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

Lenovo

AMD

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.

96%

Al won't wait

79%

of leaders say their organizations need to

adopt AI to stay competitive.²

of CIOs anticipate increased investment in AI.¹



datacenters to power AI across their organization. Public models Public services (e.g. Chat-GPT, Gemini)

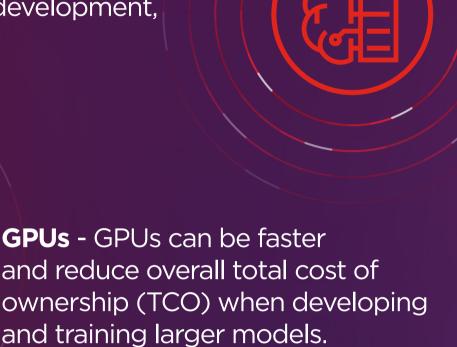
 Privately built/trained models Licensed models (enterprise, industry)

Whether in-house or contracted public/private, organizations rely on

- 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

testing, and training.

mid-sized AI model development,





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™ accelerators address the spectrum of

AMD Instinct Accelerators (GPU) and AMD EPYC processors (CPU) are designed to excel across different AI workloads:

ThinkSystem SR685a V3 Powerful 8U server

real-time inference

AMD EPYC processors

Mixed workload inference

Classic machine learning

Small to medium models

• Batch, offline & small-scale

(CPU) are ideal for...

deployments

ThinkSystem SR675 V3

- Includes Lenovo Neptune® hybrid liquid-to-air cooling³
- **CPU-based AI use cases**

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

Recommendation systems

- Mixed, Al-enabled Apps (e.g. Database, collaboration, design enhanced with AI)

ML image recognition and analysis

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



AMD Instinct accelerators

Dedicated AI deployments

Medium to large Gen Al

Large-scale, real-time

(GPUs) are ideal for...

Al training

models

inference

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

Features one or two AMD EPYC™ processors

Features two AMD EPYC[™] processors

Versatile GPU-rich 3U rack server

Includes up to 8 AMD Instinct™ accelerators

- Perfect partners
- Generative AI for enterprise-class LLM/SLM systems (e.g. HR Chatbot, document summarization)





