Cisco Secure AI Factory with NVIDIA

Key messages

Cisco Secure AI Factory with NVIDIA (Secure AI Factory) enables enterprises to accelerate the adoption of AI use cases, unlocking differentiated outcomes with confidence and unprecedented security. This scalable, high-performance, and secure AI infrastructure is purposefully designed to streamline the development, deployment, and protection of AI workloads.

Based on Cisco and NVIDIA's deep understanding of the unique path enterprises take in their AI journey, Secure AI Factory offers both a vertically integrated solution and the freedom to deploy a modular solution with networking, compute, storage, and security components tailored to each organization's unique needs.

What sets Secure AI Factory apart from AI infrastructure solutions from competitors is its foundational emphasis on security and Cisco's market leading high-performance networking that enterprises have trusted for 40 years. This enables enterprises with exceptional flexibility, enhanced security, and superior performance, equipping them to navigate the evolving landscape of AI with confidence.

Definition

Q: What is Cisco Secure AI Factory with NVIDIA?

A: The foundation of Cisco Secure AI Factory with NVIDIA is NVIDIA Enterprise Reference Architecture (ERA), or AI Factory, announced as part of our expanded partnership on February 25, 2025. The architecture consists of processors, servers, networking devices, storage, management, and software in a validated design that delivers a choice of vertically integrated or modular AI solutions to customers.

Cisco Secure AI Factory with NVIDIA builds on this reference architecture and augments it with key capabilities within Cisco's broad security portfolio, delivering a higher degree of protection than available anywhere else.

Data is the currency upon which AI and businesses run. Comprehensive security solutions are essential to address the unique security challenges at the convergence of AI and data. Traditional firewalls are critically important to protecting the data at the heart of AI. But AI requires more. It involves security solutions protecting each of the layers of the AI ecosystem—from data to models.

Q: How is this related to the partnership between Cisco and NVIDIA announced on February 25, 2025?

A: On February 25, 2025, Cisco and NVIDIA announced an expanded partnership to accelerate AI adoption in the enterprise. As part of this partnership:

- **Cisco will develop data center switches with the NVIDIA Spectrum Ethernet platform.** This open ecosystem approach will provide customers with more choices and flexibility. Organizations can standardize on the NVIDIA Spectrum-X networking platform with both Cisco® and NVIDIA switch silicon–based architectures, bringing the industry-leading technologies from both companies under a single management fabric.
- **Cisco will collaborate with NVIDIA to create and validate NVIDIA Cloud Partner (NCP) and Enterprise Reference Architectures based on NVIDIA Spectrum-X** with Cisco Silicon One®, Cisco Nexus® Hyperfabric AI, Cisco UCS® Compute, Cisco Optical Networking and other Cisco technologies.

Q: How does Cisco Secure AI Factory differ from NVIDIA AI Factory?

A: The expanded partnership between Cisco and NVIDIA announced in late February lays the foundation and framework of Cisco Secure AI Factory. By integrating Cisco security solutions, Cisco Secure AI Factory extends this framework and provides the essential security capabilities needed to address the unique security demands of AI.

Cisco Secure AI Factory with NVIDIA delivers a full-stack and validated architecture with built-in security, industry-leading networking, and the NVIDIA Enterprise AI software platform. Whether you require a vertically integrated solution or a modular AI architecture, Secure AI Factory gives you a choice that aligns with your needs and delivers the most advanced and secure infrastructure, streamlining and expediting the development and deployment of next-generation AI applications.

Q: How does Cisco Secure AI Factory differ from other AI Factory partnerships that NVIDIA has with Dell and HPE?

A: Cisco Secure AI Factory differs in two key areas. First, Cisco Secure AI Factory builds upon our expanded partnership with NVIDIA to give customers a networking choice between Cisco Nexus switches with Silicon One or Cisco switches with NVIDIA Spectrum-X silicon—all managed by Cisco Nexus Dashboard or Nexus Hyperfabric AI. Dell, HPE, and other OEMs in the market that build AI Factories with NVIDIA don't provide this flexibility. Second, Cisco Secure AI Factory is the only available solution that makes AI security foundational at every level in the solution stack—from AI models, frameworks, applications/workloads, network, and storage to compute.

