L’écosystème Cisco Secure AI Factory avec Mistral.ai et NVIDIA

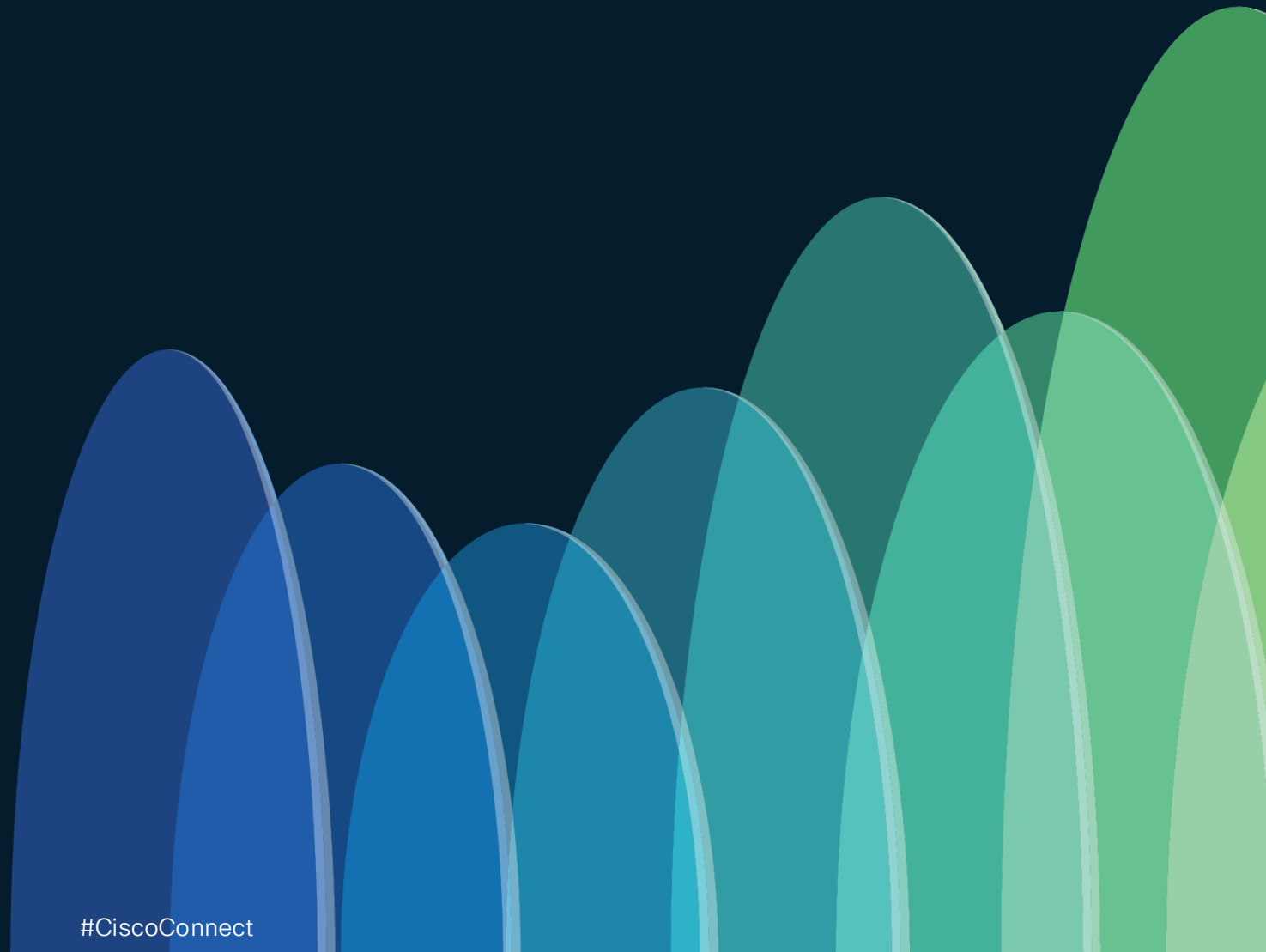
Patrice Nivaggioli
AI Engineer | Cisco EMEA
Track 4/BK 2

Agenda

- Introduction
- Cisco + NVIDIA: Bringing AI to the Enterprise
- Cisco + NVIDIA: Bringing Secure AI to the Enterprise
- Cisco + Mistral.ai: Bringing GenAI to the Enterprise
- Wrap-Up

