CISCO Engage | Tech Day

From Vision to Value: Cisco's Partnership Model for your AI Journey

Arun Manku
AI GTM Specialist, Cisco

Matthew Levesque
VP, OnStak

December 17th, 2025

AI is no longer a technology strategy. It is a sovereign capability

1990s
Machine learning

2022
ChatGPT

2024
Assistants

2025
Agentic AI

2026
Physical AI

Why Customers Are Struggling with AI

Rapid model & GPU Churn

Too many pilots, not enough production

Security, compliance, sovereignty pressure

The Cisco AI Ecosystem

OEMs: Compute & Acceleration

- NVIDIA and AMD provide hardware foundation, etc.

MINT Partners: AI Readiness, MLOps & AI Workflows

- Enable RAG and ML operations, etc.

VAR/GSI: Transact, Rack & Stack, etc.

- AI Readiness, Consulting, Resellers

ECO SYSTEM: Customization

- VAST Data, RedHat, Nutanix

From Vision to Value

Accelerated Time to Value

- Weeks to value, not months
- Reduces implementation timelines

Risk Mitigation Through Validation

- Lower risk via validated architectures
- Proven deployment models

Mission-Aligned AI Outcomes

- AI tied directly to mission outcomes
- Converts concept to measurable ROI

Building AI Together

No one wins alone

- Ecosystems drive success
- Silos of information decreases chances of successful pilots

Ecosystems Win

- Collaboration multiplies success
- Shared Innovation accelerates growth

Let’s Build Together

- Partner-first approach
- Strong, confident commitment

OnStak- Operationalize AI

Matthew Levesque VP

Date placeholder

Agenda

1. OnStak
2. Mint Program
3. The Challenge in AI
4. How we approach
5. Design Elements
6. Process
7. Real World Examples
8. Call To Action

About OnStak

Cisco Application Experience Partner of the Year – Americas

2022 | 2023

DSI Partner of the Year - East

2022

DSI Partner of the Year – West

2023 | 2024

150+ Global Partner Leadership Sessions

11+ Years of Operations

500+ Advance Technology Implementations

90% B2B Service Renewal

80% Certified Staff

100 + FTE Professionals

2 Global Innovation Centers

8% Invested in R&D

Key Partnerships

CISCO · NVIDIA · aws

Our Focus

AI · Data · Observability · Cloud

cisco

Engagement Map

Excellence, Innovation and Trust

MINT

Program Overview

WHAT

- Program started in 2017
- Non resellers 3rd party companies providing on-demand services
- 18+ Partners
- Worldwide coverage

HOW

- BU vetted - Rigorous qualification program
- Fixed Services MINT SKU on Cisco GPL
- Cisco Branded or Partner White-Labeled
- Revenue share model with the revenue split 70% for partner and 30% for Cisco

WHY

MINT delivers high value outcomes:

to Cisco			to Channel Partners		
Cap case avoidance	Scale BU TMEs	Drives profitable sales motion via Cisco GPL	Builds partner proficiency	Drives customer lifecycle adoption	De-risks new Cisco solutions

The Challenge

AI is Moving Faster Than Enterprise Infrastructure

Organizations are accelerating toward AI adoption, but face critical challenges around governance, data readiness, security, scalability, and operational maturity. While AI innovation races ahead, enterprise systems struggle to keep pace with the demands of production-grade AI deployment.

The Enterprise AI POD Program provides a modular, deployable, production-grade architecture that addresses these systemic challenges head-on, enabling organizations to harness AI's potential while maintaining enterprise-grade governance and control.

Five Critical Blockers Preventing AI Success

AI Risk + Governance Gaps

Shadow AI proliferates across the organization, unsafe LLM usage increases exposure, and missing auditability creates compliance vulnerabilities that put the enterprise at risk.

Fragmented Data Systems

Inconsistent access patterns, complex sovereignty requirements, and limited data lineage prevent organizations from leveraging their data assets effectively for AI initiatives.

