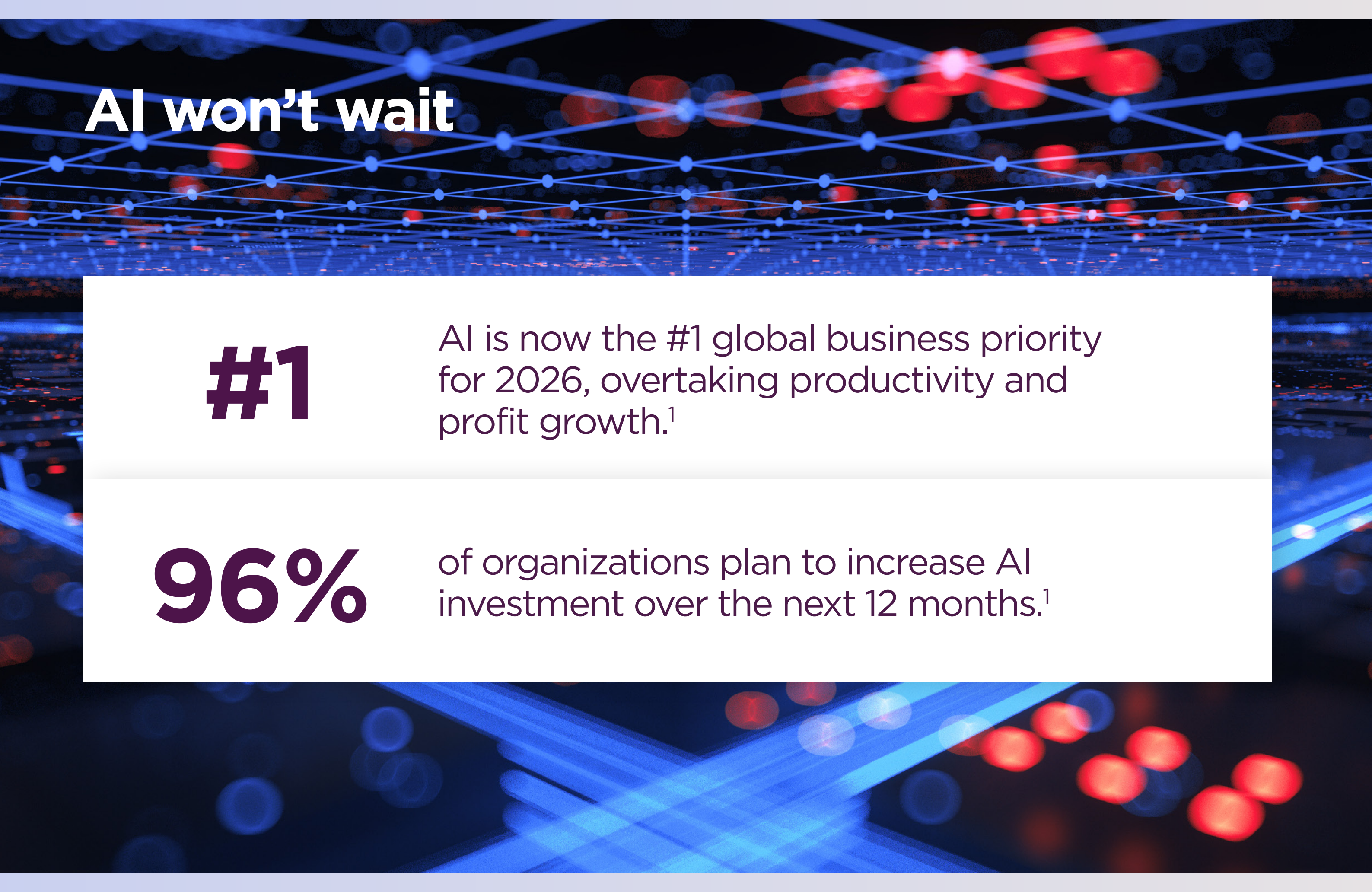
From CPU to GPU: building the right datacenter for your AI workload needs

Are you ready to seize the opportunity of AI?

If you want to run demanding AI inferencing workloads you need a reliable, high-performance platform that can meet a range of evolving needs while balancing power and efficiency.

And you need it fast.