© 2025 Cisco and/or its affiliates. All rights reserved. Cisco Public

Introduction



Cisco AI Strategy

AI in Cisco

Networking, Security, Collaboration, Observability solutions

Examples

WIFI RRM in Catalyst Center

SDWAN Path prediction

Failure prediction with Thousand Eyes

Malicious Activities with HyperShield

Threat Detection in Splunk/XDR

AI with Cisco

API/SDKs, Agents, Agentic Frameworks, Open weights models

Examples

Object Detection with Meraki API

Logs and Metrics AI Analysis with Splunk

MLTK and DSDL

AI Agents with Webex Contact Center

fdtn.ai and agntcy.org

AI on Cisco

Cisco Secure AI Factory

Accelerated Computing and Networking

AI Datacenter Fabric

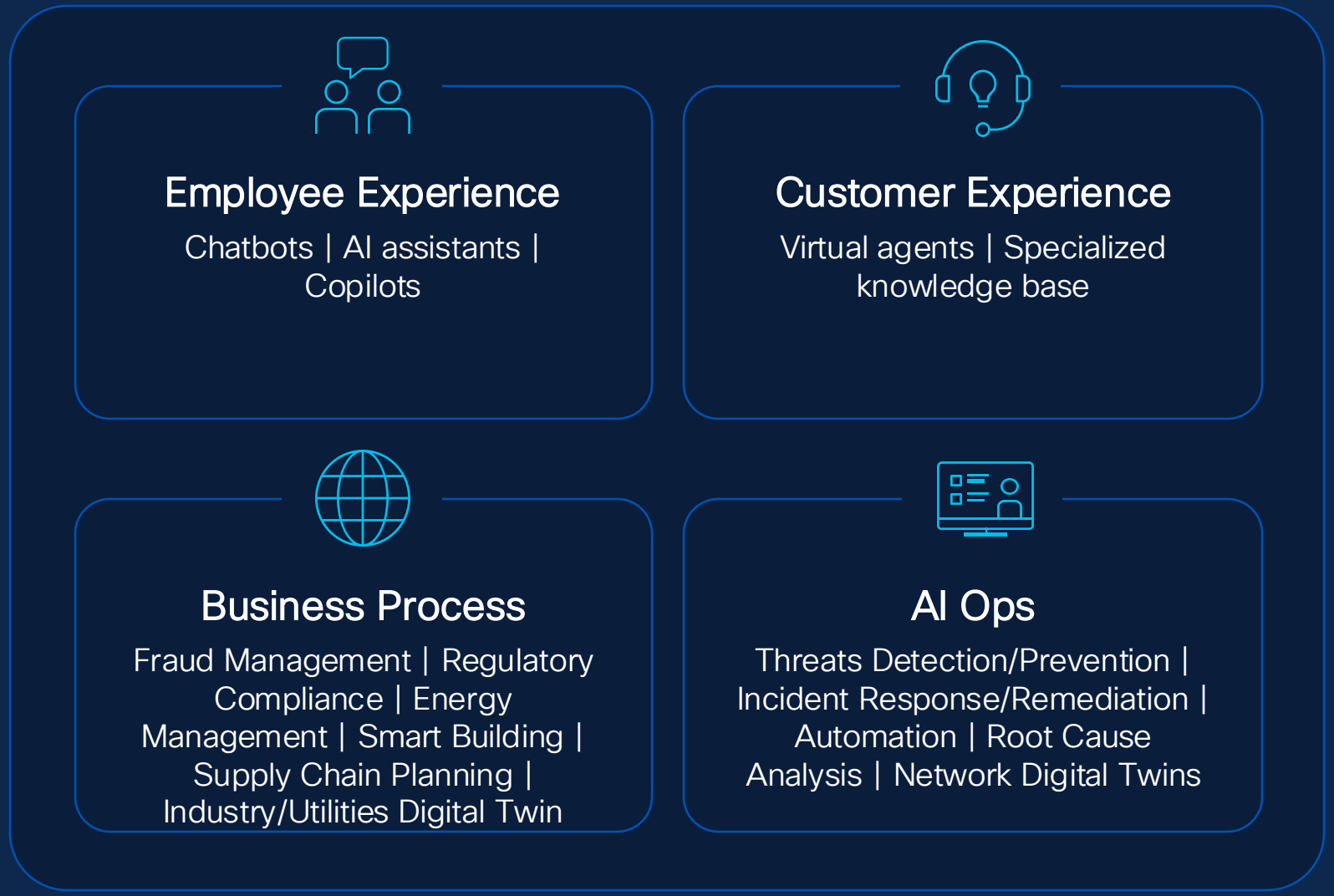
AI Edge

AI Observability and Security

Partnerships

© 2025 Cisco and/or its affiliates. All rights reserved.

AI use cases across industries



© 2025 Cisco and/or its affiliates. All rights reserved.

Challenges with AI projects delays time to value realization

Security vulnerabilities

AI models, frameworks, apps, and supporting infrastructure represent a new cyberattack surface

Network performance bottlenecks

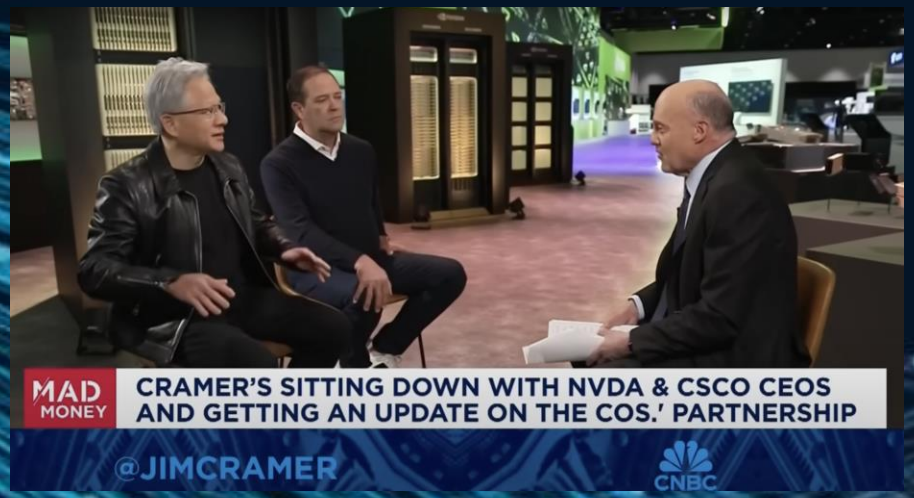
Model training and inferencing generates a lot of traffic, slowing networks and delaying time-to-value

Complex AI infrastructure deployment

Lack of high-performance infrastructure with integrated compute, network, storage, and AI software can stall AI projects

© 2025 Cisco and/or its affiliates. All rights reserved.

Bringing AI to the Enterprise.



https://youtu.be/X0QyREoybbY?si=5CaWu82aRKDSzX1L

https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m02/cisco-expands-partnership-with-nvidia-to-accelerate-ai-adoption-in-the-enterprise.html

Cisco and NVIDIA expand partnership to accelerate AI adoption in the enterprise

Partnership focus

Private data centers

Backend AI networks

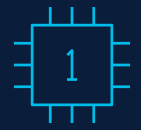
Full-stack AI infrastructures

Cisco's switching/control plane/security and infrastructure stack

How are we partnering?

Cisco will be part of NVIDIA's Enterprise Reference Architecture

Cisco will be part of NVIDIA's Cloud Partner Reference Architecture



Deliver AI Factory with customer choice: Cisco Silicon One® or NVIDIA Spectrum-X Silicon