Insufficient GPU Readiness

Complex sourcing challenges, difficult orchestration, and performance tuning requirements create bottlenecks in deploying AI workloads at scale.

No Enterprise Observability

Existing monitoring tools don't cover LLM interactions, data access patterns, prompt engineering, or latency issues critical to AI operations.

Lack of Repeatable Patterns

Every AI initiative becomes bespoke and difficult to scale, leading to inconsistent results, duplicated efforts, and wasted resources across the organization.

The Bottom Line: Enterprises don't need point solutions — they need a cohesive, governed AI operating model that scales.

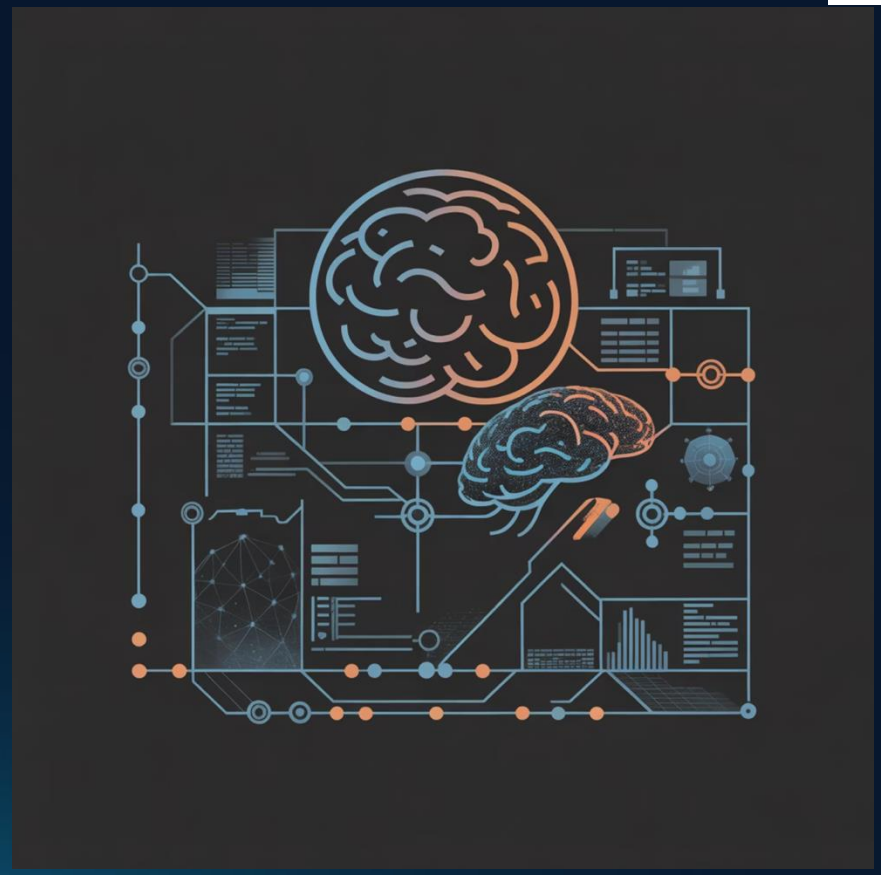
AI Requires a New Foundation

The evolution from traditional applications to machine learning to large language models demands a fundamental shift in enterprise architecture. This isn't just about adding AI capabilities — it's about building an entirely new operational foundation.

Five Critical Requirements

- **Stronger governance** at the model and data boundary to ensure responsible AI deployment
- **Full telemetry** across prompts, responses, and inference paths for complete visibility

- **Hybrid architectures** that support data sovereignty and optimal performance
- **GPU-ready infrastructure** that scales dynamically with demand
- **Repeatable, modular patterns** that teams can deploy predictably across the organization



AI is no longer an experiment. It's an architecture challenge that requires enterprise-grade thinking.

