



NEWS RELEASE

Cisco Secure AI Factory with NVIDIA Unlocks Enterprise Data for Agentic AI

2025-09-04

Cisco Secure AI Factory with NVIDIA to offer a fully integrated solution with VAST Data, using the NVIDIA AI Data Platform reference design, to accelerate RAG pipelines and enable agentic AI at scale

News Summary:

- Cisco to offer a validated solution with VAST Data to enable faster data extraction and retrieval for agentic AI.
- Cisco AI PODs leverage VAST InsightEngine to offer an NVIDIA AI Data Platform solution, providing real-time data pipelines for the Secure AI Factory.
- Cisco Secure AI Factory with NVIDIA will extend its capabilities, providing a validated architecture that addresses a variety of use cases to speed enterprise AI adoption and keep every token secure with AI Defense

SAN JOSE, Calif., Sept. 4, 2025 /PRNewswire/ -- Cisco (NASDAQ: CSCO) today unveiled a solution for building AI infrastructure designed to support workload data fabrics, further enabling enterprises to securely use their data for agentic AI at enterprise-scale. With this new solution, the [Cisco Secure AI Factory with NVIDIA](#) expands to new use cases, including the acceleration of retrieval-augmented generation (RAG) pipelines with faster data extraction and retrieval. This new capability ensures AI agents have instant, secure access to the data they need, when they need it.

[Cisco AI PODs](#), the AI Infrastructure building blocks of the Secure AI Factory, are now available with [VAST InsightEngine](#), a core capability of [VAST Data AI OS](#). These AI PODs deliver a fully integrated solution using the [NVIDIA AI Data Platform](#) reference design to transform raw data into AI-ready datasets. Within the AI PODs, the Cisco UCS server portfolio with [NVIDIA RTX PRO 6000 Blackwell Server Edition GPUs](#) together provide exceptional performance for next-generation AI applications. RTX PRO Servers from Cisco are some of the first systems to deliver the NVIDIA AI Data Platform reference design.

NVIDIA accelerated computing and AI software ensures low-latency model interaction, and Cisco's high-performance ethernet networking connects compute and data seamlessly. This unified solution enables AI agents to operate with near-real-time business insights, backed by the security, governance and flexibility of the Cisco Secure AI Factory with NVIDIA architecture. Through this architecture, customers can gain unique visibility into their environments with Splunk, and apply safety and security guardrails to every token using Cisco AI Defense.

"Agentic AI has the potential to unlock the value of AI for enterprises around the world. Moving beyond chatbots to agents that can help solve true business challenges is revolutionary, but only if enterprises can effectively leverage the right data at the right times. Cisco, NVIDIA and VAST are working together to give customers a simple path to unlocking the value of their data," said Jeremy Foster, senior vice president and general manager, Cisco Compute. "We are designing the architecture for how the enterprise will build the next generation of AI factories."

"The next wave of agentic AI will be fueled by enterprise data, enabling agents to tap into business knowledge during inference for precise, up-to-date insights," said Justin Boitano, vice president, Enterprise AI at NVIDIA. "Bringing together Cisco Secure AI Factory with NVIDIA and VASTInsightEngine creates an integrated platform for running powerful AI agents at scale."

"By integrating the VAST Data InsightEngine into the Cisco Secure AI Factory with NVIDIA, we're giving enterprises the first integrated design for RAG acceleration at scale," said John Mao, vice president of strategic alliances at VAST Data. "This collaboration with Cisco and NVIDIA represents a major milestone in the evolution of enterprise AI. The integration of the VAST InsightEngine into the Secure AI Factory architecture sets the stage for a new era where intelligent agents can operate securely, collaboratively, and at unprecedented scale."

AN ARCHITECTURE FOR ENTERPRISE AGENTIC AI

Agentic AI workloads place unique demands on IT infrastructure. Enterprises across industries are looking to deploy AI agents that can communicate with knowledge workers and other AI agents to solve complex challenges. However, this requires support for workload data fabrics that remove data bottlenecks and lower latency, so agents have access to the right data, while providing the security and governance necessary to ensure organizations stay safe.

The new capabilities unveiled today offer customers a secure AI infrastructure solution for fast data extraction and retrieval to unlock agentic AI use cases. VAST Data will be the first vendor to integrate with Cisco AI PODs to offer enterprise customers an NVIDIA AI Data Platform reference design. Customers can now experience:

- **Faster time to insights** by reducing RAG pipeline latency from minutes to seconds for near-real-time AI responses.
- **Agentic AI at enterprise scale** by enabling AI agents to operate continuously, learn dynamically and deliver contextualized business outcomes. The high throughput of data unlocks multi-step reasoning, and the architecture is designed for scale by supporting multiple agents and workloads simultaneously.
- **Security and governance are at the core**, designed to protect sensitive data while also accelerating AI innovation. With role-based access control and compliance and audit readiness, enterprises can trust their infrastructure to keep sensitive information safe.

Cisco AI PODs with VAST InsightEngine, offering an NVIDIA AI Data Platform solution, are orderable from Cisco now. The AI POD designed for RAG acceleration with NVIDIA and VAST is the first in a series of AI services PODs built to support the growing number of use cases in the enterprise.

Additional Resources:

- For more information on the Cisco Secure AI Factory with NVIDIA, [click here](#)
- VAST's Shared Everything Podcast: [A Blueprint for Enterprise Agents with Cisco, NVIDIA and VAST](#)
- VAST Data Blog: [Powering the Cisco Secure AI Factory with NVIDIA: VAST InsightEngine](#)
- VAST Webinar: [Unlock Your Data: Powering Secure, Enterprise-Scale NVIDIA NIMs with Cisco and VAST](#)

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](#).

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at <http://www.cisco.com/go/trademarks>. Third-party trademarks mentioned are the property of their respective owners. The use of the word 'partner' does not imply a partnership relationship between Cisco and any other company.

View original content to download multimedia: <https://www.prnewswire.com/news-releases/cisco-secure-ai-factory-with-nvidia-unlocks-enterprise-data-for-agentic-ai-302546076.html>

SOURCE Cisco Systems, Inc.