AMD Keynote

Dr. Lisa Su



OCP GLOBAL SUMMIT

October 12-15, 2026
San Jose, California


AMD
Solving The World's
Most Important Challenges
Accelerating AI from Silicon to Solutions

AMD × OPEN Compute Project®

15 Years of Collaboration

Open Server Platforms → Open AI Systems → Open AI Infrastructure

2012

AMD OPEN 3.0 MODULAR SERVER SPECIFICATION



2017

"NAPLES" + MICROSOFT PROJECT OLYMPUS



2023

ULTRA ETHERNET CONSORTIUM FORMS

Ultra Ethernet Consortium

2024

YOSEMITE V4



ULTRA ACCELERATOR LINK CONSORTIUM FORMS

ULTRA ACCELERATOR LINK™

2025

AMD HELIOS DEBUTS AT OCP

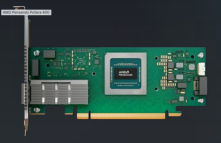


ESUN COLLABORATION FORMS

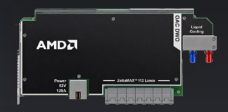
ESUN

2026

MULTIPATH RELIABLE CONNECTION



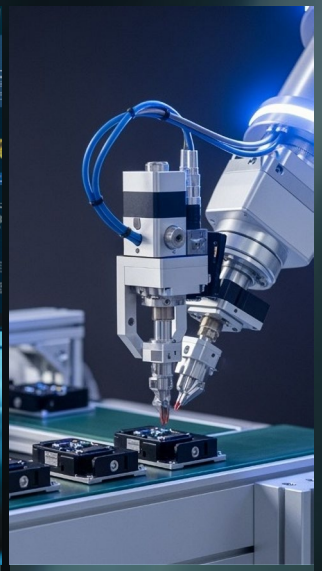
OPEN ADAPTER CARD



AI Is Transforming Every Industry



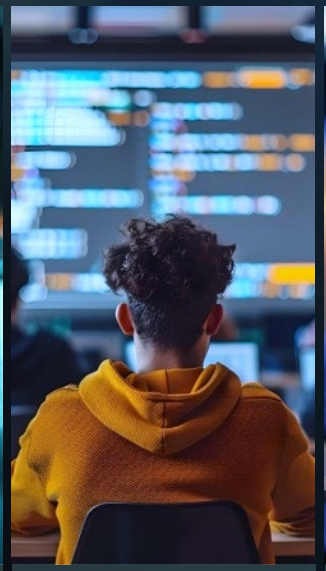
HEALTHCARE



MANUFACTURING



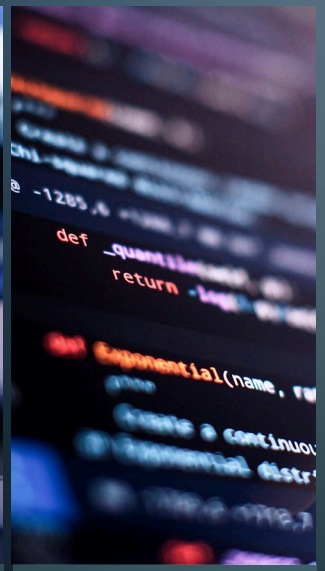
FINANCE



EDUCATION



SCIENCE



SOFTWARE



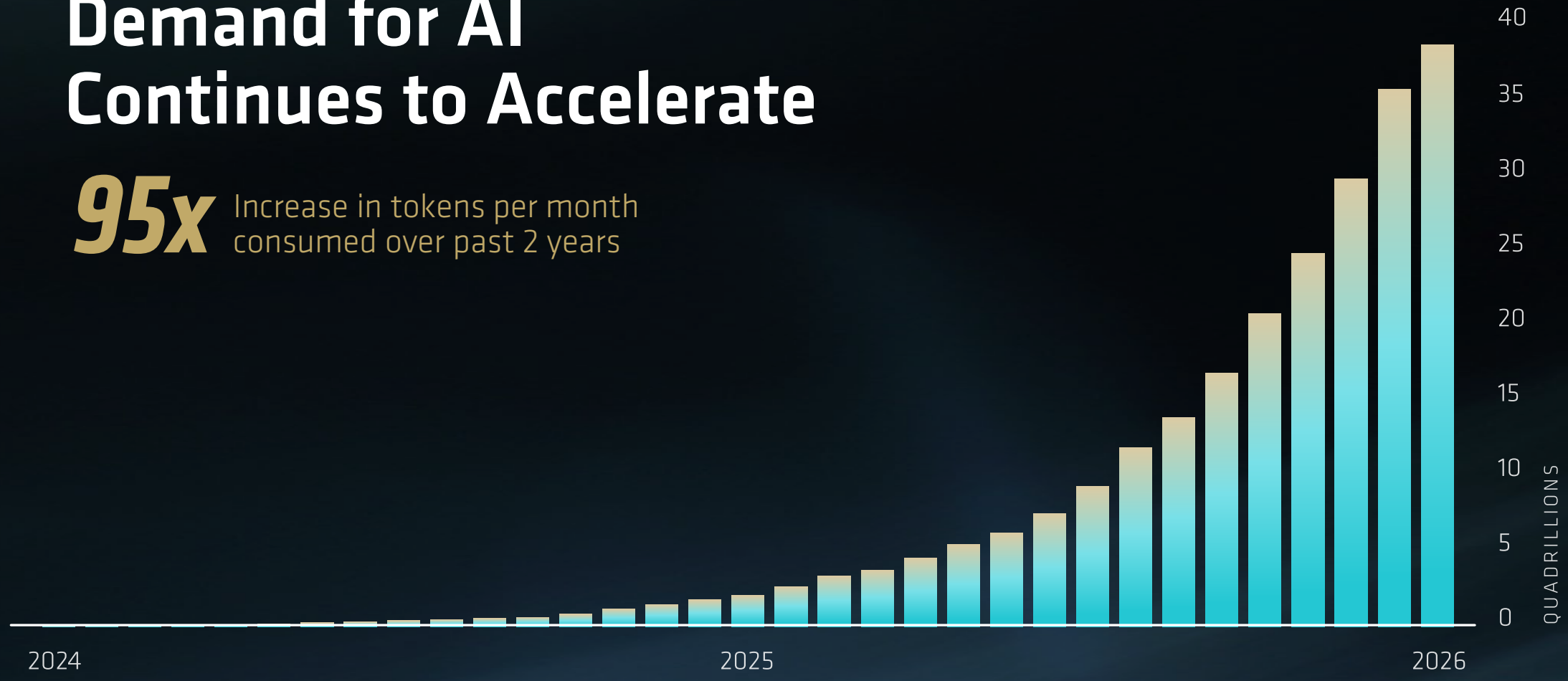
AGRICULTURE



MEDIA & ENTERTAINMENT

Demand for AI Continues to Accelerate

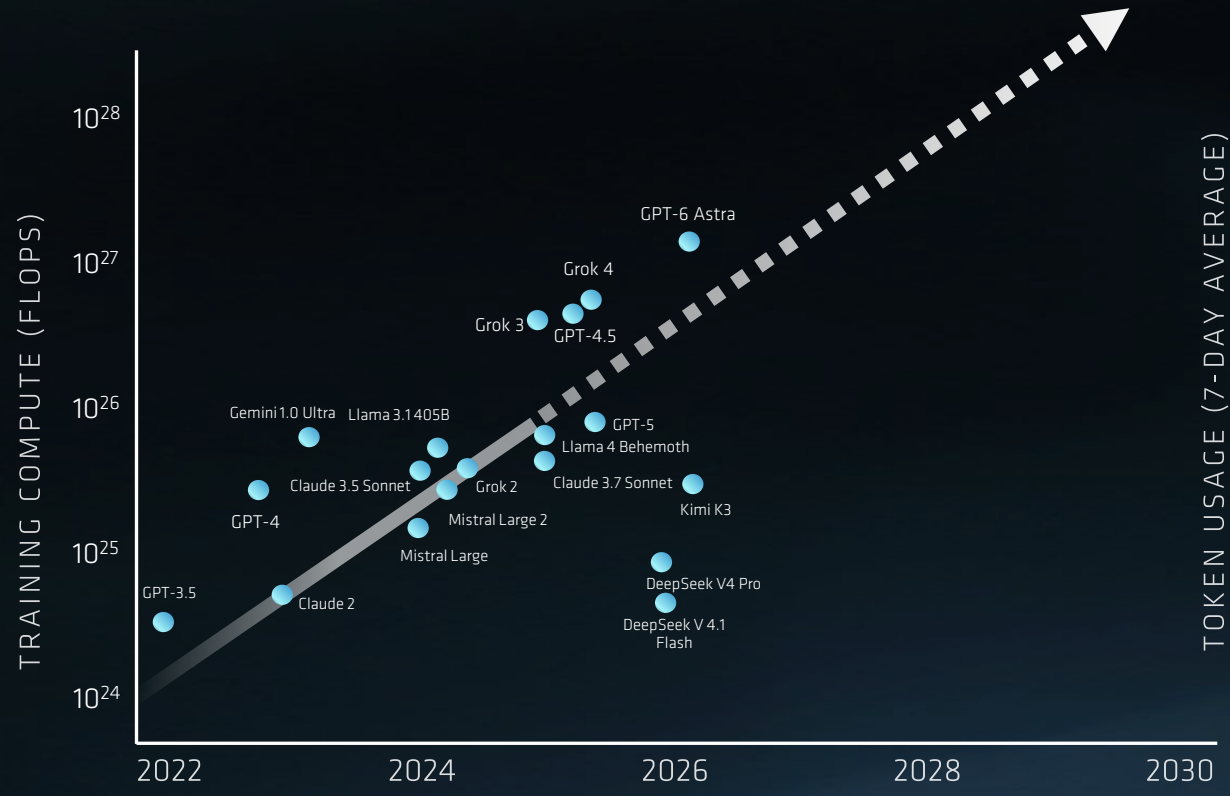
95x Increase in tokens per month consumed over past 2 years