AI won't wait

#1 AI is now the #1 global business priority for 2026, overtaking productivity and profit growth.[1]

96% of organizations plan to increase AI investment over the next 12 months.[1]

Datacenters share capabilities

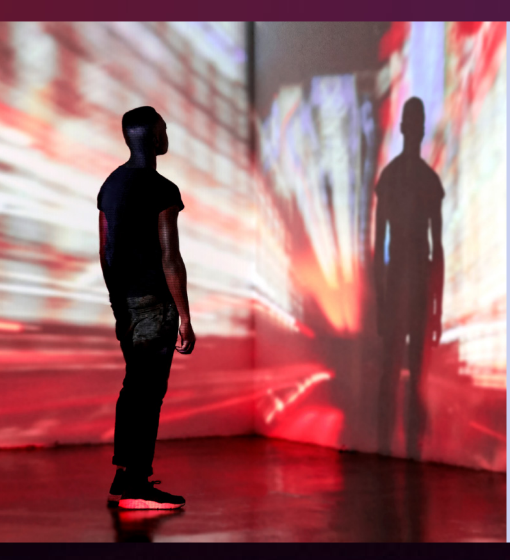
Whether in-house or contracted public/private, organizations rely on datacenters to power AI across their organization.

- Public models
- Public services (e.g. Chat-GPT, Gemini)
- Privately built/trained models
- Licensed models (enterprise, industry)

How will you power ever-growing AI?

CPUs - The latest CPUs are being designed for AI. They can be more efficient than GPUs for small- to mid-sized AI model development, testing, and training.

GPUs - GPUs can be faster and reduce overall total cost of ownership (TCO) when developing and training larger models.



Get the best of both worlds

Why compromise? Embrace a flexible platform that offers all the benefits of CPUs with the ability to efficiently scale up to GPU power as needed.

AMD EPYC™ processors and AMD Instinct™ GPUs address the spectrum of AI workloads

Deploy with the confidence of open standards and leadership security features.

AMD Instinct GPUs and AMD EPYC processors CPUs are designed to excel across different AI workloads:

AMD EPYC processors (CPUs) are ideal for...

- Mixed workload inference deployments
- Classic machine learning
- Small to medium models
- Batch, offline & small-scale real-time inference

AMD Instinct GPUs are ideal for...

- AI training
- Dedicated AI deployments
- Medium to large Gen AI models
- Large-scale, real-time inference

ThinkSystem SR685a V3

- Powerful 8U server
- Features two AMD EPYC processors
- Includes up to 8 AMD Instinct GPUs

ThinkSystem SR675 and SR675i V3

- Versatile GPU-rich 3U rack server
- Features one or two AMD EPYC processors
- Includes Lenovo Neptune® hybrid liquid-to-air cooling[2]

ThinkEdge SE455 V3

- Powerful edge server
- Features one AMD EPYC processor
- Includes up to 8 GPUs

The right CPU host is essential to maximize the value of large GPU Investments

Perfect partners

CPU Inferencing helps CIOs achieve:

- AI value in production
- Real-time decisions from data
- Lower latency, faster operations
- Control data. Strengthen governance.

AMD EPYC CPUs complement the GPUs with:

- High frequency and performance for massive volumes of data pre-processing
- High memory bandwidth, high performance I/O

CPU-based AI use cases:

- Recommendation systems
- ML image recognition and analysis
- Generative AI for enterprise-class LLM/SLM systems (e.g. HR Chatbot, document summarization)

Lenovo and AMD offer performance and efficiency leadership for both CPUs & GPUs.

Together, we will work with you to help identify the best path to solve your AI datacenter challenges.

Learn More

---

[1] IDC & Lenovo. (2026). CIO Playbook 2026: The Race for Enterprise AI. IDC. https://www.lenovo.com/us/en/solutions/cioplaybook2026/
[2] Lenovo. (2025). Lenovo ThinkSystem SR675 V3 Server Product Guide. Lenovo Press. https://lenovopress.lenovo.com/lp1611-thinksystem-sr675-v3-server

© 2026 Lenovo. All rights reserved. Lenovo and the Lenovo logo are trademarks of Lenovo. AMD, the AMD Arrow logo, EPYC, Instinct, and combinations thereof are trademarks of Advanced Micro Devices, Inc. All other trademarks are the property of their respective owners.