THE AI-READY ENTERPRISE

How enterprises are modernizing IT infrastructure for AI

AMD
together we advance_

TABLE OF CONTENTS

INTRODUCTION

THE ENTERPRISE AI IMPERATIVE

There's little historical precedent for the rapid growth of AI, which has rapidly swept through the enterprise technology landscape. AI adoption is now a business imperative, with virtually every core function standing to benefit from fast decisions, increased productivity, and decreased operational costs.

But the rapid rise of AI is not a simple or seamless process for any organization. The challenges range from meeting new data center compute and energy demands, to aligning investments in resources with objectives around productivity gains or accelerated innovations.

This eBook explores how enterprises can overcome these challenges to build scalable, efficient, and cost-effective infrastructure that's ready for AI-driven modernization.

THE OBSTACLES TO AI ADOPTION

To realize the full potential of AI, enterprises must first recognize and confront the three main challenges in their path.

#1 RETHINKING INFRASTRUCTURE FOR AI

The first challenge is modernization: updating infrastructure to improve power efficiency, optimize space, and refresh aging assets.

Enterprises need to accommodate AI workloads, and to do so rapidly and cost-effectively. Yet these workloads demand significant compute resources. They can't simply be layered onto already crowded data center infrastructures, where energy demands are a constant concern and CPU resources, storage, and network bandwidth are operating at, or near, full capacity.

Existing or outdated infrastructure can also bring high opportunity costs; resources spent on upkeep can't be redirected to AI innovation.

#2 UNDERSTANDING WHERE THE ORGANIZATION IS ON THE AI ADOPTION SPECTRUM

The second challenge is understanding where the organization is on the AI adoption spectrum. While some enterprises are still focused on efficiency and automation, others are looking to drive innovation and create new business models. Understanding an organization's position requires visibility into where data resides, how it's structured, and how effectively it can be used across the enterprise.

Achieving high-performance AI often relies on heterogeneous hardware (e.g., CPUs, GPUs, and specialized accelerators) spread across multiple nodes and even multiple sites.

#3 OPENING THE DOOR TO AI EXPANSION WITH SMARTER ARCHITECTURE

The third challenge is building an open and flexible architecture that allows integration of multiple ecosystem components minimizing the need to rewrite code, hardware compatibility and open standards in software. AI readiness depends on how well companies can modernize their enterprise stacks to connect data sets, data flows and workflows to make them compatible for AI workloads.

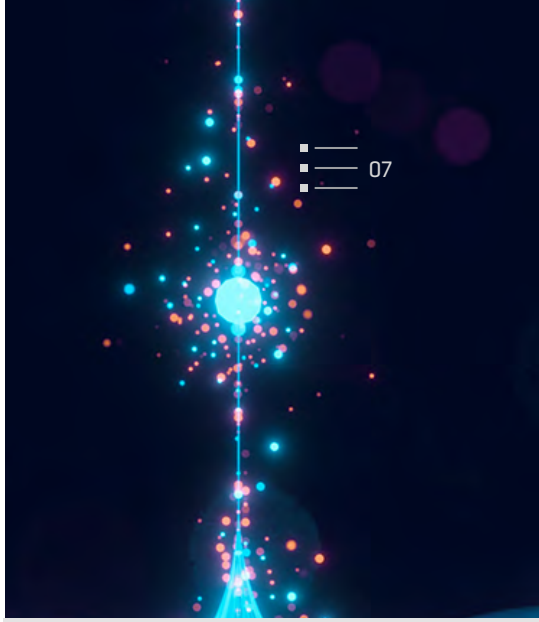
Balancing the portfolio of AI investments is a core strategic challenge. Advances in AI workloads, hardware, and software are arriving quickly, with each new generation of hardware or machine learning framework forcing an update to cost-performance calculations. Accelerating business outcomes with AI means addressing challenges through a series of targeted modernizations.

