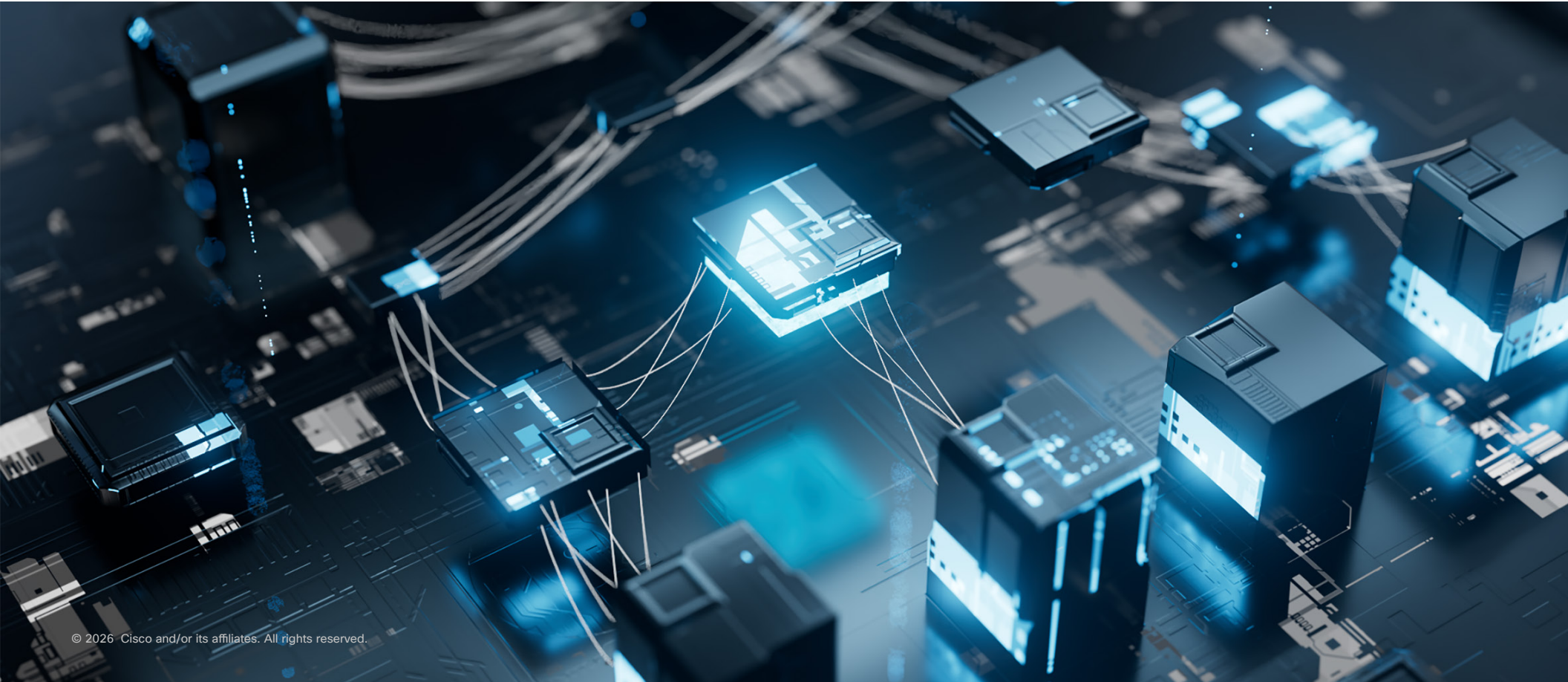


# Cisco AI PODs with Nutanix Agentic AI: Enterprise-Ready AI Infrastructure



## Benefits

- **Accelerate AI time-to-production** with pre-validated, full-stack infrastructure building blocks
- **Create a turnkey experience** offering a simple UI and “one-click” deployment of LLMs, eliminating the need for deep data scientist expertise for AI services
- **Simplify AI model governance** with a unified control plane, bringing operational discipline to AI initiatives
- **Scale AI workloads predictably** from pilot to production with a modular architecture that grows seamlessly with your GPU capacity
- **Enforce built-in security and data privacy** with platform RBAC (role-based access control) to control which users and teams can access specific models and endpoints

## A unified foundation for enterprise AI success

Cisco® AI PODs with Nutanix Agentic AI deliver a seamless, full-stack infrastructure that fuses high-performance hardware with intelligent AI orchestration. By tightly integrating GPU-dense Cisco UCS® compute and high-performance Cisco Nexus® AI networking with the Nutanix Kubernetes Platform (NKP) and the Nutanix Enterprise AI (NAI) for AI model inference and management, this is a complete, cohesive ecosystem engineered for the rigors of modern AI.

This joint architecture standardizes your stack from fabric to lifecycle, using modular building blocks to simplify provisioning, automate

Kubernetes, and enforce governance from the data center to the edge.

Gain a predictable, scalable platform that empowers you to move from experimental pilots to production-grade AI with total operational clarity and confidence. From model onboarding and RAG enablement to day-2 observability, Cisco AI PODs with Nutanix operationalizes your AI strategy, turning complex, fragmented workflows into a high-performance engine for innovation.



Figure 1. Scale AI solutions with Cisco AI PODs with Nutanix

## Learn more

### Achieve rapid deployment, consistent operations, and unified model governance

Cisco is reimagining data-center design, procurement, and operations for AI. To learn how to accelerate your transition to production-ready AI with Cisco AI PODs with Nutanix, visit <https://www.cisco.com/site/us/en/products/computing/hyperconverged/nutanix/index.html>.

## Operationalize AI with a unified, scalable foundation

Merge Cisco's high-density compute and high-performance, integrated AI networking with Nutanix's orchestration to create a turn-key environment. Remove the friction of assembling disparate components with a proven architecture for model deployment and inferencing.

Cisco AI PODs with Nutanix enable you to:

- **Deploy validated scale-unit building blocks** that eliminate design guesswork and ensure consistent performance with architecture built specifically for enterprise AI
- **Standardize cluster operations through the Nutanix Kubernetes Platform**, which provides consistent provisioning, scaling, and policy enforcement across your data center and edge environments
- **Govern model lifecycles using Nutanix Enterprise AI (NAI)**, a centralized control plane that standardizes how LLMs and open-source models are onboarded, tuned, and monitored for peak performance

Unlike disjointed hardware and software stacks, this joint solution provides a single, cohesive ecosystem. It transforms AI from experimental silos into a repeatable, enterprise-grade platform, ensuring your infrastructure scales seamlessly with evolving AI ambitions.