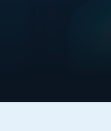
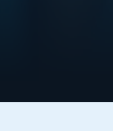


Scaling Agentic Intelligence: The Case for Unified AI Infrastructure



Madhumitha Sathish
Research Manager, High Performance Computing, IDC



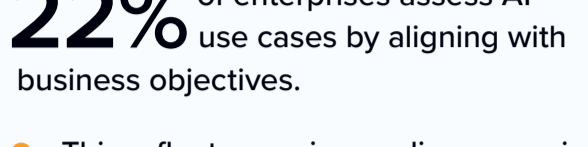
Peter Rutten
Research Vice-President, Performance Intensive Computing, Worldwide Infrastructure Research, IDC

Navigating Agentic AI Uncertainty and AI Readiness: Awareness to Impact

Most enterprises are still exploring agentic AI, launching it cautiously and evaluating its impact with strategic intent.

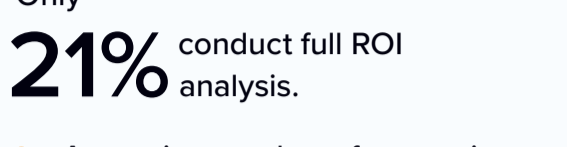
75% of enterprises either haven't explored or lack clarity on agentic AI use cases.

- This reflects a major readiness gap in understanding and defining agentic AI applications.



83% of enterprises launch fewer than 10 AI use cases simultaneously.

- This indicates cautious scaling, suggesting limited resources, risk aversion, or a need for clearer ROI.



22% of enterprises assess AI use cases by aligning with business objectives.

- This reflects a major readiness gap in understanding and defining agentic AI applications.



Only **21%** conduct full ROI analysis.

- A growing number of enterprises are applying structured, strategic methods to evaluate AI's business impact.



n = 410
Source: IDC's AMD AI Survey, July 2025

Securing AI: Policies and Protections

Organizations are prioritizing layered access and robust security for AI model development and deployment.

What types of access policies and security measures does your organization have in place for protecting AI model development and deployment?

77% Use data encryption at rest and in transit.

- Encryption is now a baseline expectation for protecting sensitive AI data.



74% Enforce minimum necessary access for users.



73% Follow compliant data storage and processing policies.

- Regulatory compliance is a major driver in AI infrastructure decisions.



80% Implement multi-factor authentication for added security.

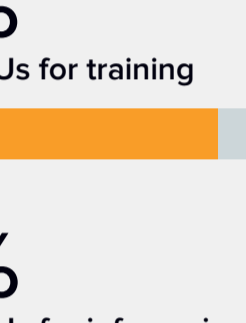
- Identity verification is a top priority in AI environments.



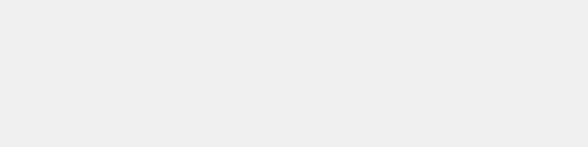
n = 410
Source: IDC's AMD AI Survey, July 2025

Optimizing AI Training and AI Inferencing Infrastructure

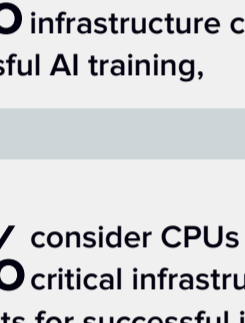
Power, speed, and precision: Specialized compute, memory, and interconnect — the right infrastructure stack drives smarter, faster AI model training and inferencing.



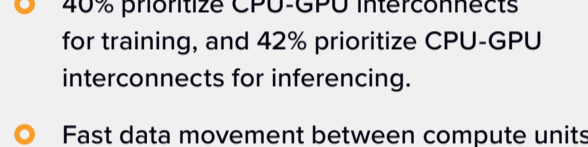
70% rely on GPUs for training



72% rely on GPUs for inferencing.



Almost **25%** consider CPUs to be critical infrastructure components for successful AI training.



32% consider CPUs to be critical infrastructure components for successful inferencing.



- 40% prioritize CPU-GPU interconnects for training, and 42% prioritize CPU-GPU interconnects for inferencing.
- Fast data movement between compute units is essential to avoid bottlenecks.

n = 410
Source: IDC's AMD AI Survey, July 2025

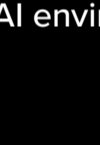
Building Resilient AI Infrastructure

1. In which locations does your organization develop (train or fine-tune) and deploy its AI models?

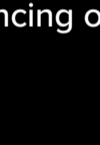
Organizations are adopting hybrid strategies to balance scalability, control, and compliance in AI training and inferencing workloads.



72% develop AI models in public cloud



50% still use their own datacenters.



63% deploy AI models in public cloud.



50% use their own datacenters.

- While cloud dominates, on premises remains vital for control, customization, and compliance.

