AMD
together we advance_

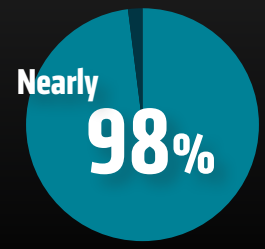
MAXIMIZE AI PERFORMANCE AND SCALABILITY WITH AMD

Unlock performance, efficiency, and innovation with the industry's broadest AI compute portfolio

THE NEED FOR ENTERPRISE AI TRANSFORMATION

Enterprises are under increasing pressure to optimize their IT infrastructure to meet the growing demands of AI workloads. This is the top priority identified by buyer research, ranking higher than sustainability and cybersecurity.

While CIOs and ITDMs look to go beyond AI pilots to enterprise-wide deployments that deliver significant business results, 70% of GenAI pilots fail to move into production due to inadequate data, lack of governance, high risk or inability to justify the business value[1].



of ITDMs reported implementing AI-ready infrastructure as the top priority when it comes to data center investments.[2]

THE AMD ADVANTAGE: AI SOLUTIONS BUILT FOR YOUR BUSINESS OUTCOMES

AMD is the trusted, open, and flexible partner that enables enterprise AI transformation. Our broad portfolio, open ecosystem, and comprehensive full-stack solutions help enterprises:

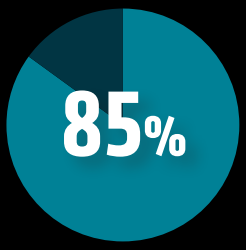
- Optimize compute for every workload
- Accelerate innovation
- Seamlessly scale AI initiatives to drive tangible business results and expand competitive advantage

> **AMD is no longer just catching up—it is forging a differentiated path with openness, developer-first engagement, and full-stack integration.[3]**
>
> **IDC**

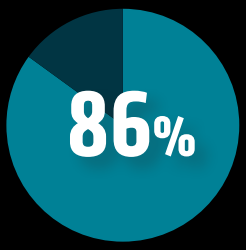
AMD
together we advance_

BROAD PORTFOLIO OF LEADERSHIP COMPUTE ENGINES

Enterprises must address core infrastructure gaps when scaling AI. ITDMs report insufficient GPU/CPU compute capacity, and a lack of future-ready infrastructure. This makes it imperative for them to modernize their data centers.



of ITDMs cite insufficient GPU or CPU resources as a major challenge across various aspects of their data centers.[4]



of ITDMs report inflexible infrastructure not designed for AI as a major challenge for supporting AI or ML workloads in their data centers.[5]

AMD has hardware to support AI wherever it runs, from data center to cloud, edge, or endpoint. It's a portfolio trusted at scale, adopted by AI leaders for production training and inference.

AMD EPYC™ CPUs	Scalable, energy-efficient, high-performance compute for AI data prep, inference, and orchestration at enterprise scale.
AMD Instinct™ GPUs	Accelerate AI training and HPC with breakthrough performance per watt and open-software flexibility.
AMD Pensando™ DPUs	Offload networking, security, and storage tasks to free CPU/GPU cycles for AI workloads.
AMD Versal™ Adaptive SoCs	Provide reconfigurable acceleration for edge AI, 5G, and embedded intelligence.
AMD Ryzen™ AI	Brings dedicated AI processing to the PC, enabling fast, efficient on-device inference.

AN OPEN ECOSYSTEM FOR ACCELERATED INNOVATION

Freedom to choose across hardware, software, and systems is critical to IT leaders. Nearly 93% of ITDMs report that security and compliance capabilities are the most important factors when evaluating new data center solutions, and AMD enables this flexibility.[6] AMD ROCm™ software is an open, performance-optimized software stack that gives developers the versatility to build, scale, and optimize their AI/ML workloads.

AMD
together we advance_

A CUSTOMER-FIRST APPROACH TO COLLABORATION

As a key contributor to open-standards groups such as the Open Compute Project and Ultra Ethernet Consortium, AMD helps drive industry-wide breakthroughs that benefit all.

The result: enterprises can accelerate innovation through openness and interoperability.

FLEXIBLE, FULL-STACK AI SOLUTIONS

AMD helps enterprises adapt and scale their infrastructure, from silicon to systems. Optimized, integrated platforms offer wide-ranging memory and core densities to process agentic and reasoning workloads. CPUs and GPUs work together to help lower latency, keeping cores and memory fully utilized. The AMD "Helios" AI Rack is a turnkey solution purpose-built for frontier training and distributed inference.

SOLUTIONS TO MINIMIZE DEPLOYMENT FRICTION

Enterprises can monitor their performance with AMD tools as they scale up their data centers, integrating new systems that are compatible with their existing infrastructure. The broad AMD portfolio delivers full-stack solutions to accelerate deployment and help enterprises scale AI fast.

- AMD Instinct™ MI355X GPUs accelerate model training by up to 3x vs. Instinct™ MI300X GPUs[7]
- AMD EPYC™ 9965 CPUs are the highest performance server CPU in a 2P configuration[8]
- AMD Ryzen™ AI PRO processors: The world's most powerful NPU for next-gen AI PCs designed for Enterprise[9]

READY TO SCALE YOUR ENTERPRISE AI TRANSFORMATION AND DELIVER BUSINESS VALUE?

DISCOVER HOW
WITH AMD ENTERPRISE AI SOLUTIONS.

[1]Gartner®, 1Q25 The CIO Report, 2025. GARTNER is a registered trademark and service mark and IT Symposium/Xpo is a trademark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and are used herein with permission. All rights reserved. [2]5-Point Likert Scale; Watchtower AI-powered research platform powered by Intercept, Global, n=161,242. [3]IDC, Advancing AI 2025: AMD's Full-Stack Infrastructure Vision, doc # US53602925, June 2025. [4]5-Point Likert Scale; Watchtower AI-powered research platform powered by Intercept, Global, n=147,984. [5]5-Point Likert Scale; Watchtower AI-powered research platform powered by Intercept, Global, n=453,097. [6]5-Point Likert Scale; Watchtower AI-powered research platform powered by Intercept, Global, n=45,592. [7]MI350-041: Based on AMD internal testing as of 6/5/2025. Using 8 GPU AMD Instinct MI355X Platform measuring text generated offline inference throughput for Llama4 Maverick chat model (FP4) compared 8 GPU AMD Instinct MI300X Platform performance with (FP8). MI355X ran 8xTP1 (8 copies of model on 1 GPU) compared to MI300X running 2xTP4 (2 copies of model on 4 GPUs). Tests were conducted using a synthetic dataset with different combinations of 128 and 2048 input tokens, and 128 and 2048 output tokens. Server manufacturers may vary configurations, yielding different results. Performance may vary based on use of latest drivers and optimizations. [8]As of 10/17/2025, a 2P server powered by EPYC 9965 processors is the highest scoring 2P server in SPECrate®2017_int-base published results with a score of 3230, www.spec.org/cpu2017/results/res2025q2/cpu2017-20250324-47086.html. SPEC® and SPECrate® are registered trademarks of Standard Evaluation Performance Corporation. Learn more at SPEC.org. [9]STXP-06: Based on AMD product specifications and competitive products announced as of Oct 2024. AMD Ryzen™ AI PRO 300 Series processors' NPU offers up to 55 peak TOPS. This is the most TOPS offered on any system found in enterprise today. AI PC is defined as a laptop PC with a processor that includes a neural processing unit (NPU).