Exponential View, State of the AI Economy, Sept 30, 2026

Training Continues to Scale

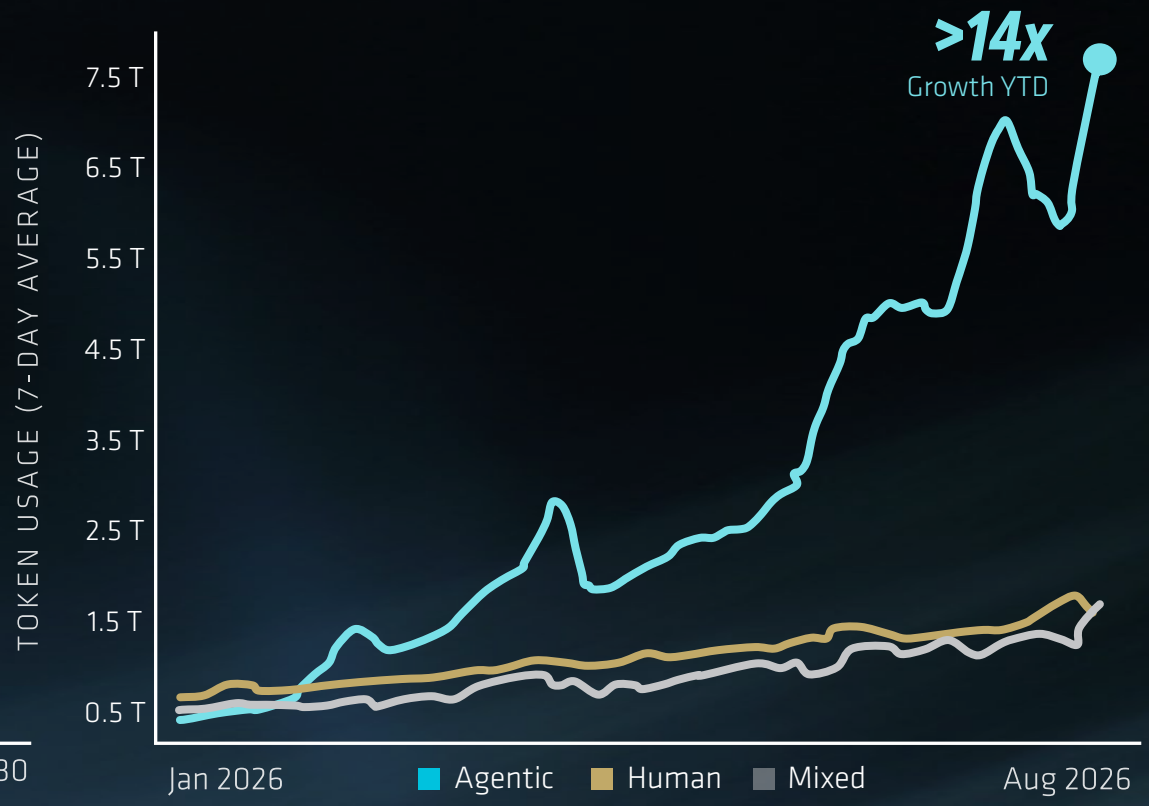
TRAINING FLOPS INCREASING 5X EVERY YEAR SINCE 2020