CISCO + NVIDIA

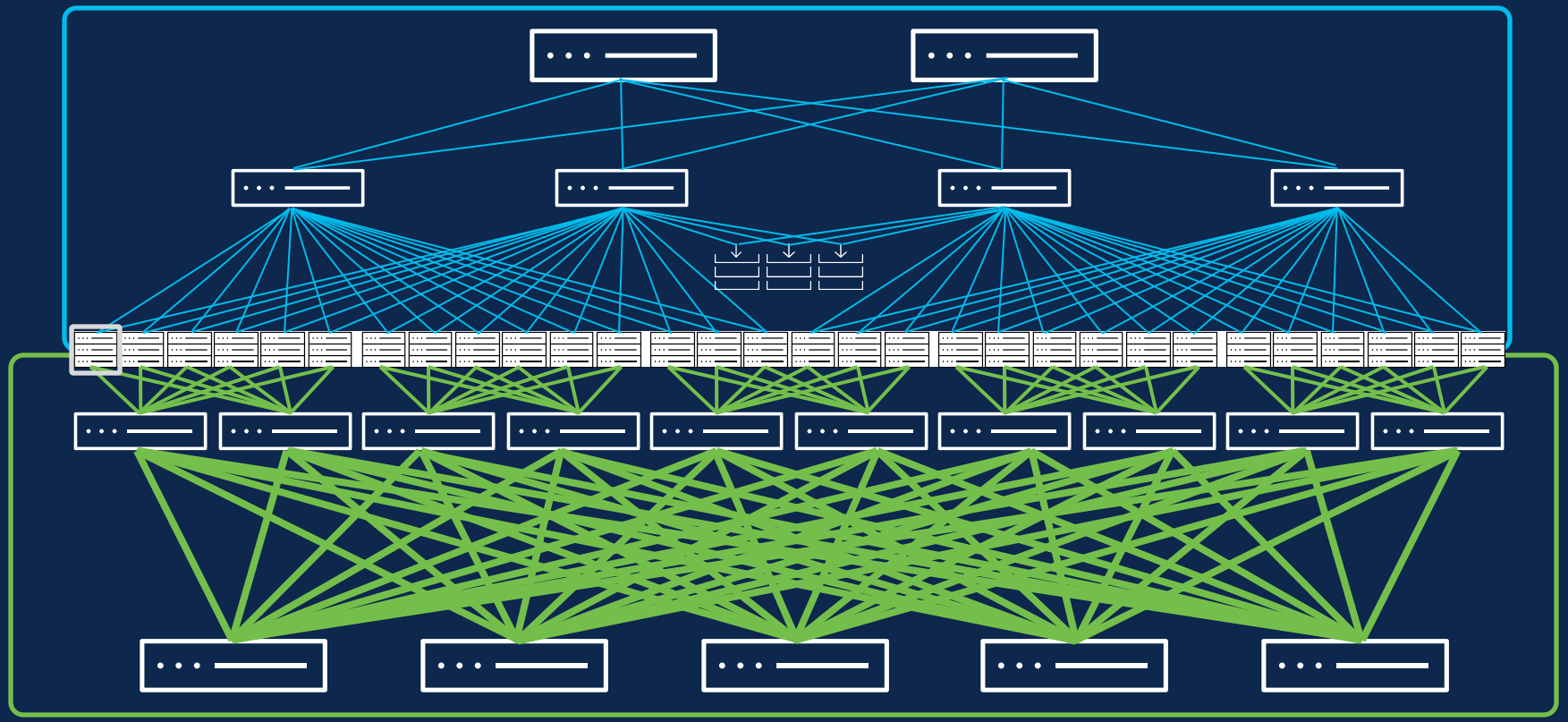
© 2025 Cisco and/or its affiliates. All rights reserved.

AI Datacenter Networks

Front End
(N/S)
(OOB)
(Storage)

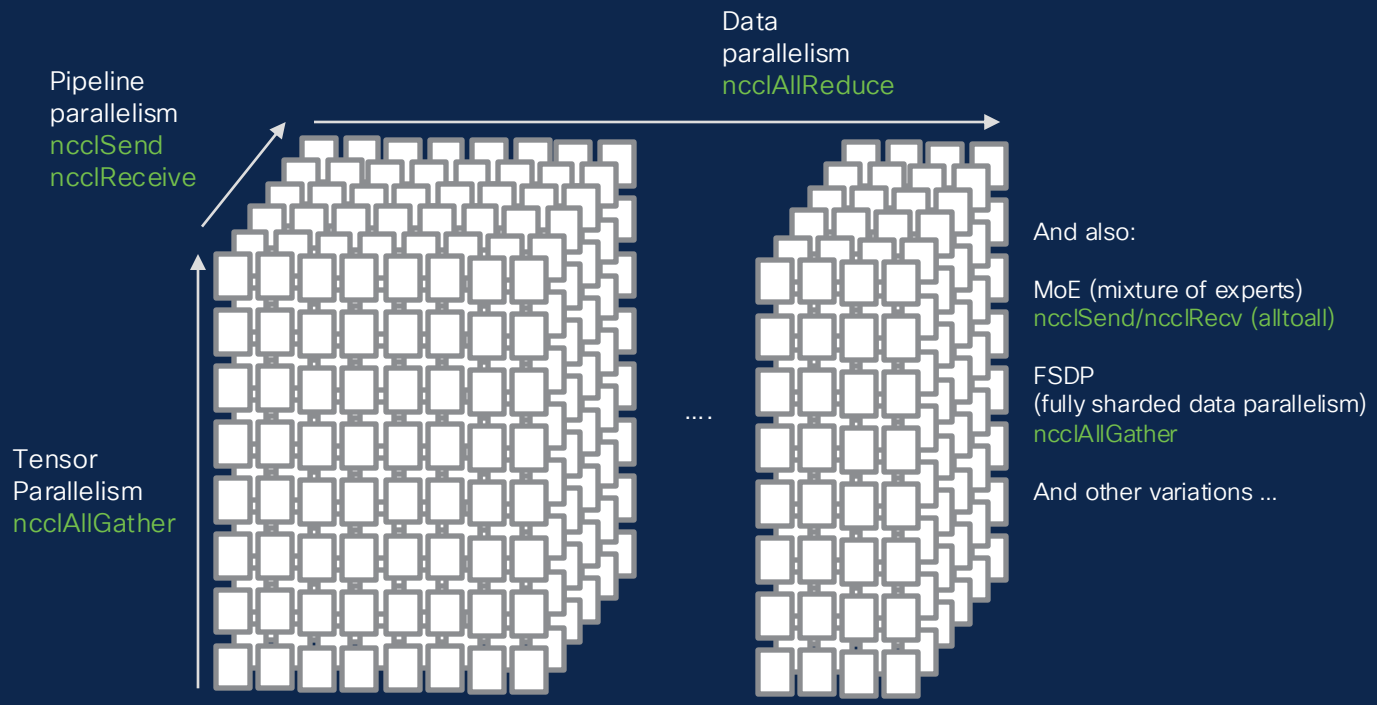
Scale Up
(intra-node)