90% of leaders cite implementing AI-ready infrastructure as a key focus area for current data center investments.[1]

While **62%** identify security concerns related to AI data usage as a barrier to supporting AI and ML workloads.[2]

BUILDING THE FOUNDATION FOR AI MODERNIZATION

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



AMD EPYC™ SERVER CPUs

THE LEADING CPU FOR AI[3]

Many data centers are already running at or near capacity in terms of available space or power or both.[4] AMD EPYC™ server CPUs offer leadership performance and efficiency to enable material workload consolidation, allowing more space and energy to support new AI workloads in your existing data centers. For example, a server powered by two 192-core AMD EPYC™ 9965 CPUs delivers ~1.9x the inference throughput than 2P Intel® Xeon™ 6980P processors.[5*]

*Average runs/hour of 2P servers running the popular gradient boosting model XGBoost (Higgs Data Set) at FP32 precision

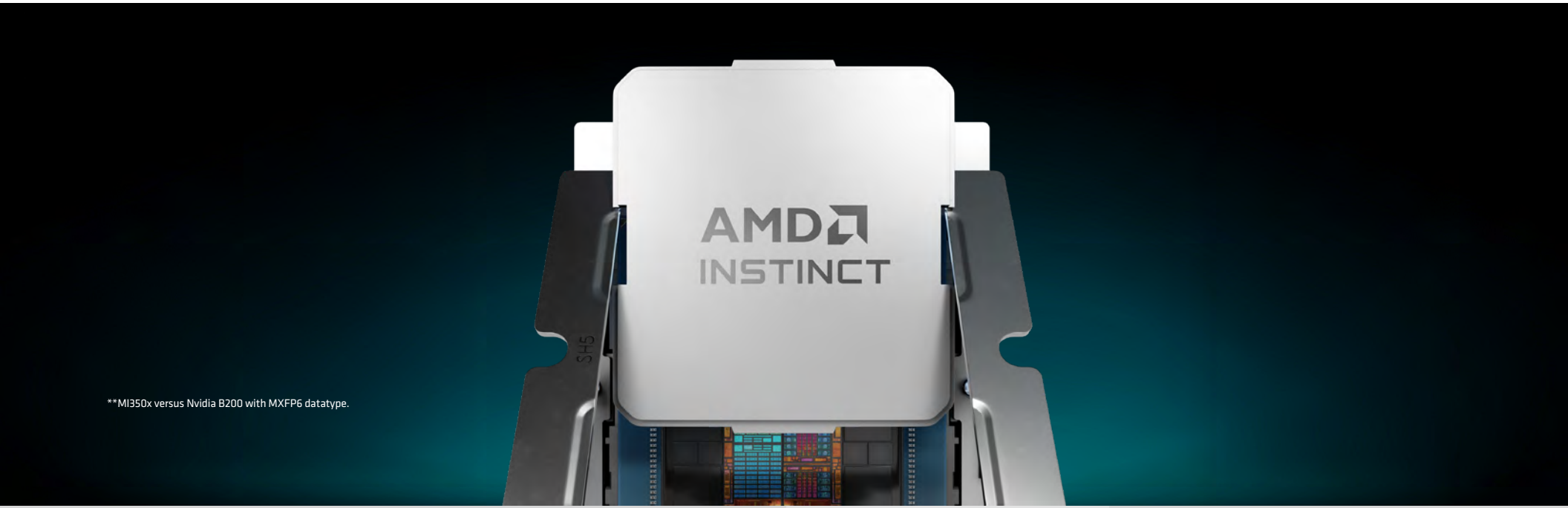
> **OUR AMD-BASED SYSTEMS ARE INCREDIBLY RELIABLE, ENSURING ZERO DOWNTIME IN MISSION-CRITICAL OPERATIONS.**

Ali Ray
Group Head of
Technology Platforms,
Emirates NBD

AMD INSTINCT™ GPUs

ACCELERATE AI TRAINING AND HPC WITH BREAKTHROUGH PERFORMANCE PER WATT

Built on 4th Gen AMD CDNA™ architecture, AMD Instinct™ MI350 Series GPUs deliver exceptional AI performance, with up to 2.2x the peak theoretical AI performance compared to competitive accelerators[6]**, massive 288GB HBM3E memory, 8TB/s bandwidth, and expanded datatype support including MXFP6, and MXFP4.