Source: Epoch AI, Trends in Artificial Intelligence, October 5, 2026

But Inference is Now the #1 AI Workload

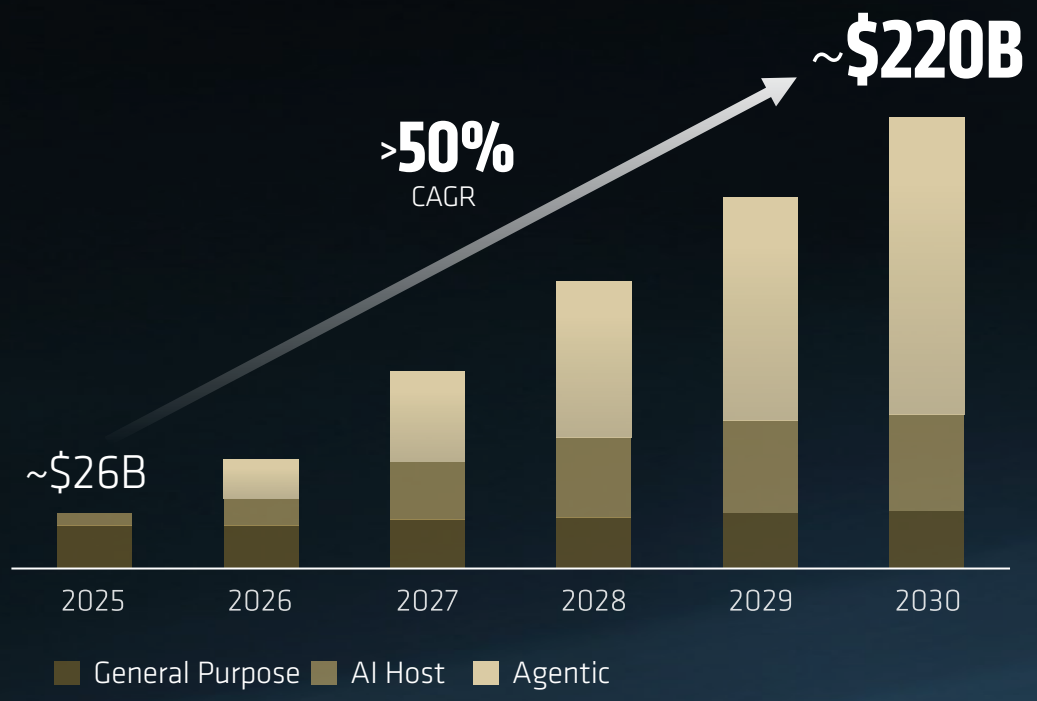
DRIVEN BY ACCELERATED AGENTIC ADOPTION



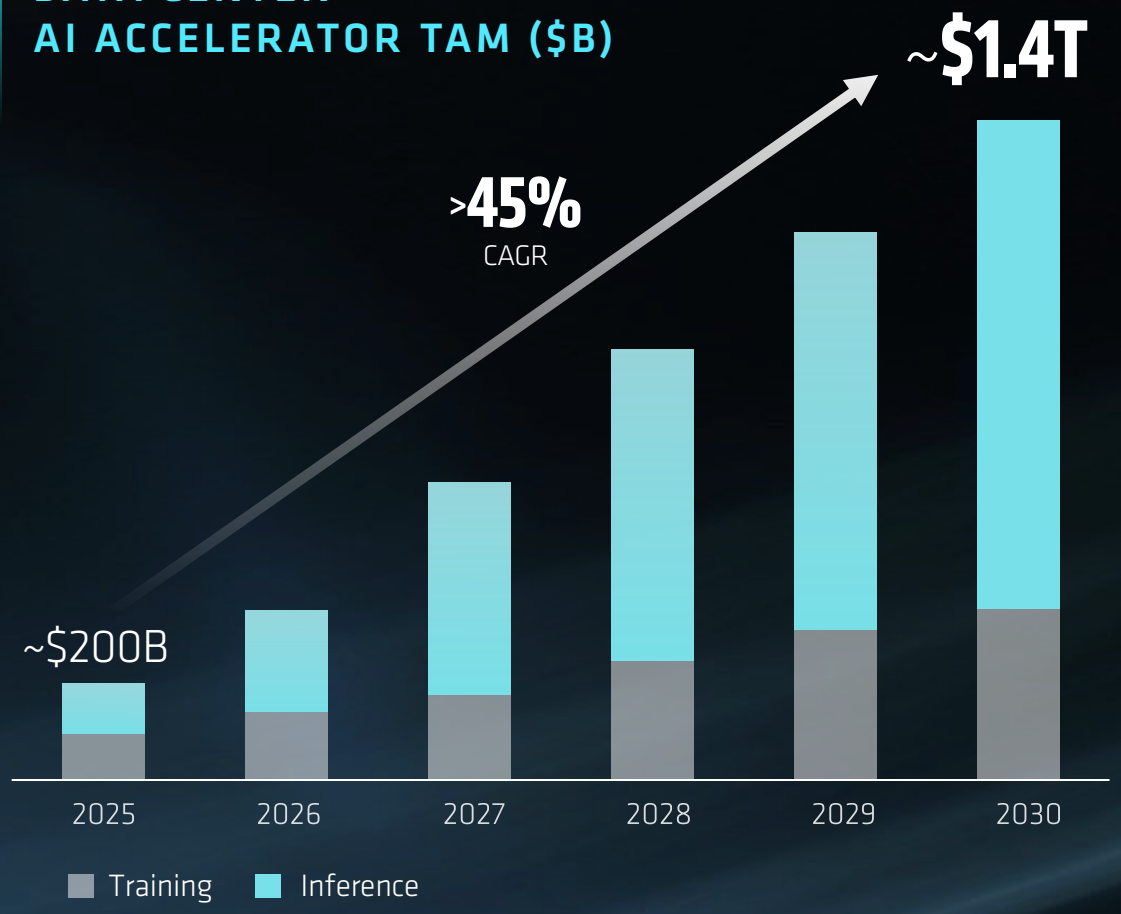
Source: Openrouter.ai/ratings

Demand Accelerating Across CPUs and GPUs

DATA CENTER CPU TAM ($B)



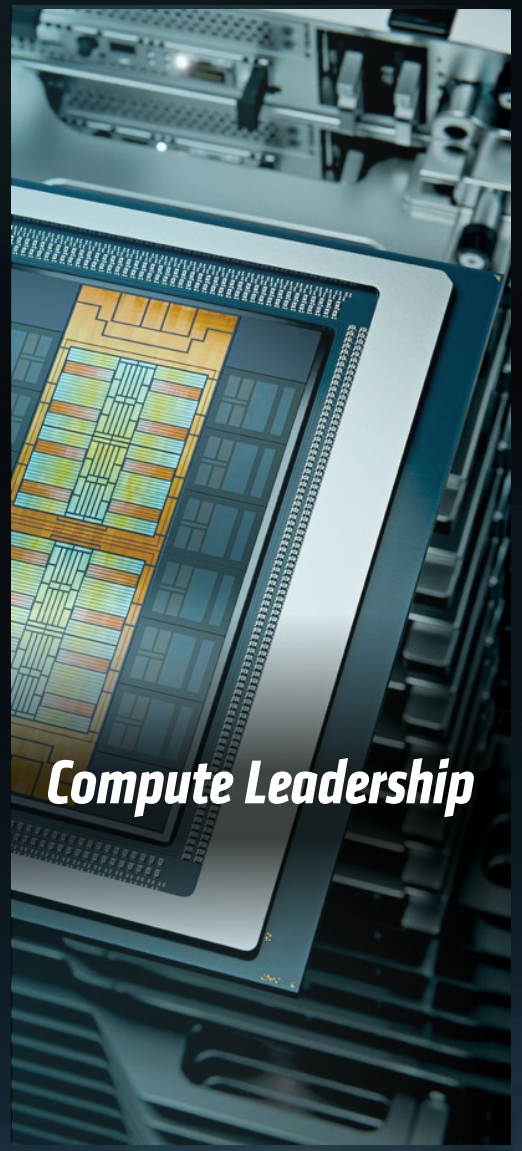
DATA CENTER AI ACCELERATOR TAM ($B)



