

# Cisco AI PODs for Collaboration

## Secure, On-Premises AI for Sovereign Environments

Bring enterprise AI capabilities to your collaboration infrastructure—without compromising data sovereignty.

### Contents

Product overview .....	2
Features and benefits.....	2
Prominent features .....	2
Product specifications.....	3
Platform support .....	4
Licensing.....	4
Integrations and ecosystem .....	5
Ordering information .....	5
Cisco Capital.....	6
Document history .....	6

## Product overview

**Cisco® AI PODs for Collaboration** (also known as **Collaboration AI PODs**) deliver production-grade AI services as a turnkey on-premises appliance, purpose-built for defense, government, and regulated enterprise customers who cannot leverage cloud-based AI due to compliance, data sovereignty, or security requirements.

Cisco AI PODs for Collaboration provide real-time transcription services for Cisco collaboration solutions (refer to Table 5), enabling customers to unlock the productivity benefits of AI while keeping sensitive data within their controlled environment. With support for air-gapped deployments and no external connectivity requirements, organizations can deploy AI-powered transcription in even the most restricted networks.

## Features and benefits

Table1. Features and Benefits

Feature	Benefit
<b>Data Sovereignty</b>	All data processed and stored on premises; no cloud dependency or external data transmission
<b>Air-gap Ready</b>	Deploy in fully isolated networks with zero external connectivity requirements, except licensing
<b>Enterprise Scale</b>	Support for 700–800 concurrent transcription sessions per Collaboration AI PODs with multi-Collaboration AI PODs load balancing and fault tolerance
<b>Low Latency</b>	~600 ms latency for real-time transcription
<b>Ecosystem Integration</b>	Software Development Kit (SDK)-enabled integration with Cisco solutions

## Prominent features

- **Real-Time Transcription:** Automatic Speech Recognition for live closed captions/transcription with near-instant results. Available in five languages; English, French, Spanish, Italian, and German.
- **Secure by Design:** Transport Layer Security (TLS) 1.3 encryption, API key authentication, encrypted storage.
- **Multi-POD Architecture:** Scale horizontally with intelligent load balancing across multiple appliances, no clustering required.
- **Load Balancing and Fault Tolerance:** The Collaboration AI PODs SDK-implemented by integrations-features intelligent, built-in load balancing that automatically distributes sessions across multiple PODs using a least load and top-down approach, ensuring optimal resource utilization, seamless scalability, and continuous service availability without manual intervention.

- **Logging and Analytics:** Comprehensive logging for efficient troubleshooting and analytics to track utilizations/usage.
- **Stateless Operations:** All closed captions data is processed in-memory only and permanently discarded at session end—nothing is stored, logged, or recoverable between sessions.
- **User Experience:** Transcription accuracy is comparable to Webex®; however, as with all AI-based solutions, results may vary due to the inherent non-deterministic nature of AI models.

## Product specifications

Table 2. Product Specifications

Specifications	Details
<b>Management</b>	<ul style="list-style-type: none"> <li>▪ Backup and restore</li> <li>▪ Email notifications (via Simple Mail Transfer Protocol [SMTP])</li> <li>▪ Alerts management</li> </ul>
<b>Security and auditing</b>	<ul style="list-style-type: none"> <li>▪ All control data is authenticated and encrypted (industry-standard TLS1.3)</li> <li>▪ Encrypted storage</li> <li>▪ Extensive audit logging</li> <li>▪ Unified Extensible Firmware Interface (UEFI) Secure boot</li> <li>▪ Certificate Key Types/Algorithms <ul style="list-style-type: none"> <li>- RSA (2048, 3072 and 4096)</li> <li>- ECDSA (prime256v1 &amp; secp384r1)</li> </ul> </li> </ul>
<b>Diagnostics and Troubleshooting</b>	<ul style="list-style-type: none"> <li>▪ Syslog – local and remote</li> <li>▪ Log bundle (Encrypted) <ul style="list-style-type: none"> <li>- System logs</li> <li>- System configuration</li> <li>- Application configuration</li> <li>- System information</li> <li>- Crash logs</li> </ul> </li> </ul>
<b>High Availability</b>	<ul style="list-style-type: none"> <li>▪ No clustering required, achieve redundancy and fault tolerance by adding additional PODs</li> </ul>
<b>Account management and permissions</b>	<ul style="list-style-type: none"> <li>▪ Local accounts (with Admin and Support roles)</li> </ul>

## Platform support

Cisco AI PODs for Collaboration (Collaboration AI PODs) software is optimized to run on Cisco Unified Computing System (Cisco UCS®) C240 Series AI PODs with reference configuration (refer to the ordering information section).