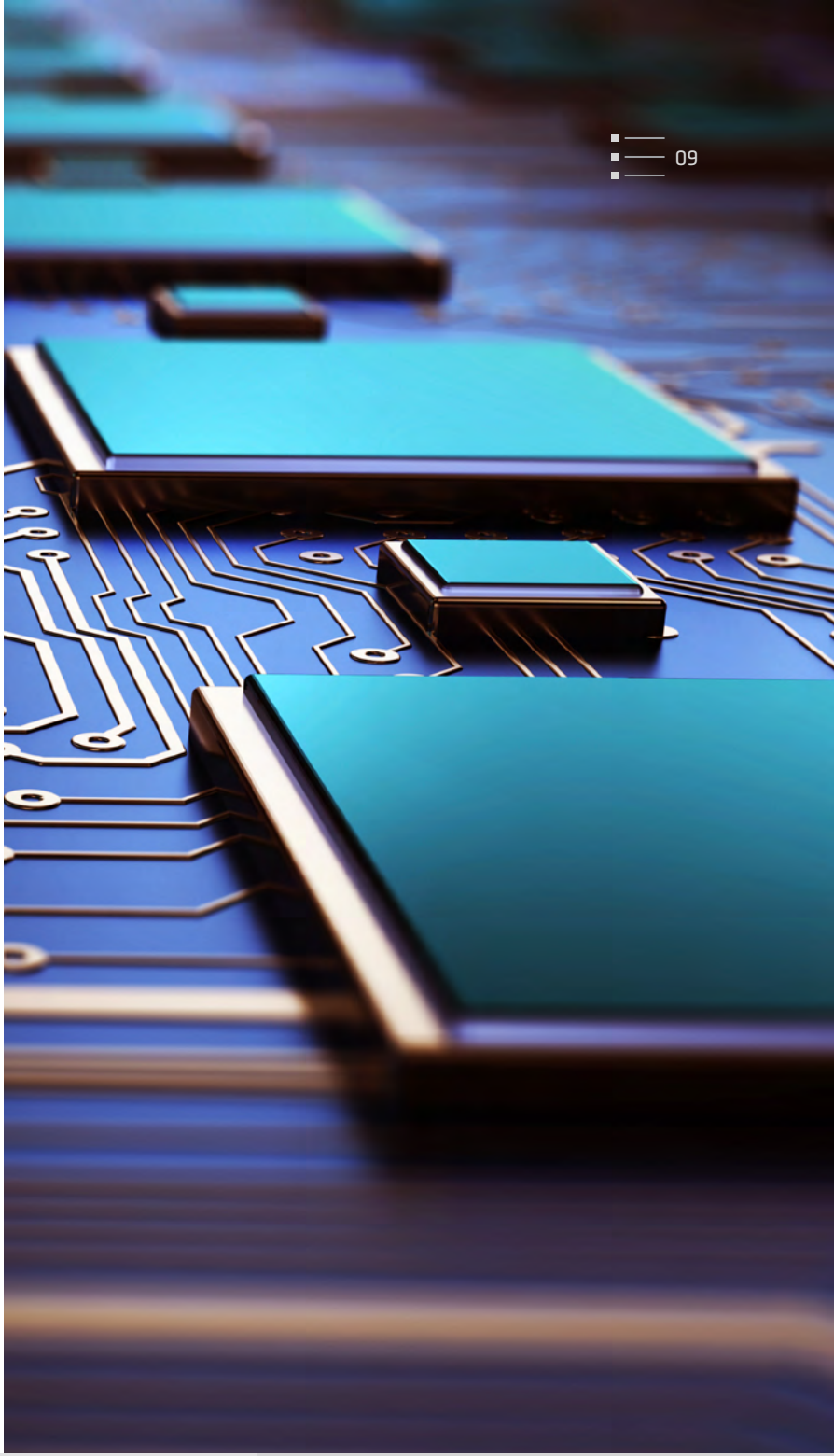


**MI350x versus Nvidia B200 with MXFP6 datatype.

AMD ROCm™ SOFTWARE

OPTIMIZED GPU SOFTWARE STACK

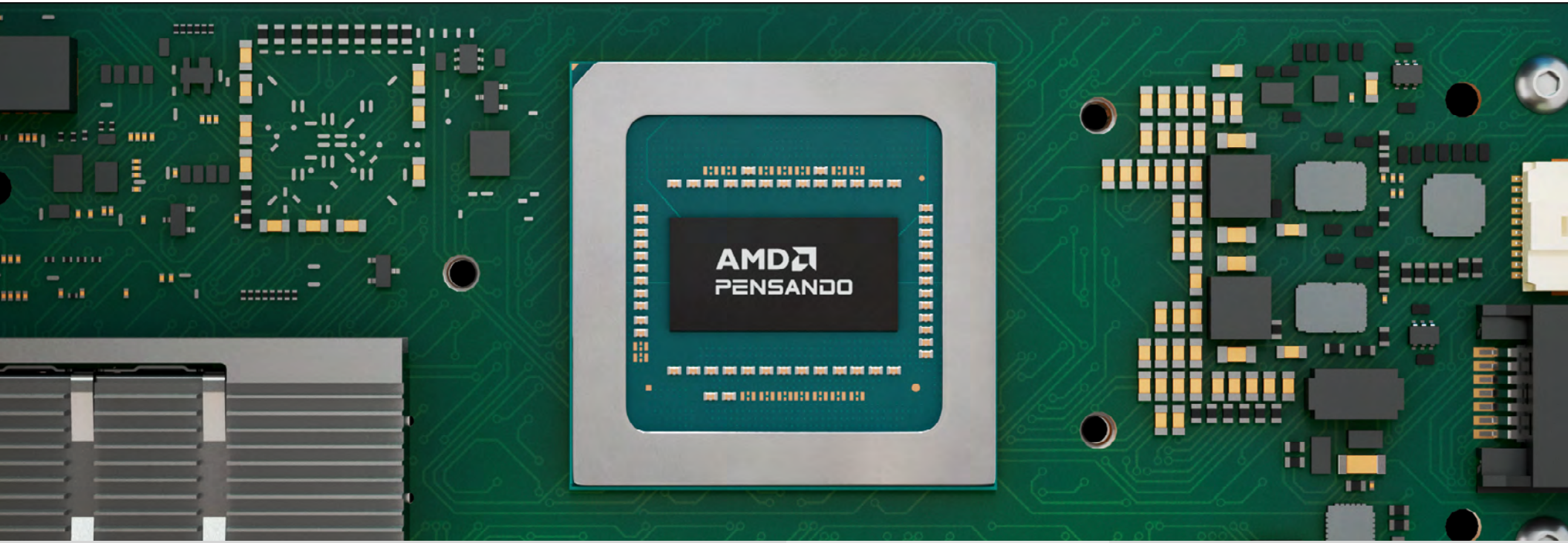
AMD ROCm™ is an open software stack including drivers, development tools, and APIs that enable GPU programming from low-level kernel to end-user applications. Developers can now train and deploy large models locally, enhancing privacy, performance, and cost efficiency while reducing cloud dependency. AMD ROCm™ offers a suite of optimizations for AI workloads and supports the broader AI software ecosystem including open frameworks, models, and tools.



AMD PENSANDO™ NETWORKING

ADVANCING PERFORMANCE IN AI NETWORKS

AMD Pensando™ Salina 400, a fully P4 programmable DPU, provides 2x performance and scale compared to previous generations.[7] And the AMD Pensando™ Pollara 400 NIC uses UEC-ready RDMA, which has a 6x faster message completion time and a 5x faster collective completion time than RoCEv2.[8]