Introducing the Enterprise AI POD Program

A modular system of four interoperable PODS, each solving a critical part of the enterprise AI maturity journey. Each POD deploys independently to reduce risk and accelerate time to value, but together they form a cohesive enterprise AI fabric that addresses every layer of the AI stack.

1

Unified Edge Intelligence POD

Real-time telemetry and edge AI capabilities that transform network infrastructure into an intelligent, observable edge layer.

2

AI Defense & Governance POD

Comprehensive guardrails and policy enforcement that ensure responsible AI usage from day one.

3

Data Sovereignty Fabric POD

Governed, sovereign, RAG-ready data foundation that unifies enterprise data access while maintaining compliance.

4

On-Prem AI POD

GPU-accelerated private LLM and RAG infrastructure for organizations requiring complete data sovereignty.

What Makes This Unique

Governance-First Approach

Cisco AI Defense and NeMo Guardrails enforce responsible AI usage from day one, ensuring compliance and reducing risk before models ever reach production.

Observability Baked Into Architecture

Splunk, ThousandEyes, and Cisco Observability capture every AI telemetry signal — from prompt engineering to model responses to data access patterns.

Data Sovereignty Without Friction

VAST Data OS provides unified, sovereign, RAG-ready data access that accelerates innovation while maintaining strict data governance and compliance requirements.

Modular, Scalable, Future-Proof

PODS interoperate seamlessly but can be deployed individually, allowing organizations to start small, prove value, and scale systematically.

This is not a toolkit — it's a complete AI operating model designed for enterprise scale, compliance, and long-term success.

The Four-POD Architecture in Detail

Each POD contributes distinct, essential capabilities to create the full lifecycle of enterprise AI — from edge intelligence to data governance to model deployment to continuous observability.

Unified Edge Intelligence POD

Transforms Meraki infrastructure into an intelligent, observable edge that captures real-time telemetry, enables edge AI processing, and provides unprecedented visibility into network-based AI interactions.

AI Defense & Governance POD

Provides a comprehensive control plane for AI safety, auditability, and compliance. Enforces policies, monitors usage, and ensures responsible AI deployment across the organization.

Data Sovereignty Fabric POD

Creates governed, sovereign, enterprise-wide data zones that enable rapid AI development while maintaining strict data residency, privacy, and compliance requirements.

On-Prem AI POD

Delivers private, sovereign LLM and RAG infrastructure with GPU acceleration, enabling organizations to deploy cutting-edge AI capabilities without cloud dependencies.

How the PODS Work Together

The Complete AI Operating Model

User Application

Business users interact with AI-powered applications across the organization

Guardrails & Governance

Requests pass through AI Defense POD for policy enforcement and safety checks

Model Layer

LLMs and RAG systems process requests with full observability

Data Fabric

VAST sovereign zones provide governed, high-performance data access

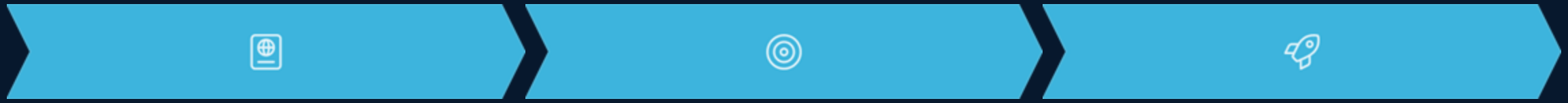
Observability

Continuous monitoring feeds improvements back into the system

The Result: AI becomes predictable, safe, scalable, and compliant — transforming from experimental projects into production-grade enterprise capabilities.

The Anatomy of a MINT Engagement

Our AI Governance engagement with a leading US healthcare company is delivering significant value through our joint partnership.



Assessment

Audit AI systems, identify compliance risks, stakeholder interviews.

Planning

Develop governance policies, ethical guidelines, roadmap.

Execution

Deploy frameworks, monitor, train, optimize.