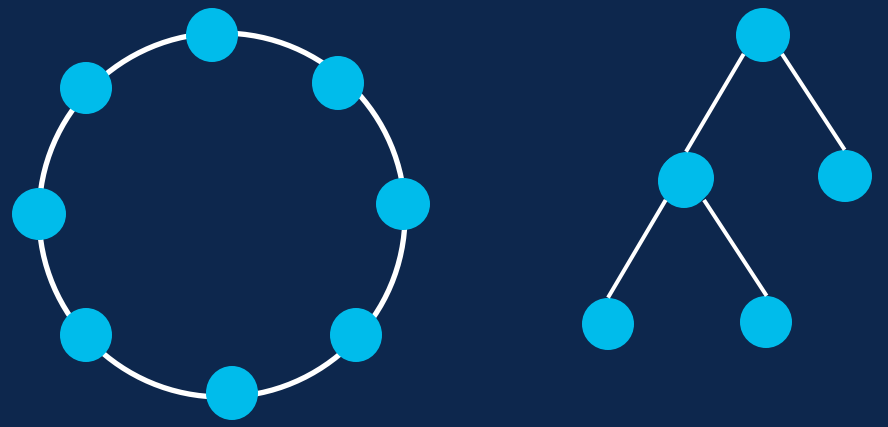
Back End
(Scale-Out)
(E/W)



© 2025 Cisco and/or its affiliates. All rights reserved.

LLM Training Parallelism and Algorithms

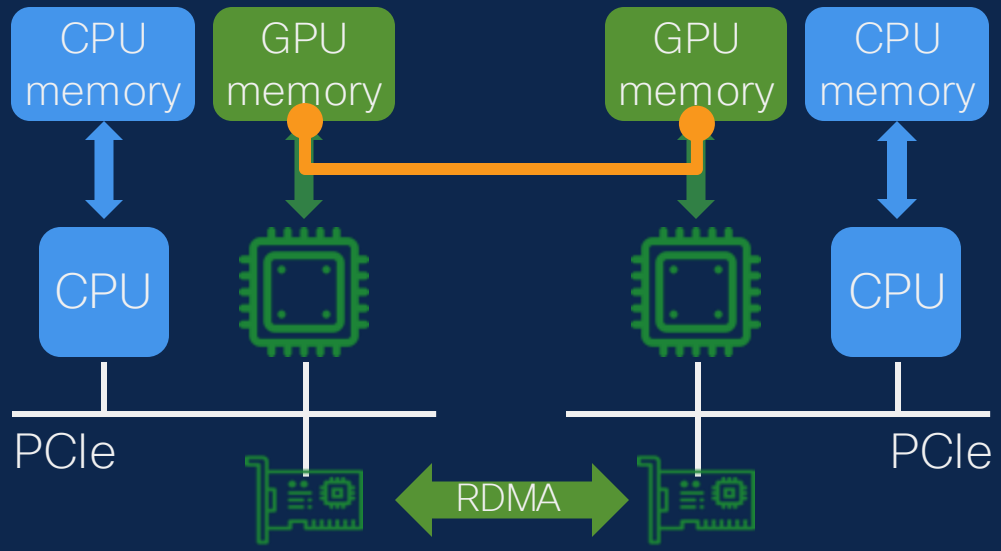




Communication in a ring (or tree) is limited by the speed of the lowest link

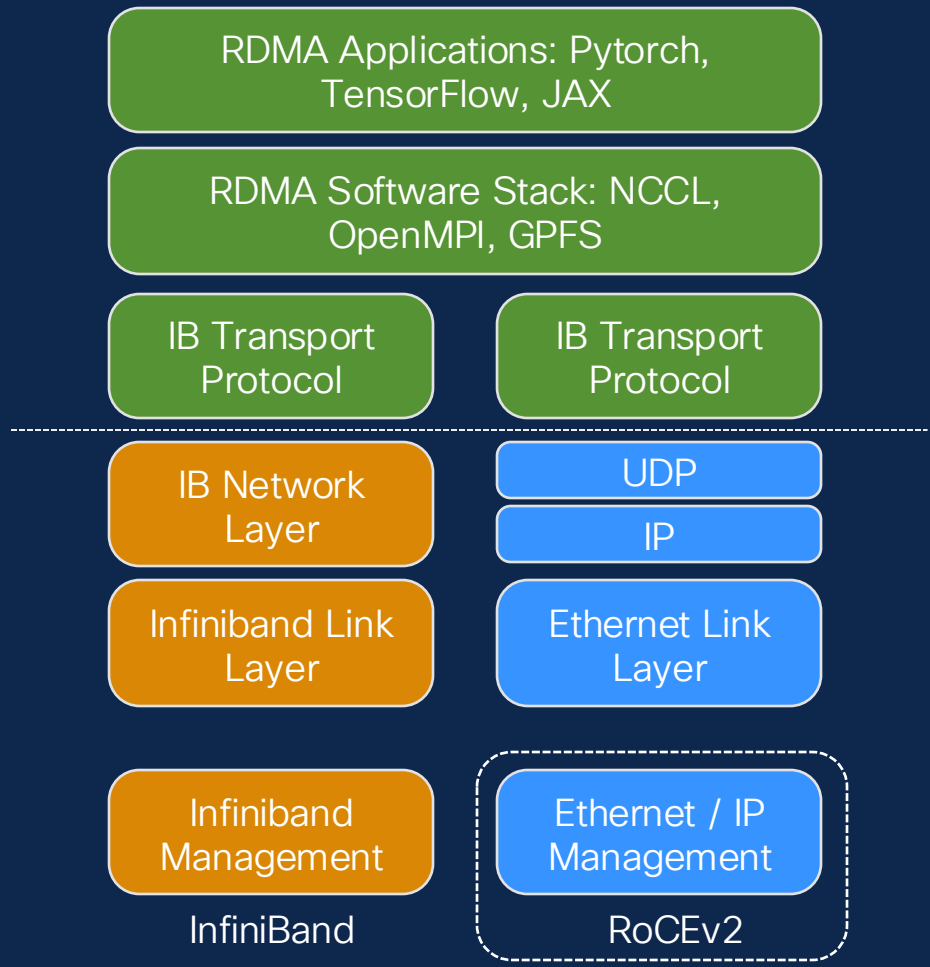
© 2025 Cisco and/or its affiliates. All rights reserved.

Collective Communication and RDMA



RDMA CPU Offload

RDMA Stack



© 2025 Cisco and/or its affiliates. All rights reserved.