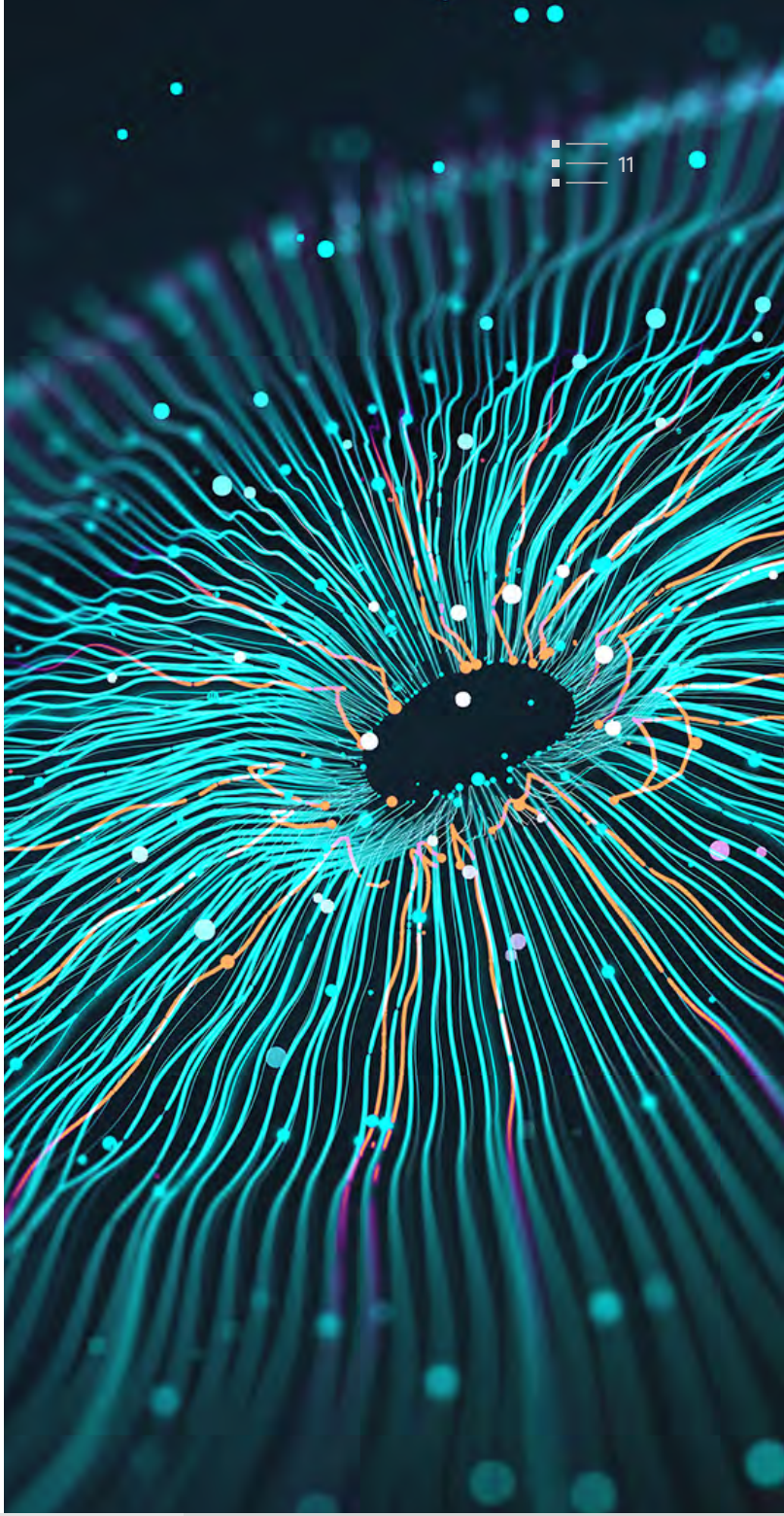


AMD

AMD INFINITY GUARD

SECURITY CAPABILITIES FOR A DATA-DRIVEN WORLD

AMD Infinity Guard[9] is a sophisticated suite of system-level security features designed for a data-driven world. Built into AMD EPYC™ server CPUs, its state-of-the-art capabilities are designed to help defend against internal and external threats and keep data safe.



BEYOND THE DATA CENTER

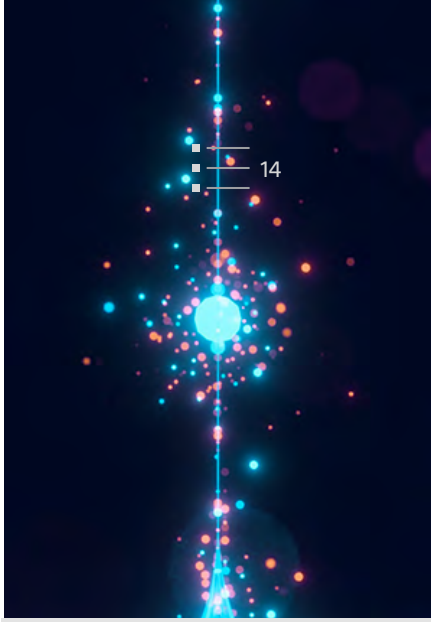
AI is not reserved for large-scale cloud and data center server deployments. For many end users, AI-powered PCs will provide many day-to-day business benefits. They can deliver real-time performance, even when network connections aren't optimal.