Components

Q: How do Cisco security products integrate into the factory?

A: What sets Secure AI Factory apart from AI infrastructure solutions from competitors such as Dell, HPE, Lenovo, etc., is its foundational emphasis on network choice and security. The Secure AI Factory inserts Cisco security solutions directly into this solution—not simply bolting them on. With the choice of multiple deployment options, the integration of those solution components will be determined by each customer's needs and the management and data requirements of Cisco security products. The security integrations will depend on the selected Cisco security option. Some products, such as Cisco AI Defense, are delivered through SaaS and are more easily integrated. Other solutions, such as Cisco Hypershield, may require additional infrastructure, such as dedicated processing units or appliances.

Q: What products will be included in Cisco Secure AI Factory with NVIDIA?

A: Under the hood of Secure AI Factory, there are two pillars—the underlying reference architecture and the added security solutions. The AI Factory reference architecture may include the following categories of products and solutions:

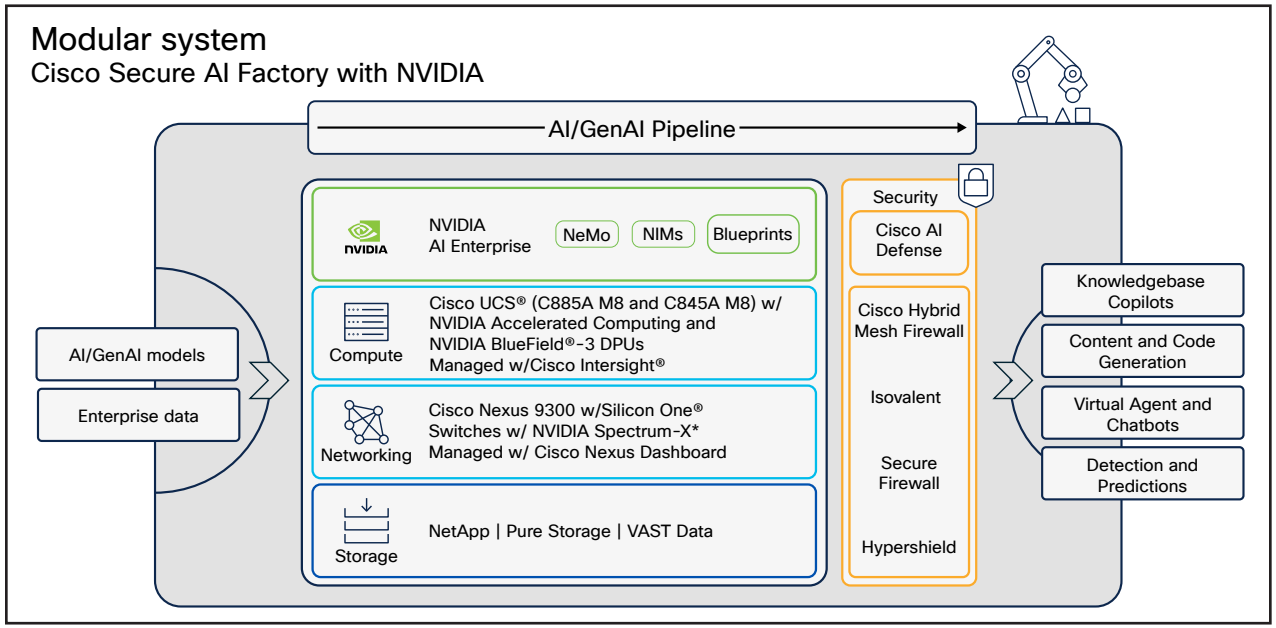

Figure 1. Modular Solution Option

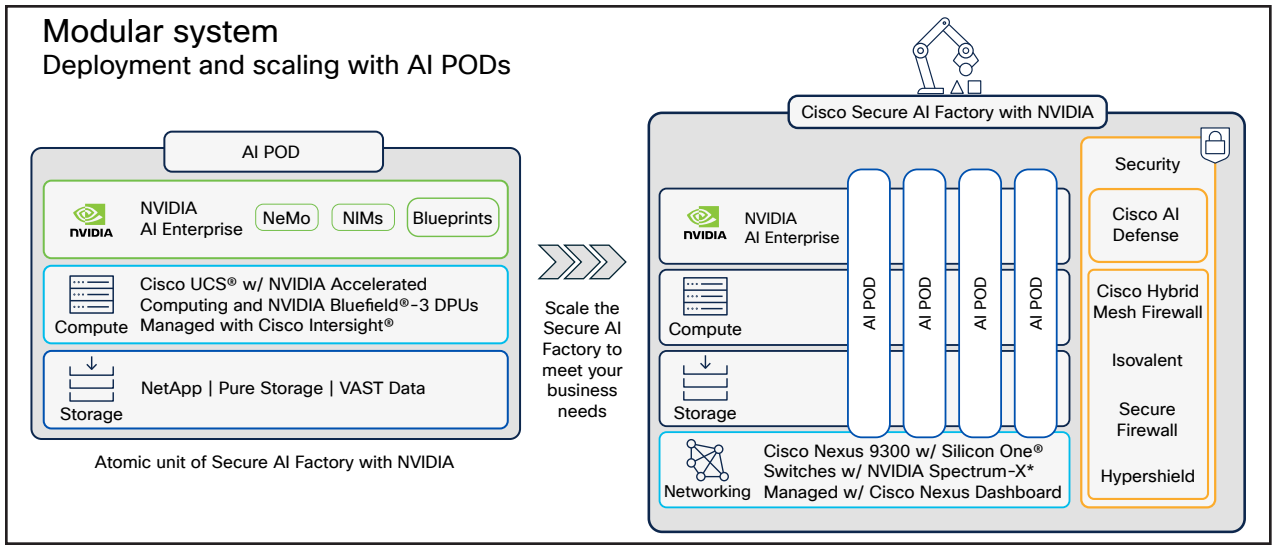

Figure 2. Modular Solution Option with Cisco AI PODs

The framework of this solution provides two deployment tracks—a vertically integrated solution and a modular solution. As a result, you have greater choice to select products according to your needs and preferences.

For instance, should you prefer an easy-to-deploy, cloud-based solution, Cisco Nexus Hyperfabric AI and Cisco 6000 Switches would be an excellent, vertically integrated option for networking. For customers who prefer on-premises infrastructure, Cisco Nexus 9000 Series Switches and Nexus Dashboard would be the preferred solution. In addition, Cisco AI PODs represent an atomic unit that can be used to build the solution, providing validated designs that incorporate NVIDIA processors and third-party storage, virtualization, and containers.