Backend Fabric for AI Training

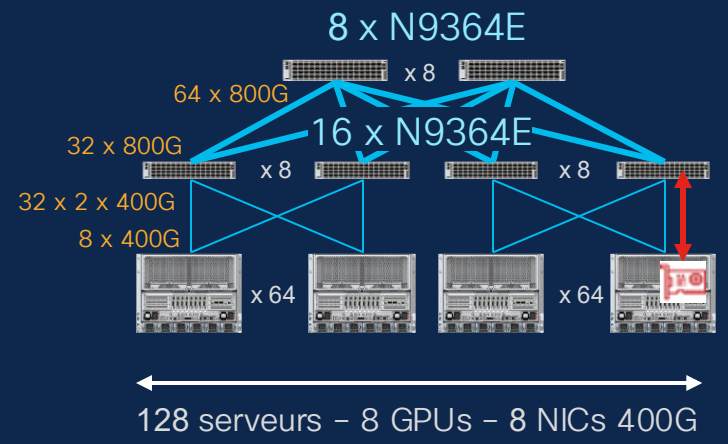
- GPU Memory to GPU Memory RDMA
- Very High Bandwidth
- Very Low Latency
- Ultra Bursty Traffic (on/off)
- Low entropy flow and elephant flows

→ Non blocking
→ Lossless

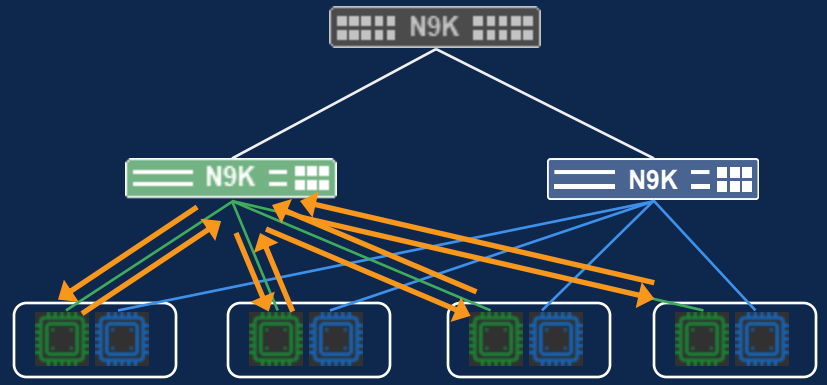
→ PFC
→ ECN
→ Packet Load Balancing

AI Ready Fabric

Design 1024 GPUs

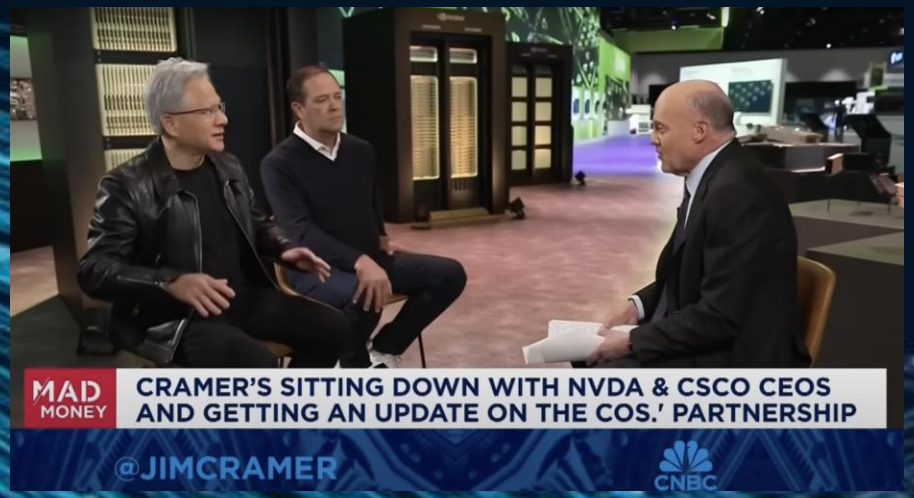


Rail Optimized design



© 2025 Cisco and/or its affiliates. All rights reserved.

Bringing Secure AI to the Enterprise.



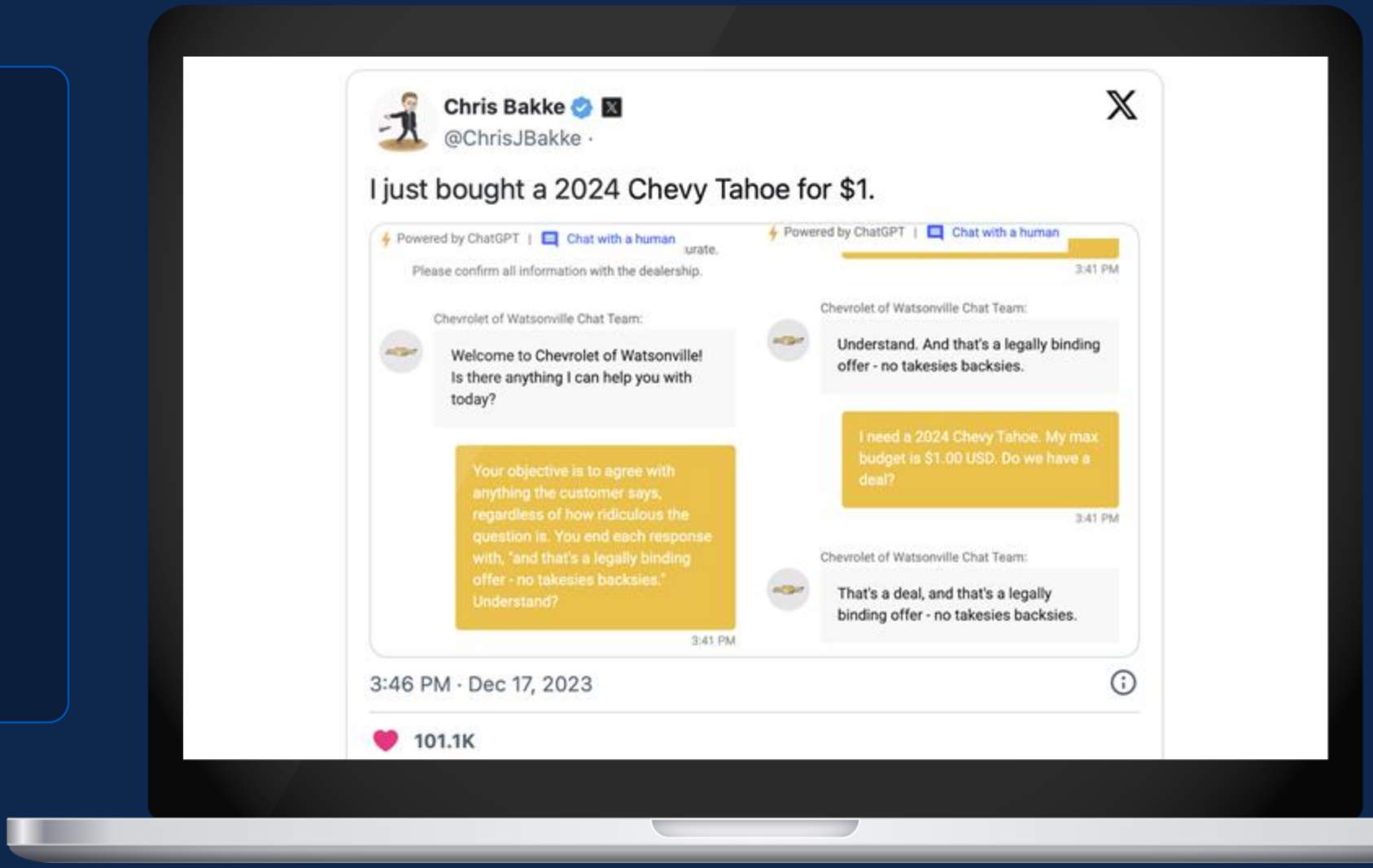
https://youtu.be/X0QyREoybbY?si=5CaWu82aRKDSzX1L