Regardless of the approach, reliability, availability, and serviceability (RAS) features are becoming standard requirements. When asked if the organization has installed RAS features in the AI environment for inferencing on the AI model:



65% of organizations have already installed RAS features in their AI inferencing environments.



38% with robustness equal to or greater than their IT landscape.



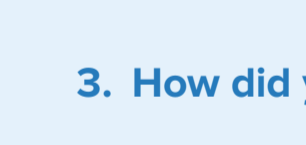
27% with less robustness than their IT landscape.

n = 410
Source: IDC's AMD AI Survey, July 2025

Enterprises balance performance, flexibility, and scale across training, inferencing, and deployment environments.

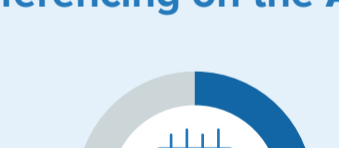
2. How did your organization obtain the AI infrastructure for training the AI model?²

25% rely on VARs or integrators for streamlining AI infrastructure setup.



37% build all or portions of the AI infrastructure in-house.

- Custom setups allow organizations to tailor systems to specific AI workloads and business needs.



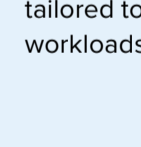
37% opt for a fully configured turn-key solution from a vendor.

- Pre-integrated systems help accelerate time to value and reduce deployment complexity.



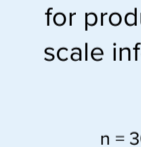
n = 299
Source: 2 IDC's AMD AI Survey, July 2025

3. How did your organization obtain the AI infrastructure for inferencing on the AI model?³



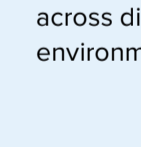
38% build infrastructure in-house

- Many organizations prioritize control and customization for inferencing systems tailored to specific workloads.



36% opt for turn-key vendor solutions.

- Pre-configured systems offer fast deployment and reduced complexity for production-scale inferencing.



24% rely on VARs or integrators

- Expert partners help streamline setup and ensure optimized performance across diverse environments.

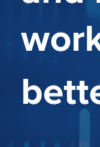
n = 304
Source: 3 IDC's AMD AI Survey, July 2025

Integrated AI Infrastructure: Unlocking Speed, Scale, and Smarter Innovation

Would an end-to-end AI enterprise infrastructure solution that integrates CPUs, GPUs, AI software, networking (DPUs, UALink, and Ultra Ethernet Networking), adaptive computing, and cluster-level systems design benefit your organization?

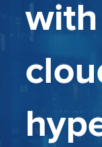
As AI adoption accelerates, enterprises are looking to unified systems that simplify deployment, optimize resources, and reduce reliance on costly cloud services.

What Respondents Are Saying



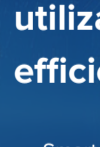
“A unified cluster-level design will help us scale efficiently and manage workloads better.”

— Scalability and workload management are top priorities for enterprise AI.



“Deploying this infrastructure in our datacenters would decrease our spend with specialty clouds and hyperscalers.”

— Cost control and faster deployment are key drivers for on-prem AI infrastructure.



“Cluster-level optimization and resource pooling would boost resource utilization efficiency.”

— Smarter resource use leads to better performance and ROI.



“Such a solution would enhance our ability to develop and deploy agentic capabilities more efficiently”

— Enterprises want infrastructure that supports rapid innovation and dynamic business needs.

Conclusion — Scaling AI with Strategic Infrastructure

Organizations must optimize cost, security, and flexibility to scale AI across diverse environments.

Plan for long-term resilience and ROI by integrating forecasting, sustainability, and cost-aware infrastructure decisions.

Strengthen governance and security practices with encryption, access controls, and compliance policies widely adopted.

Adopt hybrid infrastructure models for flexibility, blending cloud, datacenter, and turnkey systems to meet evolving AI needs.

Strategically manage rising infrastructure costs with end-to-end AI solutions. As AI adoption grows, enterprises face mounting costs tied to model updates, inferencing, and specialized hardware such as GPUs. End-to-end infrastructure solutions that include cluster-level optimization and resource pooling offer a path to streamline operations, improve resource efficiency, and reduce reliance on costly external cloud services.

n = 410
Source: IDC's AMD AI Survey, July 2025

Message from the Sponsor



AMD is the trusted, open, and flexible partner that accelerates enterprise AI transformation. Our broad portfolio of compute engines provides choice, flexibility and proven performance enabling enterprises to optimize to compute for every workload including training and inference. Our commitment to an open ecosystem empowers industry collaboration, providing enterprises with faster access to innovations. And our full-stack solutions include advanced orchestration capabilities that help enterprises easily deploy and scale AI initiatives. AMD is the customer-first AI leader that meets the evolving needs of enterprises. Together, we transform AI potential into measurable business outcomes and sustainable competitive advantage.

[Learn more about AI solutions from AMD](#)