In terms of scale, Table 3 shows the capacity of real-time transcription service running on UCS C240 Cisco AI PODs.

Table 3. Real-Time Transcription Capacity

Language	Max concurrent sessions*
English	800
Non-English (Spanish, Italian, French, German)	700

- **\*Scalability:** Capacity defaults to the lowest-rated language when multiple languages are enabled.
- **\*Concurrency:** Session maximums are non-configurable, tested limits. Scale out by adding PODs if requirements exceed published values.
- **\*Throughput:** Real-time transcription supports a high-throughput capacity of up to 80 concurrent sessions per minute, ensuring reliable performance at scale.

## Licensing

Simple, Concurrent, and Predictable

**Cisco AI PODs for Collaboration (Collaboration AI PODs) use a straightforward concurrent session model. You license concurrent capacity—not consumption.**

One Shared AI License = One Concurrent Session.

Collaboration AI PODs connects to Cisco Smart licensing to simplify purchasing, registering, and monitoring licenses. Smart Licensing is supported via direct mode only.

Table 4. License Definitions

Term	Definition
Session	Any single meeting or call (live or recorded) being processed with AI features
Coverage	Real-time transcription

**Example:** With one license, run unlimited meetings throughout the day—one at a time. Need two meetings with AI features running simultaneously? Add a second license.

## Integrations and ecosystem

Cisco AI PODs for Collaboration integrate directly with Cisco collaboration solutions, delivering AI capabilities where your meetings and calls happen—without requiring data to leave your premises.

Table 5. Integrations

Release version	Integration
Version 1.0	Cisco Meeting Server v4.0 – Real-time transcription for on-premises meeting deployments

### Expanding the Ecosystem

Additional Cisco solutions and certified partner integrations are on the roadmap, with SDK access enabling AI across meetings, recordings, and contact center workflows.

## Ordering information

Below are the two main steps for ordering Collaboration AI PODs hardware and software licenses. To place an order, visit the [Cisco Ordering Home Page](#).

1. Select your preferred hardware platform.
  - a. Cisco AI PODs for Collaboration (Collaboration AI PODs) software is optimized to run on Cisco UCS C240 AI PODs spec-based configuration. No third-party servers are currently supported.
    - i. Under [AI Services POD](#), pick Collaboration AI PODs
      1. Pick your preferred server (with GPU or without GPU), description in Table 6.
      - ii. Alternatively, directly order **AIPOD-COLLAB SKU** and pick from hardware choices mentioned in Table 6.

**Note:** The Real-time transcription feature (or v1.0 of Collaboration AI PODs) does not require a GPU to function. A GPU can be added at any time as an optional, plug-and-play upgrade.

2. Select Software subscription license.
  - a. Collaboration AI PODs licenses are available as a subscription, under A-CMS top level SKU with monthly or yearly billing which includes Cisco Software Support Services (SWSS) bundled together.
    - i. Under **A-CMS** SKU configuration, section **Cisco AI PODs for Collaboration**, provide the max concurrent sessions count.

Table 6. Collaboration AI PODs Hardware/Software PIDS

Part #	Product Description
<b>AIPOD-COLLAB</b>	Cisco AI PODs for Collaboration (Hardware)
<b>UCSC-C240-M8-CL-G</b>	UCS C240 M8 2U Rack Server for Collaboration AI (GPU)
<b>UCSC-C240-M8-CL</b>	UCS C240 M8 2U Rack Server for Collaboration AI
<b>A-COLLAB-AIPOD-SAL</b>	Collaboration AI PODs Shared AI License (Software License)

## Cisco Capital

**Flexible payment options make it easier than ever to get the Cisco technology you need.**

Cisco Capital® delivers leading-edge payment solutions, allowing you to stay focused on what's most important—your business. We can help you drive business outcomes, accelerate innovation and digital transformation, and adapt to market dynamics faster with flexible payment options tailored to your specific business needs. Reduce the total cost of ownership, conserve capital, and accelerate growth. We help you realize the full benefits of Cisco technology today, and in the future, and pay for it in the way that best suits your business requirements. Whether you are looking for a pay-as-you consume model, or need to bundle Cisco hardware, software, services, subscriptions, and third-party solutions, [learn more](#) about how Cisco Capital can help.

## Document history

Table 7. Document history

New or Revised Topic	Described In	Date
<b>Initial Version</b>	-	May 30, 2026