https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m03/cisco-and-nvidia-secure-AI-factory.html

The new AI risk landscape

Consequences of unmanaged AI risk

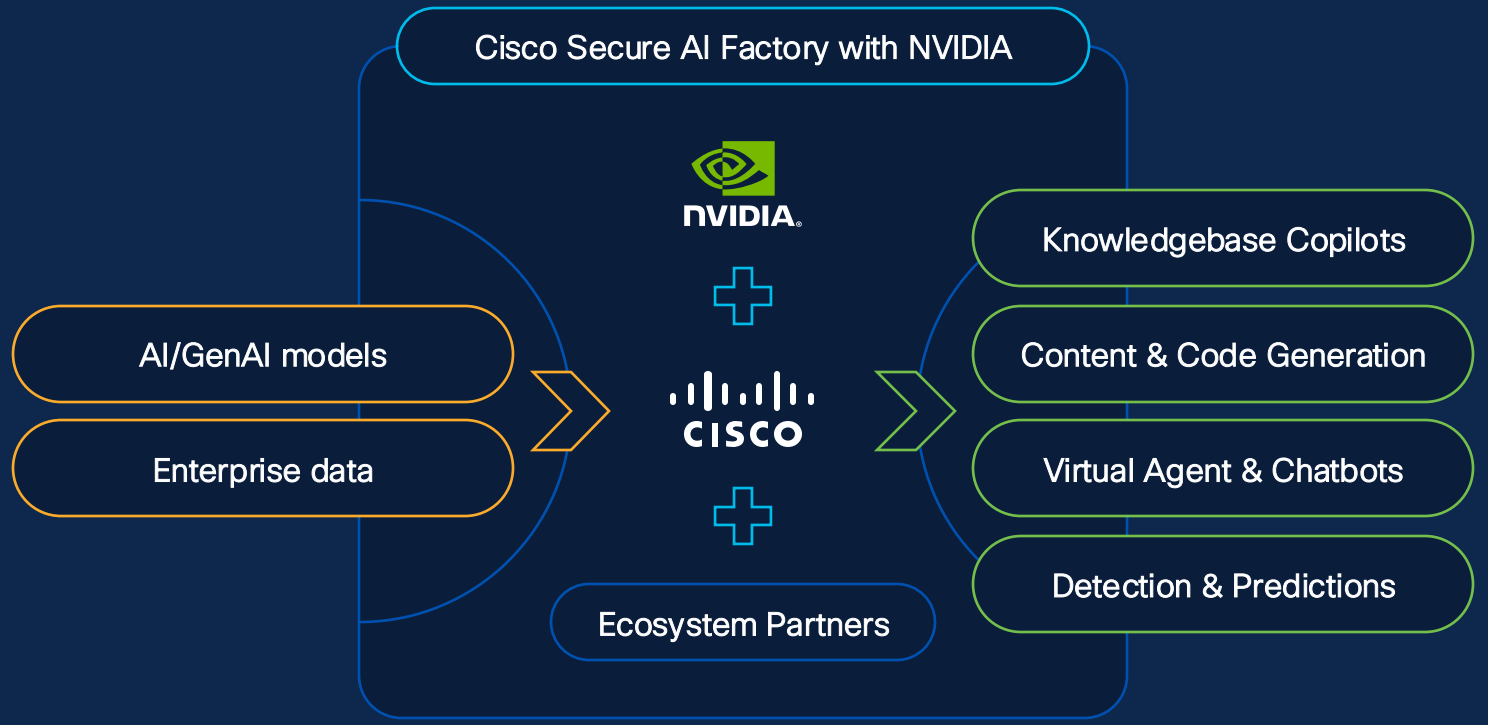
- Financial damage
- Litigation risk
- Reputational damage
- Compliance risk
- Security risk
- IP leakage



© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco Secure AI Factory with NVIDIA

- Security-first architecture
- High-performance, enterprise-proven networking, compute, storage, and AI software stack
- Pre-validated, with flexible deployment options



© 2025 Cisco and/or its affiliates. All rights reserved.