Source: AMD Internal Projections

AMD

Strategy







AMD × OCP GLOBAL SUMMIT

Collaboration at Scale



Rack-Scale Innovation

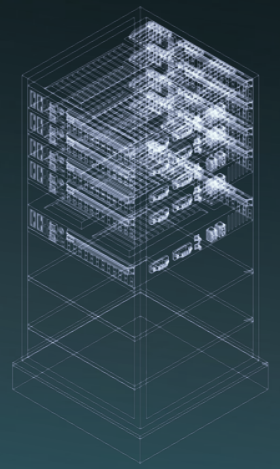


Agentic Infrastructure

Frontier AI Requires Integrated Rack Architectures



Leadership Compute Capability



Open Rack Architecture

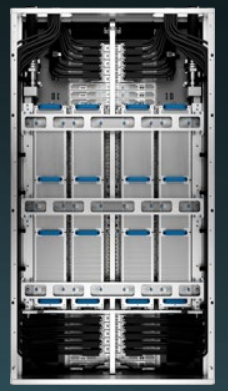
AMD PENSANDO

Ultra Ethernet Consortium

ULTRA ACCELERATOR LINK™

ESUN

Open Fabric to Scale AI



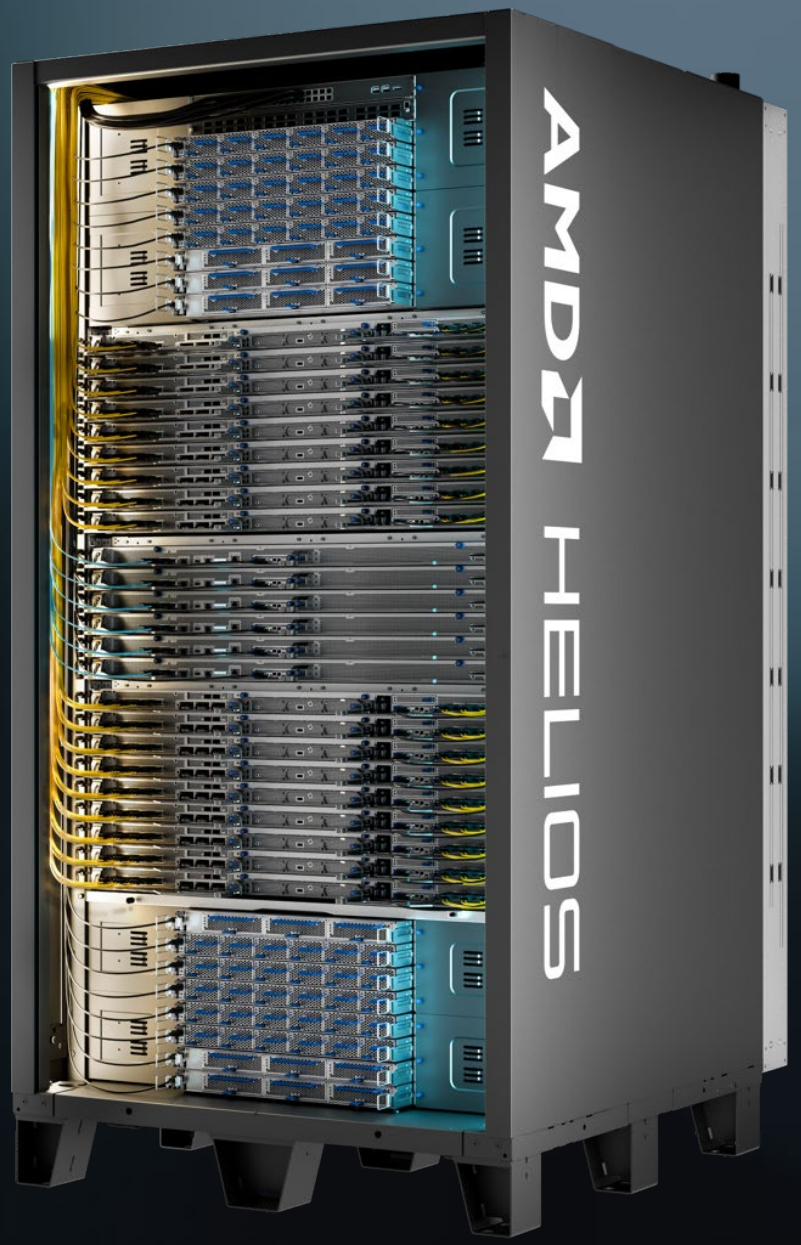
Rack Scale Efficiency & Serviceability



Turnkey Solutions

AMD HELIOS

The Highest Performance Rack-Scale AI Solution



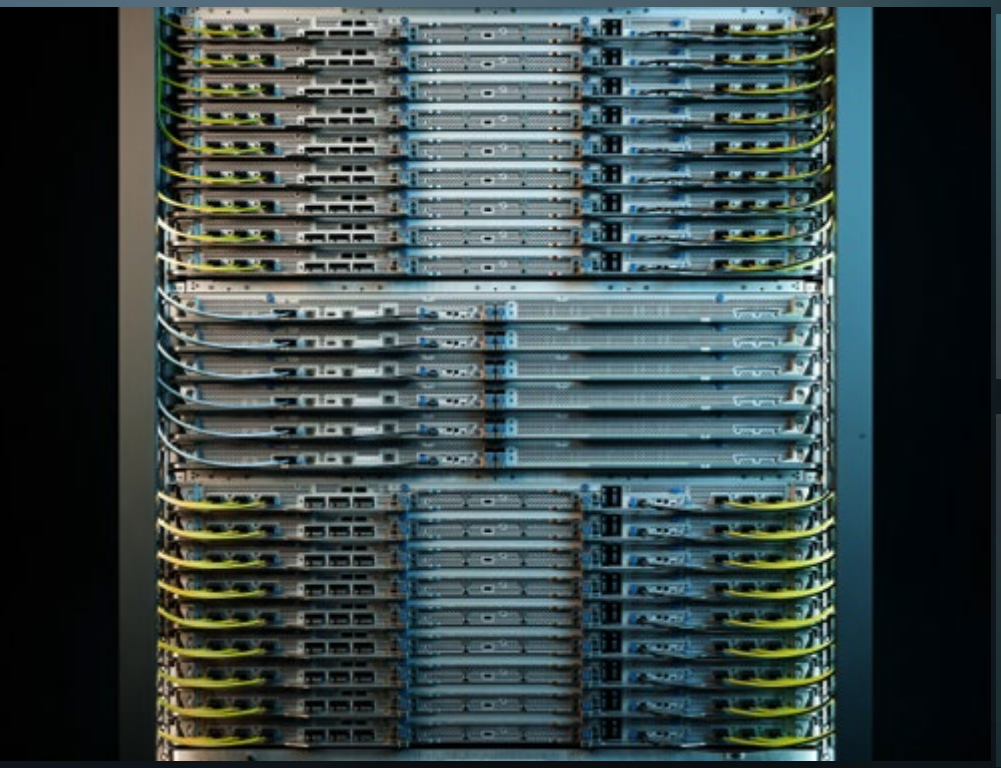
Helios: High-Performance Rack
Co-Engineered with AI Leaders



AMD +	Meta +	OPEN Compute Project®
Open Rack Wide	OCP Power Shelves	OCP GPU Management

Helios: High-Performance Networking

Delivered Through An Open Ecosystem



Open + Performant + Programmable

FRONT-END
Security, Storage, SDN Acceleration

SCALE-UP
UALoE, ESUN

SCALE-OUT
UEC, MRC, MetaRoCE

BROADCOM CISCO ARISTA HPE DRIVENETS

ROCm: High-Performance Software
Powered By An Open Ecosystem

OPEN SOURCE COMMUNITY ENABLING AMD



AMD Helios™ Delivers Leadership Rack-Scale Solution

#1 COMPUTE
2.9 Exa Flops FP4
1.48 Exa Flops FP8

#1 MEMORY
31 TB HBM4 Capacity
1.7 PB/s HBM4 Bandwidth

#1 NETWORKING
260 TB/s Scale Up Bandwidth
43 TB/s Scale Out Bandwidth

DELIVERS

Up to 30% More Tokens/$

On leading AI Models

17,000 tokens/s/GPU

DeepSeek R1 Inference (MLPerf)

