With its industry leading AI-powered solutions and services, Cisco enables its customers, partners and communities to unlock innovation, enhance productivity and strengthen digital resilience. With purpose at its core, Cisco remains committed to creating a more connected and inclusive future for all. Discover more on The Newsroom and follow us on X at @Cisco.

Disclaimer: The timeline for the release of some products, features and integrations is subject to change given ongoing evolution in development and innovation.

View original content to download multimedia:https://www.prnewswire.com/news-releases/cisco-reimagines-security-for-the-agentic-workforce-302721788.html

SOURCE Cisco Systems, Inc.