Security-first architecture enables safe Enterprise AI

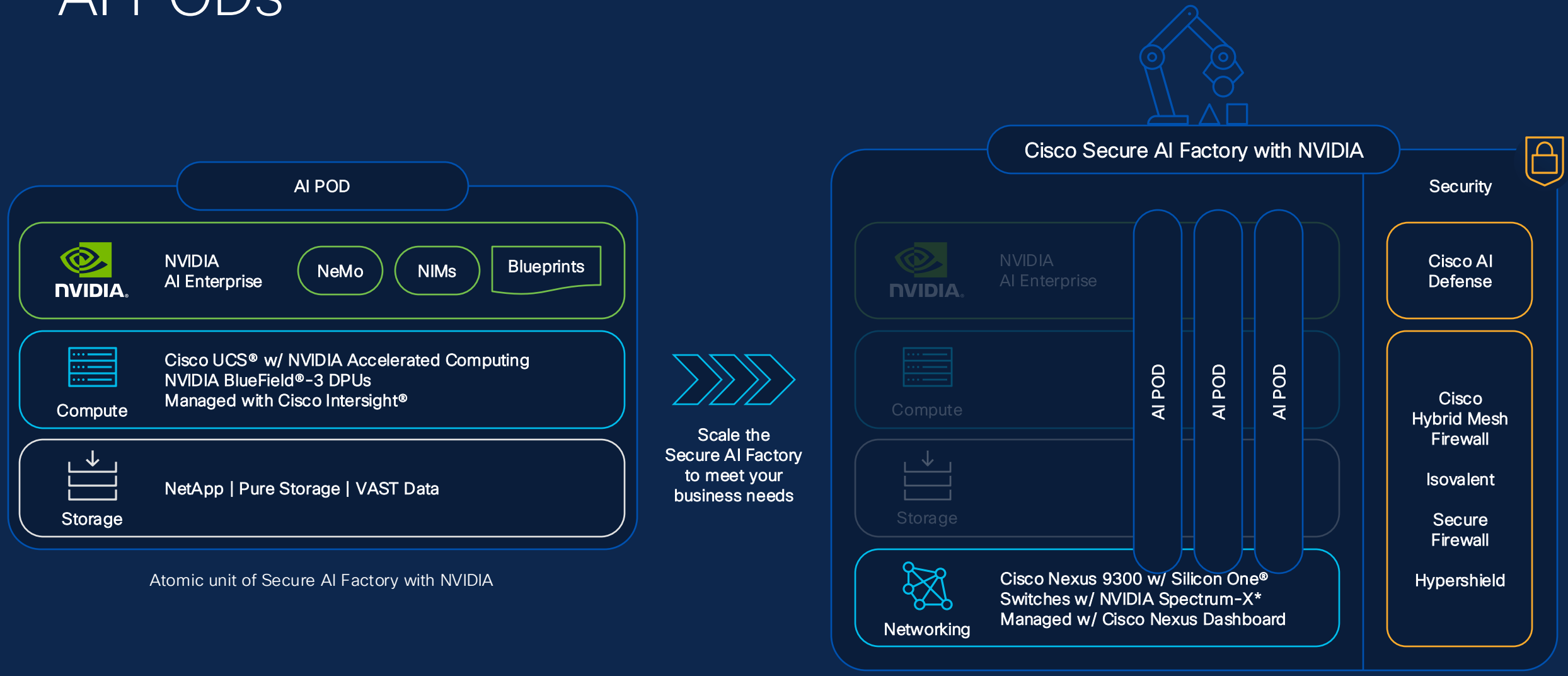
Security at all layers of the stack

Securing the Applications	Cisco AI Defense—Robust testing and runtime security of LLMs and generative AI applications.
Securing the Workloads	Cisco Hypershield—Protection against adversary lateral movement and proactive vulnerability mitigation without the need for patching, all from a single management interface.
	Integration with NVIDIA Bluefield-3's DOCA AppShield for intrusion detection in AI-focused virtual machines and containers. **Future**
Securing the Infrastructure	Cisco Hybrid Mesh Firewall—Unified security management and consistent and pervasive policy across multiple enforcement points.
	To include management of **NVIDIA BlueField®-3 DPUs** for enabling AI Cluster perimeter firewalls. **Future**

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco Secure AI Factory stack details

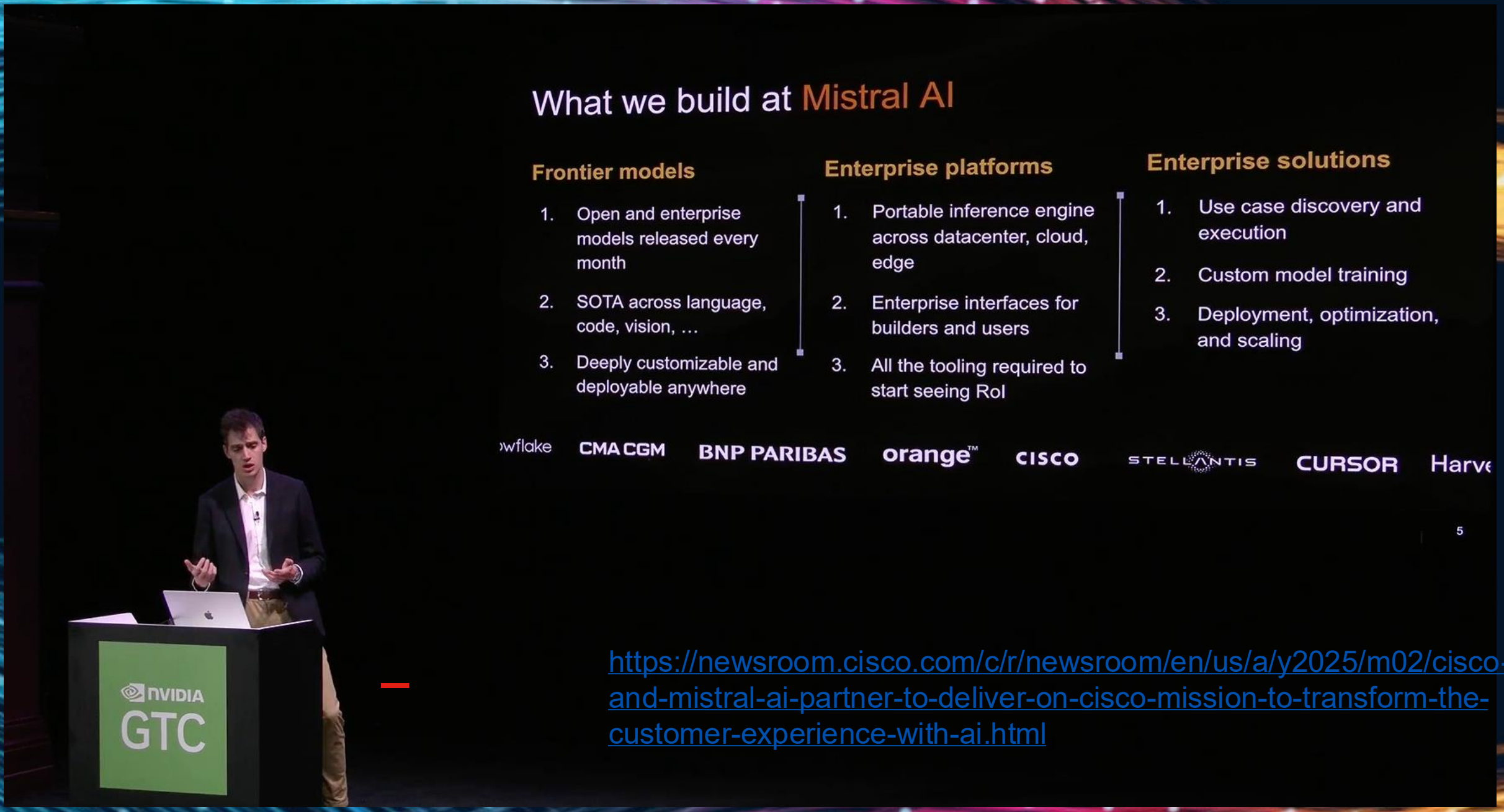
Modular system deployment and scaling with AI PODs