This approach also keeps personal or sensitive data on the device, enhancing privacy protections. Business users benefit from AI-assisted apps that streamline productivity tasks, automate data processing, and provide personalized insights based on usage patterns, all at lightning-fast speeds.

AMD RYZEN™ AI PRO PROCESSORS

THE WORLD'S MOST ADVANCED PC PROCESSORS

Built for faster, smarter, and more efficient AI-powered workflows, AMD Ryzen AI™ PRO 300 Series processors offer the world's most powerful NPU for next-gen AI enterprise PCs.[10]



AMD

AMD RYZEN™ AI PRO PROCESSORS

THE WORLD'S MOST ADVANCED PC PROCESSOR

Laptops

AMD Ryzen™ AI PRO 300 Series processors ship in a wide range of Microsoft Copilot+ PC laptops and offer up to 1.4x the multithreaded performance compared to competitive offerings[11] and up to 23 hours of multiday battery life[12].

Mobile Workstations

Leveraging hardware-accelerated generative AI or Copilot+ PCs with Windows® 11 and AMD Ryzen™ AI Max PRO Series processors to get real-time AI assistance, and experience up to 50 TOPS of efficient AI performance.[13]

Experience up to

102%

faster performance than Intel on Chaos V-Ray.[14]***

***AMD Ryzen™ AI Max+ PRO 395 processor versus Intel Core Ultra 9 185H processor.

Desktops

Empower employees to fuel new levels of productivity with the first dedicated AI engine for desktop PCs.[15] AMD Ryzen™ 7 8700G processors deliver up to 46% higher performance compared to an Intel® Core™ i7 14700 processor.****

****Passmark 11 overall benchmark.

> **FOR US, IT'S THE ONE SOLUTION THAT JUST WORKED, AND IT MET OR EXCEEDED ALL OF MY EXPECTATIONS. IT'S ALMOST RARE WHEN A TECHNOLOGY PROVIDER LIVES UP TO ALL THE PROMISES IT MAKES.**

BJ Meyer
CIO, Higginbotham

AMD PRO

AMD PRO delivers enterprise-grade performance, manageability, and security for modern business computing. The suite is designed to help IT and business leaders simplify fleet management, protect sensitive data, and extend device lifecycle value, enabling productivity, reliability, and sustainability across diverse work environments.

AMD PRO Security

AMD PRO Security empowers your business with a multi-layered security infrastructure, providing reliable protection against modern cyber threats.

AMD PRO Manageability

AMD PRO Manageability provides the highest level of compliance with open standards for remote systems management[16] and empowers IT managers with the best control, flexibility, and simplicity[17]. It is offered on the widest range of 3rd-party network devices in the industry.[18] AMD Ryzen™ AI PRO processors provide a unified software image and consistent BIOS for streamlined deployment.

FROM EVOLUTION TO ACCELERATION: POWERING THE AI FUTURE

AI has become ubiquitous across the enterprise. It's tailored to specific tasks, varying by use case and department, and increasingly specialized by domain and industry.

AMD provides choice, flexibility, and proven performance with a reliable roadmap that enterprises can count on. AMD solutions address today's enterprise needs and combine advanced technologies that simplify deployment across the AI infrastructure, accelerating scalability and enabling faster time-to-value.

As enterprises navigate the next wave of AI-driven transformation, partnering with AMD means working with a technology leader who continually invests in research and development, with a long history of delivering product roadmaps reliably. AMD stands as a trusted partner, delivering the technologies, ecosystem, and innovation freedom needed to build the foundation for tomorrow's intelligent, efficient, and resilient enterprise.

PLAN FOR AI WITH CONFIDENCE

Get in touch to discuss how working with AMD can accelerate business outcomes with AI for an organization, and deliver scalable, sustained success for the data center and fleet.