AMD measured MLPerf offline config

Helios Deployments Underway at Leading AI Customers

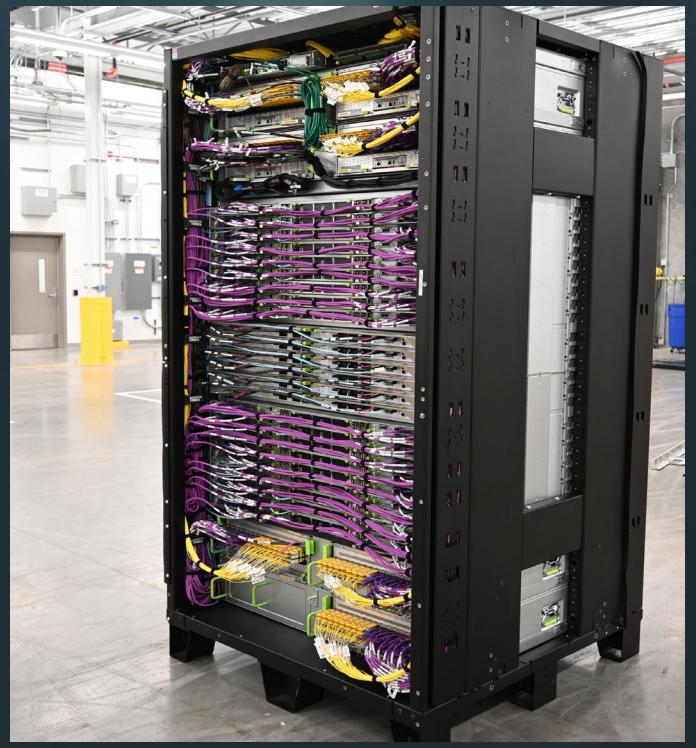
VIDEO

Helios Deployments Underway at Leading AI Customers

OpenAI Meta ANTHROP\C Microsoft ORACLE

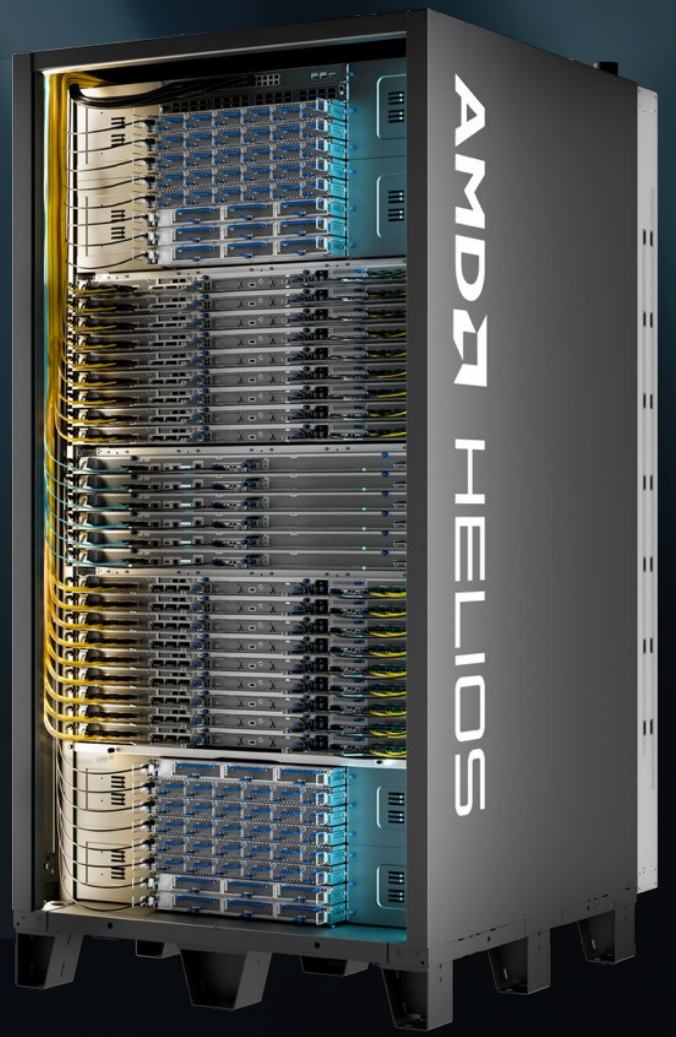






AMD Helios
Growing Adoption by AI Leaders

OpenAI
Meta
ANTHROPIC
Microsoft
ORACLE
together.ai
Fireworks
river
ODYSSEY
DELL Technologies
HPE
Lenovo
SUPERMICRO
Bull
amp PBC
HUMAIN THE END OF LIMITS.
IBM
TENSORWAVE
VULTR
Crusoe
Alignment Engine
Cirrascale
CORVEX
OneQode
neysa
NATIONAL COMPUTE
Fluidstack
SANMINA
wiwynn



AMD × OCP GLOBAL SUMMIT

Collaboration at Scale



Rack-Scale Innovation



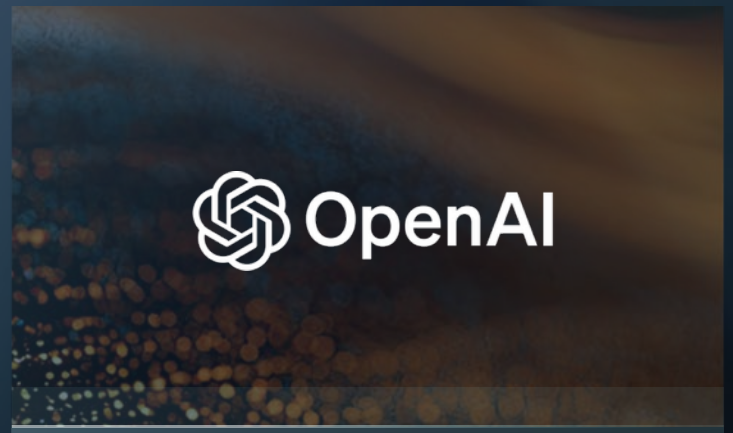
Agentic Infrastructure

Entering an Era of High-Volume Agentic Usage



5 Million

Downloads of Meta Muse
in first few weeks



10,000

Agents to solve the Navier–Stokes
Millennium Prize Problem



30,000

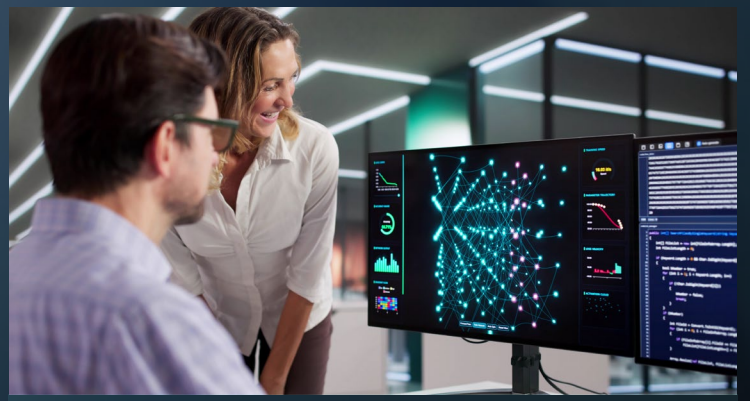
R&D Agents building the
next Frontier Model

