

# Cisco Services for Secure AI Factory with NVIDIA

Deploy, optimize, and support, enterprise-ready AI infrastructure with unmatched efficiency

## Outcomes

- **Accelerate** Time to First Intelligence (TTFI).
- **Optimize** Cost to True Scale (CTTS).
- **Secure** your AI environment from day one.

## Simplify and secure AI adoption

You want your AI initiatives to succeed quickly, securely, and at scale – but over 70 percent<sup>1</sup> of enterprise AI deployments stall before production due to infrastructure complexity, security challenges, and skills gaps, leading to wasted GPU investment and lost competitive advantage.

Cisco® Secure AI Factory with NVIDIA delivers modular, validated AI infrastructure built on Cisco UCS® compute, NVIDIA GPUs, and Nexus® networking – brought together and operationalized through Cisco Services, with integrated security and observability aligned to NVIDIA Enterprise Reference Architectures.

Our services are designed to accelerate Time to First Intelligence (TTFI) – the time from hardware delivery to first actionable AI insights – while optimizing Cost to True Scale (CTTS), so infrastructure costs grow predictably with AI demand, not exponentially.

Security and observability are built in from day one. Cisco AI Defense protects against 200+ threat categories, Hybrid Mesh Firewall provides advanced perimeter security, and Splunk® delivers end-to-end observability – capabilities unmatched by other vendors.

One validated solution. One secure path. One trusted deployment. With Cisco Services, organizations deploy faster, scale predictably, and operate with confidence across the full AI lifecycle.

## What you get

Move your AI infrastructure deployments forward with tailored solutions and expert guidance.

- **Planning and design:**  
Align AI infrastructure design to your workloads, security posture, and scaling objectives across compute, networking, storage, and observability.
- **Implementation:**  
Deploy a customized AI infrastructure with integrated fabrics, security, and observability – built for high-performance AI workloads.
- **Optimization and testing:**  
Maximize GPU efficiency and eliminate bottlenecks with environment-specific tuning of network, compute, and storage—validated using your AI workloads.
- **Operational readiness:**  
Prepare your teams for Day 2 with runbooks, dashboards, monitoring, and hands-on knowledge transfer.
- **Post-implementation support:**  
Get expert guidance to sustain performance, resilience, and confidence as you scale.

## Tailor your AI infrastructure with modular options

Build your AI infrastructure to fit your unique workloads and business goals. Our services for Cisco Secure AI Factory with NVIDIA offer flexible, validated building blocks that simplify deployment, boost performance, ensure security – helping you scale confidently and speed up time to value.

Choose the infrastructure, security, and observability components that fit your workload and scaling needs.

### Infrastructure options

Options	Use case	Examples
Compute only	Rapid inferencing	Pre-trained models, quick wins
+Frontend	Customer AI apps	Chatbots, image detection
+Backend	Model training	LLMs, proprietary AI, research
Full stack	Enterprise AI	End-to-end development at scale

### Security and observability options

Options	Use case	Examples
AI Defense	Secure AI Workloads	Prompt injection, data leakage, model validation
Hybrid Mesh Firewall	Protect the Network	Zero trust, perimeter security, compliance
Splunk Observability	Optimize Operations	Monitoring, alerting, cost optimization

## Proven results

*A leading telecommunications provider deployed and validated a 1,000+ GPU cluster in a matter of weeks with Cisco Services – a process that typically takes months in the industry.<sup>2</sup>*

## Maintain peak performance with support

Cisco Support is built for AI environments where uptime and performance are critical. With solution-level expertise across AI infrastructure, networking, security, and observability, our support goes beyond break-fix to proactively protect resiliency and performance. You get direct access to experts who understand AI workloads – eliminating delays, finger-pointing, and fragmented support. The result is faster resolution, fewer disruptions, and greater operational confidence.

With support, you get more value from every engagement through proactive, predictive, and personalized support.

## Why choose Cisco Services?

Cisco Services simplifies AI deployment through deep expertise, advanced automation, and proven tools. Our tailored approach ensures secure, scalable AI infrastructure and faster business value. Key benefits include:

- **Accelerate Time to First Intelligence** and business value with AI.
- **Realize 90%+ GPU utilization from day one in large-scale AI infrastructure deployments<sup>2</sup>** – no lengthy ramp-up periods.<sup>2</sup>
- **Benefit from proven expertise** supporting service provider and enterprise AI deployments.
- **Leverage integrated security** for comprehensive AI workload protection.
- **Gain end-to-end observability** for proactive monitoring and operational resilience.



← Time to First Intelligence: From months to weeks →

## Get started

Schedule a readiness assessment with our AI infrastructure experts. Contact your Cisco account team, partner, or visit [www.cisco.com/go/profservices](https://www.cisco.com/go/profservices).

<sup>1</sup>Gartner, 2024

<sup>2</sup>2025, Leading Telco and Digital Services Provider in APJC