Explore more

CLAIMS:

[1]5-Point Likert Scale; Watchtower AI-powered research platform powered by Intercept, Global, n=161,242.
[2]5-Point Likert Scale; Watchtower AI-powered research platform powered by Intercept, Global, n=453,097.
[3]EPYC-029D: Comparison based on thread density, performance, features, process technology and built-in security features of currently shipping servers as of 10/10/2024. EPYC 9005 series CPUs offer the highest thread density, leads the industry with 500+ performance world records including world record enterprise leadership Java® ops/sec performance, top HPC leadership with floating-point throughput performance, AI end-to-end performance with TPCx-AI performance and highest energy efficiency scores. Compared to 5th Gen Xeon, the 5th Gen EPYC series also has more DDR5 memory channels with more memory bandwidth and supports more PCIe® Gen5 lanes for I/O throughput, and has up to 5x the L3 cache/core for faster data access. The EPYC 9005 series uses advanced 3-4nm technology, and offers Secure Memory Encryption + Secure Encrypted Virtualization (SEV) + SEV Encrypted State + SEV-Secure Nested Paging security features. For additional details, see https://www.amd.com/en/legal/claims/epyc.html#q=epyc5#.
[4]Analysis based on AMD internal data.
[5]9xx5-162: XGBoost (Runs/Hour) throughput results based on AMD internal testing as of 04/08/2025. XGBoost Configurations: v1.7.2, Higgs Data Set, 32 Core Instances, FP32.
2P AMD EPYC 9965 (384 Total Cores), 1.5TB 24x64GB DDR5-6400 (at 6000 MT/s), 1.0 Gbps NIC, 3.84 TB Samsung MZWLO3T8HCLS-00A07, Ubuntu® 22.04.5 LTS, Linux 5.15 kernel, BIOS RVOT1004A, (SMT=off, mitigations=on, Determinism=Power), NPS=1
2P Xeon 6980P (256 Total Cores), 1.5TB 24x64GB DDR5-8800 MRDIMM, 1.0 Gbps Ethernet Controller X710 for 10GBASE-T, Micron_7450_MTFDKBG1T9TFR 2TB, Ubuntu 22.04.1 LTS Linux 6.8.0-52-generic, BIOS 1.0 (SMT=off, mitigations=on, Performance Bias)
Results:
CPU Throughput Relative
2P 6980P 400 1
2P 9755 436 1.090
2P 9965 771 1.928
Results may vary due to factors including system configurations, software versions and BIOS settings.
[6]MI350-009A: Based on calculations by AMD Performance Labs in May 2025, to determine the peak theoretical precision performance for the AMD Instinct™ MI350X / MI355X GPUs, when comparing FP64, FP32, FP16, OCP-FP8, FP8, MXFP6, FP6, MXFP4, FP4, INT8, and bfloat16 datatypes with Vector, Matrix, or Tensor with Sparsity as applicable, vs. NVIDIA Blackwell B200 accelerator. Server manufacturers may vary configurations, yielding different results.
[7]PEN-012: Measurements conducted by AMD Performance Labs as of Aug 27, 2024 on the current specification for the AMD Pensando™ Salina DPU accelerator designed with AMD Pensando™ 5nm process technology, projected to result in delivering 400Gb/s line-rate estimated performance. Estimated delivered results calculated for AMD Pensando™ Elba DPU designed with AMD Pensando 7nm process technology resulted in 200Gb/s line-rate performance. Actual results based on production silicon may vary. Salina projected performance: Bandwidth: 400Gbps Connections per second: 10M Packets per Second: 100MPPS Encryption Offloads: 400 Gbps Storage IOPS: 4 Million Actual results and specifications may vary based on production silicon.
[8]Yanfang Le, Rong Pan, Peter Newman, Jeremias Blendin, Abdul Kabbani, Vipin Jain, Raghava Sivaramu, Francis Matus. Case Study: STrack: A Reliable Multipath Transport for AI/ML Clusters (https://arxiv.org/pdf/2407.15266). Results not specific to the Pollara AI NIC 400, actual performance may vary.
[9]GD-183A: AMD Infinity Guard features vary by EPYC™ Processor generations and/or series. Infinity Guard security features must be enabled by server OEMs and/or Cloud Service Providers to operate. Check with your OEM or provider to confirm support of these features. Learn more about Infinity Guard at http://www.amd.com/en/products/processors/server/epyc/infinity-guard.html.
[10]STXP-06a: Based on AMD product specifications and competitive products announced as of March 2025. 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 an enterprise today. AI PC is defined as a laptop PC with a processor that includes a neural processing unit (NPU).