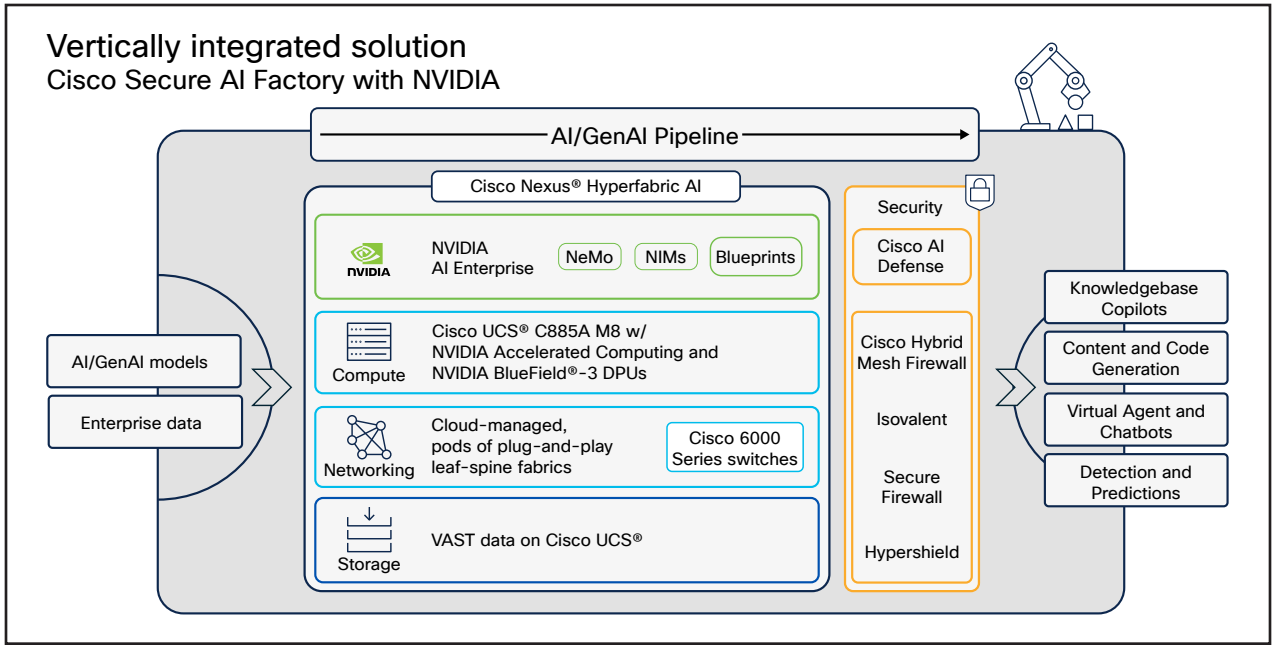

Figure 3. Vertically Integrated Solution Option

The Secure AI Factory includes security at all layers, including:

- **Securing the infrastructure:** Cisco Hybrid Mesh Firewall provides unified security management and consistent policy across multiple enforcement points, including network switches, traditional firewalls, and workload agents. This integrated approach ensures pervasive and consistent security, ranging from deep packet inspection to wide infrastructure coverage, detecting, blocking and containing adversaries. Cisco Hypershield (part of Hybrid Mesh Firewall) will, in the future, extend pervasive security enforcement to include management of NVIDIA's Bluefield 3 for enabling AI Cluster perimeter firewalls.
- **Securing the Workload:** Cisco Hypershield prevents adversary lateral movement and proactive vulnerability mitigation without the need for patching, all from a single management interface. By monitoring and controlling process executions, file access, and network activities, Hypershield delivers deep visibility and surgical runtime enforcement within AI workloads. Future enhancements will further strengthen workload protection through integration with NVIDIA Bluefield-3's DOCA AppShield for intrusion detection in AI-focused virtual machines and containers.
- **Securing the AI application:** Cisco AI Defense empowers security and AI teams with comprehensive tools for robust testing and runtime security of generative AI applications. Utilizing algorithmic red teaming techniques, AI Defense evaluates generative AI models applications against diverse security (data privacy, prompt injections, etc.) and safety (e.g. toxic behavior) risks without requiring application modifications. Additionally, AI Defense applies runtime controls to ensure applications comply with leading frameworks, including OPSWAT LLM and MITRE ATLAS.

Cisco and NVIDIA each bring a unique understanding of customer AI infrastructure needs, and by combining their insights, can offer flexible deployment models alongside proven reference architectures. Cisco's Secure AI Factory with NVIDIA will offer enterprises scalable, high-performance AI infrastructure, that supports customers at any stage of their journey and embeds security throughout.

Again, because of the modular nature of the Secure AI Factory, integrated solutions will be aligned to your requirements, including cloud or on-premises management options.

Q: How will Cisco Secure AI Factory with NVIDIA be managed?

A: Management of Secure AI Factory will ensure the optimal mix of flexibility and performance necessary to support a wide range of AI initiatives. Cisco Intersight will manage server operations. NVIDIA AI Enterprise software will support AI/ML tools for data scientists. And Cisco Nexus Hyperfabric AI will manage the Cisco networking, building on Cisco's previous work with NVIDIA. The intuitive interface of the solution will make it easy to create and manage AI fabrics. The management capabilities will include comprehensive control of the fabric and will ingest NVIDIA proprietary insights directly into Silicon One for greater visibility of workload connectivity. The integration across Cisco and NVIDIA products is expected to continue to tighten, and we plan to offer Nexus Dashboard as a management platform. Management of the security solutions will be through the Cisco Security Cloud Control, providing access to AI-relevant security solutions via a single dashboard.

Q: When will the Secure AI Factory reference architecture be available?

A: Specific timelines are being worked out.

C67-5040751-00 03/25