This partnership attracted new executive stakeholders (CEO, COO, CISO, VP Data and Analytics, Chief Data Scientist), creating fresh opportunities for Cisco and the incumbent partner.

cisco

Ethical AI Governance Framework

A leading US Healthcare Company stands at a critical juncture in digital transformation. While AI technologies have been adopted across the organization, a cohesive governance framework is essential to maximize value, ensure compliance, and drive strategic innovation.

Current State Challenges

- Fragmented AI initiatives across entities
- Lack of unified governance approach
- Algorithmic fairness concerns
- Complex regulatory landscape (HIPAA, HITECH)
- Shadow AI systems requiring oversight
- Need for ethical AI implementation

Key Goals

- Comprehensive governance framework
- AI Center of Excellence design
- Proof-of-concept implementations
- Sustainable AI operating model

Partnership Advantage

OnStak + Netsync + Cisco Partnership: Combining Cisco's and Netsync's enterprise technology leadership and services with Onstak's AI governance expertise, leveraging Cisco's Responsible AI Framework and Secure AI Factory architecture.

Phased Approach

01 Assessment

Duration: 6 weeks

Focus: Governance Maturity & Use Case Ideation

- Current state AI inventory
- Governance maturity assessment
- Use case identification
- Organizational impact analysis
- Gap analysis & recommendations

02 Planning

Duration: 8 -11 weeks

Focus: Framework & Strategic Planning

- AI governance framework design
- Healthcare-specific policies
- AI Center of Excellence design
- Use case prioritization
- Performance monitoring framework

03 Execute

Duration: 12-14 weeks

Focus: Tooling & Use Case Deployment

- Governance tooling implementation
- Proof-of-concept development
- Knowledge transfer & training
- Sustainability planning
- Future roadmap development

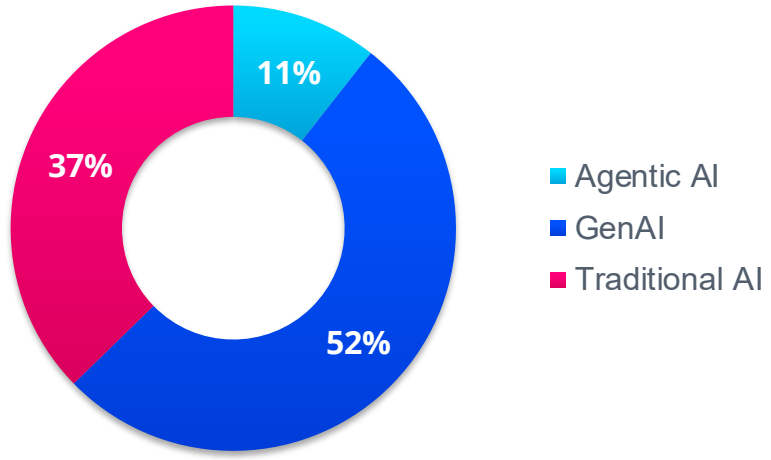
Use Case Discovery

236 Unique Opportunities Identified

Business Unit Distribution

- 83 (35%)
- 39 (17%)
- 19 (8%)
- 95 (40%)

Distribution by AI Type



Co-Delivering Value with Cisco Partners

Our AI Expertise

Comprehensive AI capabilities

Cisco Infrastructure

Trusted, scalable foundation

Customer Success

Measurable, sustainable value

Cisco Partners

Local expertise & relationships

This partnership combines our AI expertise, Cisco's infrastructure, and local partners' market understanding. Together, we deliver results greater than any single entity.

Let's Unlock AI Value Together

Unlocking tomorrow, Together

Proven Track Record

Demonstrated AI success with measurable outcomes.

Successful U.S. Partnership

Proven collaboration with Cisco, ready for global scaling.

European Opportunity

Significant potential to adapt our proven approach for Europe.

> AI unlocks new value and empowers organizations to achieve real world outcomes.

Let's collaborate to deliver measurable, AI-driven value. Unlock an intelligent, outcome-focused future together.

Start the Conversation

Learn More

cisco

Thank you