[11]STXP-12: Testing as of Sept 2024 by AMD performance labs on an HP EliteBook X G1a (14in) (40W) with AMD Ryzen AI 9 HX PRO 375 processor, Radeon™ 890M graphics, 32GB of RAM, 512GB SSD, VBS=ON, Windows 11 Pro vs. a Dell Latitude 7450 with an Intel Core Ultra 7 165H processor (vPro enabled), Intel Arc Graphics, VBS=ON, 16GB RAM, 512GB NVMe SSD, Microsoft Windows 11 Pro in the application(s) (Best Performance Mode): Cinebench R24 nT. Laptop manufactures may vary configurations yielding different results.
[12]STXP-32: Based on internal testing by AMD as of 9/23/24. Battery life results evaluated by operation of a nine-participant Microsoft Teams video conference on battery. Test configuration for AMD and Intel systems run from power level 90% > 45% @150nits brightness and power mode set to ""best power efficiency."" System config: HP EliteBook X G1a (14in) with an AMD Ryzen AI 9 HX PRO 375 processor (40W), Radeon™ 890M graphics, 32GB RAM, 512GB SSD, VBS=ON, Windows 11 Pro. System config: Apple MacBook Pro 14 with M3 Pro 12-core processor, Apple integrated graphics, 36GB RAM, 1TB SSD, MacOS 15.0. System Config: Dell Latitude 7450 with an Intel Core Ultra 7 165H processor (28W) (vPro enabled), Intel Arc Graphics, VBS=ON, 16GB RAM, 512GB NVMe SSD, Windows 11 Pro. Manufacturers may vary configurations yielding different results. Performance may also vary based on use of latest drivers.
[13]GD-243: Trillions of Operations per Second (TOPS) for an AMD Ryzen processor is the maximum number of operations per second that can be executed in an optimal scenario and may not be typical. TOPS may vary based on several factors, including the specific system configuration, AI model, and software version. https://www.amd.com/en/products/processors/laptop/ryzen-pro/ai-max-pro-300-series/amd-ryzen-ai-max-plus-pro-395.html.
[14]SHOP-01: Testing as of November 2024 by AMD Performance Labs using the following benchmarks: V-Ray Benchmark 6, Cinebench R24 (CPU, nT), SPECapc Maya (CPU Composite), SPECapc Solidworks 2024, SPECapc PTC Creo. Configuration for AMD Ryzen™ AI Max+ PRO 395 processor: AMD reference board, Radeon™ 8060S graphics, 128GB RAM, 1TB SSD, VBS=ON, Windows 11. Configuration for Intel Core Ultra 9 185H processor: Dell Precision 5490 14", Nvidia RTX 3000 Ada Graphics (8GB), 64GB RAM, 1TB SSD, VBS=ON, Windows 11. Laptop manufacturers may vary configurations yielding different results.
[15]PXD-03: As of January 2024, AMD has the first available dedicated AI engine on a desktop PC processor, where 'dedicated AI engine' is defined as an AI engine that has no function other than to process AI inference models and is part of the x86 processor die. For detailed information, please check: https://www.amd.com/en/products/processors/consumer/ryzen-ai.html. PXD-03.
[16]KRKP-7: Compared to Intel vPro, AMD PRO Manageability implements more profiles of the DASH Management Initiative to support multi-vendor management for desktop and mobile systems.
[17]KRKP-19: Statement compares Intel vPro and AMD PRO Manageability software suites as of February 2025, and is based on the AMD Pro Manageability software suite providing: (a) standard support for wireless and KVM manageability in every configuration; (b) a range of supported solutions for IT flexibility (including DASH CLI command line, lightweight UI AMC, Microsoft Configuration Manager Plugin AMPS, and the web browser-based ADMP) as opposed to one (1); and (c) a consistent user experience that adhere to well-known design principles.
[18]KRKP-20: As of February 2025, AMD PRO Manageability is supported on a wide range of networking hardware from 3rd parties consisting of wireless devices from MediaTek, Qualcomm and Realtek as well as wired devices

AMD
together we advance_

CLAIMS: (continued)

from Broadcom, Marvell and Realtek, unlike Intel vPro which only supports their own network components.
[19]GD-18u: DISCLAIMER: The information contained herein is for informational purposes only and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD Is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No licenses, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD products are set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale.

AMD
together we advance_