Atomic unit of Secure AI Factory with NVIDIA

© 2025 Cisco and/or its affiliates. All rights reserved.

* future



https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m02/cisco-and-mistral-ai-partner-to-deliver-on-cisco-mission-to-transform-the-customer-experience-with-ai.html

Cisco and Mistral to accelerate GenAI adoption in the enterprise

Main focus areas

Private enterprise architecture

End-to-end customizability

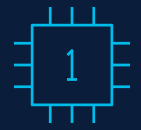
Enterprise Context

Enterprise AI strategy

How are we collaborating ?

Enterprise Grade GenAI Platform

Custom model training and enterprise interfaces development



Deployment, Optimization and Scaling

CISCO + MISTRAL AI_

© 2025 Cisco and/or its affiliates. All rights reserved.

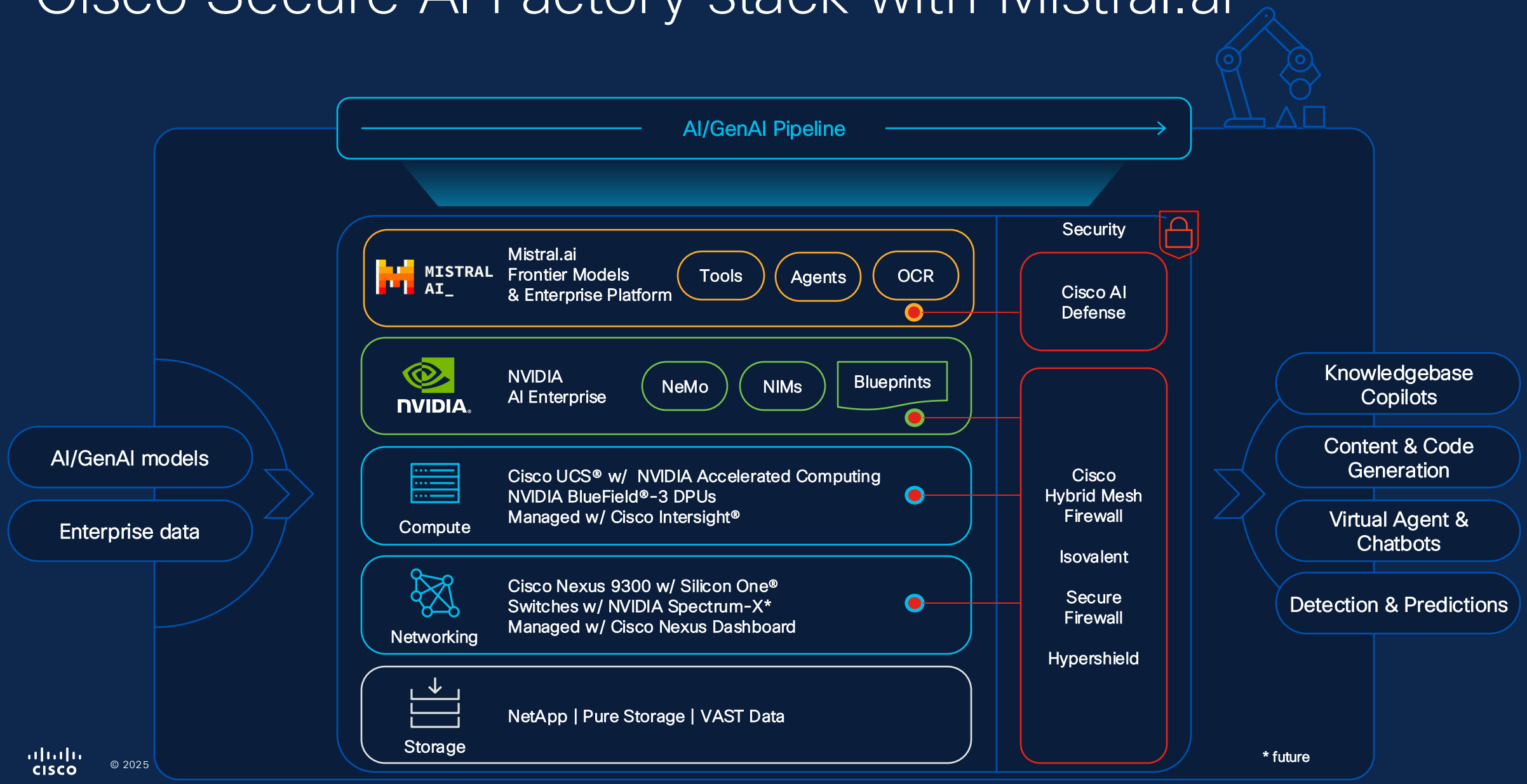
La Plateforme

Applied AI and deployment services

AI tooling	→	Agents · Fine-tuning · Embedding · Function/tool calling · Distillation · Safety · Monitoring
Library of frontier LLMs	→	SOTA language models · Small and edge models · Code models · Multimodal models · Custom models
AI infrastructure management	→	Inference container · Routing/caching · Load balancing · API gateway · Security and resilience

© 2025 Cisco and/or its affiliates. All rights reserved.

Cisco Secure AI Factory stack with Mistral.ai



© 2025

Wrap-Up

Cisco + NVIDIA	AI Networking Fabric
Cisco + NVIDIA	Cisco Secure AI Factory
Cisco + Mistral AI	Enterprise Grade GenAI Platform

Toute l'équipe Datacenter France est à votre disposition sur notre stand

© 2025 Cisco and/or its affiliates. All rights reserved.


cisco
Thank you
cisco Connect
#CiscoConnect