Agents Need the Right Compute



PERSONAL AGENTS

Dense Servers
(Max Agents per Watt)



ENTERPRISE AGENTS

Balanced Servers
(Max # of Performant Agents)



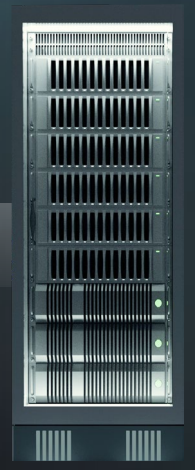
TECHNICAL COMPUTE AGENTS

Performance Servers
(Max Agent Performance)

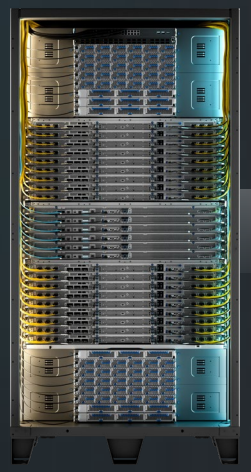
And Agents Also Need the Right Rack

HYPERSCALE
Disaggregated Agentic Rack

CPU



+



GPU

ENTERPRISE
All-in-One Agentic Rack



CPU
+
GPU

ANNOUNCING TODAY

AMD Helion Agentic AI Rack

Open & Flexible Rack-Scale Compute for Agentic AI

Match your CPU compute to your agentic workloads

Customize your mix of CPU core density, performance and power

Choose any combination of CPU & GPU nodes

CPU-only for hyperscale deployments & all-in-one CPU/GPU Agentic racks for enterprises

Integrated token router

Intelligently route your agents model calls across frontier and open source models

Open hardware & open software

OCP compliant hardware components



AMD Helion Agentic AI Rack

Up to 32,768 Zen 6
x86 CPU Cores

Up to 16 MI350P
74 PetaFlops - FP4

Up to 102 Tbps Networking
Fully Programmable



ORv3 44OU Rack
Native OCP Spec Compliant

Standard Modular Compute Board
Native OCP Spec Compliant

Open-Source Ecosystem Stack

AMD Helion Agentic Rack – Hyperscale Config

Rack-Scale Agentic Infrastructure (CPU Only)

PERSONAL AGENTS
Dense Servers
“Venice” 256c @ 400W

ENTERPRISE AGENTS
Balanced Servers
“Venice” 256c @ 600W

TECHNICAL AGENTS
Performance Servers
“Venice” 96c High Frequency



Mix and Match CPU Nodes
to Support Diverse Agentic Workloads

AMD Helion Agentic Rack – Enterprise Config

All-in-One Enterprise Agentic AI Rack (CPU+GPU)

AGENTIC NODE – CPU "SANDBOXES"

Dense – "**Venice" 256c @ 400W**
Performance – "**Venice" 96c @ 5.0GHz**

INFERENCE NODE – GPUs

Performance - **Instinct MI350P**

GENERAL PURPOSE NODE

Database, Storage – **"Venice" SP7, SP8**
Technical Compute – **"Venice-X"**



AVAILABLE NOVEMBER ‘26

AMD Helion Agentic Rack

Open & Flexible Rack-Scale Compute for Agentic AI

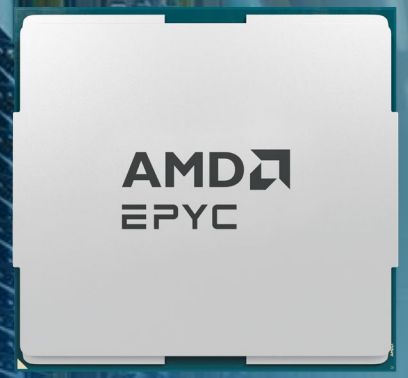
CONTRIBUTING TO OPEN Compute Project®

DELL Technologies · HPE · Lenovo · SUPERMICRO · CISCO

wiwynn® · QCT™ · FOXCONN HON HAI TECHNOLOGY GROUP · MiTAC MiTAC Computing

Inventec · ASUS · ASRock Rack · GIGABYTE™





EPYC "Venice"

The Best Server CPU Portfolio for Agentic Racks

GPU SERVERS "HOST NODE"	AGENT CPU SERVERS "SANDBOXES"	GENERAL PURPOSE CPU SERVERS		
"Venice" HF RDIMM / MRDIMM **"Verano"** LPDDR	**"Venice" 256c**	**"Venice" SP7**	**"Venice" SP8**	**"Venice-X"**
High Frequency CPU to GPU Bandwidth	High Core Count Low Power	High Performance High Density	Balanced Performance & Power	Technical Compute & HPC

See Endnote: EPYC-067

New EPYC “Venice” with LPDDR

AMD EPYC Venice 256c

Highest Density

AMD EPYC Venice HF

Highest Perf/Core

AMD EPYC Venice SP8

Balance Perf/Power



“Verano”

Lowest System Power

Fastest CPU to GPU Bandwidth

Up to 192c

AVAILABLE IN 1H 2027

See Endnote: EPYC-067

Open Ecosystem Drives Innovation

Open Hardware + Open Software + Open Ecosystem





Collaboration at Scale

Delivering the Next Era of AI Infrastructure

- Leadership Components
- Flexible, Open Platforms
- Integrated Rack Solutions


AMD


OCP
GLOBAL
SUMMIT
Scaling Innovation for the AI Era.
October 12-15, 2026 | San Jose, California