

Cisco AI POD for Splunk

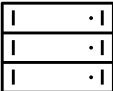
The preferred on-premises deployment for Splunk AI

Splunk AI improves team productivity and delivers more resilient operations, from Splunk AI Assistant and Agent Launchpad to Splunk hosted models, with more AI capabilities to come. These capabilities are now available to organizations that need to run AI in their own data center.

Product benefits



Unlock Splunk AI in your data center



Reduce infrastructure and support complexity



Start right-sized and scale

Cisco AI POD for Splunk is the preferred on-premises deployment for Splunk AI. It combines AI Tier, the software foundation for supported Splunk AI workloads, with Cisco Unified Computing System (UCS), NVIDIA GPU acceleration, OpenShift, and the Splunk Operator for Kubernetes in a pre-sized, pre-tested, pre-validated system backed by single-vendor support, with support for FIPS 140-3 deployments. This means less work sizing, integrating, validating, and supporting the stack yourself, while the workload stays in your environment. The value is not the server itself; it is a defined, supported path to operationalize Splunk AI in environments where you need local control.

Many Splunk Enterprise customers cannot rely on cloud-hosted AI services because of data sovereignty, security, regulatory, governance, or air-gap requirements. Running Splunk AI locally creates a second problem: the infrastructure required to run those workloads has to be sized, procured, validated, integrated, and supported across GPU compute, networking, Kubernetes, and the Splunk AI software itself.

Splunk AI experiences on the POD

Splunk AI experience	What it does	Availability
Splunk AI Assistant	Search in Splunk in natural language and agent mode	Available
Splunk hosted models	Specialized generative AI time-series model for forecasting and anomaly detection and open-weight LLM models for general reasoning	Available
Agent Launchpad	Build, run, and review custom agents invoked from searches and alerts in Splunk workflows with full context	Planned Q4'26

Agent Launchpad and support for FIPS 140-3 deployments are planned for general availability on AI Tier and Cisco AI POD for Splunk in Q4'26. Timing and capabilities are subject to change.

What is inside the AI POD

Cisco AI POD for Splunk arrives as one system. AI Tier, the software that delivers supported Splunk AI capabilities, is validated with an integrated system of Cisco compute, networking, and Kubernetes architecture rather than assembled on site.

Splunk AI experiences

What your teams interact with, running on the layers below.

AI Tier

Splunk software

AI services, models, and orchestration that the Splunk AI experiences run on.

SOK

Splunk Operator for Kubernetes

Kubernetes-native operational model for declarative deploy, upgrade, scale, and monitor of AI services, including mixed mode alongside an existing Splunk Enterprise deployment.

OpenShift

Red Hat

Kubernetes platform that runs the containerized AI services on the POD.

Cisco UCS

Server + GPU

Pre-validated Cisco UCS compute with NVIDIA RTX PRO 6000 GPU acceleration for AI inference and model workloads.

Cisco Intersight provides infrastructure management for the system.

Hardware profile

A defined minimal configuration from the general-purpose Cisco AI POD hardware catalog, sized so you can start small and scale as AI usage grows.

Component	Qty
Splunk Kubernetes control plane	
Cisco UCS C225, OpenShift cluster	3
AI Tier plane	
Cisco UCS C845	1
NVIDIA RTX PRO 6000 GPU	2
Networking	
Cisco Nexus switching	TBD

Is this the right fit

- You run Splunk Enterprise in an on-premises, private-cloud, sovereign, regulated, FIPS-compliant, or air-gapped environment.
- You want supported Splunk AI capabilities to operate locally rather than through cloud-hosted AI services.
- You need GPU-backed infrastructure for Splunk AI workloads and do not already have an approved environment available to your Splunk team.

If you have your own infrastructure

AI POD is not required to run Splunk AI on-premises. If you already have qualified GPU and Kubernetes infrastructure, or use another infrastructure standard, AI Tier can be deployed as software on your own environment instead.

Download AI Tier >

Talk to your Splunk account team to qualify the workload, confirm your infrastructure readiness, and size the right configuration for your environment.

This product brief is for informational purposes only. The products and capabilities described are governed by the applicable fully signed ordering document and any incorporated terms and conditions. Capabilities identified as planned are shown for planning purposes only; timing and capabilities are subject to change.

