AMD INSTINCT™ MI350 SERIES GPUs: ADVANCE AI AND HPC IN THE DATA CENTER

AMD Instinct™ MI350 Series GPUs set a new leadership standard for generative AI and HPC. Built on the cutting-edge 4th Gen AMD CDNA™ microarchitecture, these GPUs deliver exceptional efficiency and performance for training massive AI models, high-speed AI inference and hosting the most advanced HPC workloads.

THE CONVERSATION ROADMAP

Engage your buyers in a clear, consistent and memorable way. Assist their decision-making by sharing new perspectives, establishing yourself as a trusted partner and clarifying the case for taking action. **Why AMD Instinct MI350 Series GPUs for AI and HPC?**

ICEBREAKER

AI is in massive demand, largely because the productivity benefits of generative AI are immediately understandable and applicable across nearly every field and industry. Meanwhile, HPC applications are relied on to tackle some of today's most critical issues.

THE BREAKDOWN

Generative AI requires training on massive data sets and often inference at scale, a demand that's ramping up as enterprises and cloud providers engage in AI development and integration everywhere they can. Meanwhile, exceptional efficiency and performance are required for the new class of HPC applications needed for progress in healthcare, energy, climate science, transportation, scientific research and more.

THE TURN

What if your customers could get more GPU performance in a drop-in platform with leadership compute unit counts and advanced high-bandwidth, high-capacity memory? What if they could develop AI and HPC applications on multiple acceleration platforms without having to stress over compatibility, operational complexity, specialized design requirements or increased security issues?

THE BREAKTHROUGH

AMD Instinct MI350 Series GPUs offer your customers the opportunity to build high-performance infrastructure with leadership acceleration and increased capacity for the users and data needed to meet today's AI and HPC challenges.

KEY MESSAGES

Achieve leadership performance for AI and HPC.

AMD is introducing two powerful GPUs: the MI350X, designed for standard-density deployments, and the MI355X, purpose-built for high-density AI workloads. With enhanced memory capacity and robust datatype support, these new MI350 Series GPUs, designed for efficiency and scalability, handle diverse AI workloads from large-scale inferencing to generative AI model training.

Enjoy seamless scalability and deployment.

The AMD Instinct MI350X GPU integrates seamlessly with AMD Instinct MI300 Series platforms (MI300X and MI325X) and competitive infrastructures for cost-effective performance upgrades. For higher-density computing, AMD Instinct MI355X GPUs deliver thermal efficiency for both compact and high-capacity cooling configurations.

Leverage proven, open, capable AI software.

AMD facilitates the adoption and use of multiple acceleration platforms and cross-platform HPC and AI development with the ROCm™ 7 open-source software ecosystem and programming toolset. With a collection of drivers, development tools and APIs, it extracts the best performance from AMD Instinct GPUs while maintaining compatibility with industry software frameworks. With Day 0 support, ROCm libraries help optimize AI workloads from Hugging Face®, Meta, OpenAI and others, delivering high throughput, low latency and peak efficiency for NLP, computer vision and generative AI inference—accelerating AI innovation at scale.

Join a community of AI leaders.

AMD Instinct Series GPUs power AI, driving adoption across leading CSP and OEM partners. With support for third-party ISVs and OSVs—including partners such as Clarifai, ClearML, Fireworks, Flex.ai, Lamini, Rapt.ai, RHEL, OpenShift® and UbiOps—AMD provides customers with flexibility, choice and leading-edge tools to accelerate AI innovation. AMD Instinct MI350 Series GPUs easily integrate with JAX, Kokkos, ONNX, PyTorch, Raja, Runtime, TensorFlow, Triton, vLLM and more, enabling effortless AI and HPC deployment with minimal code changes.

UNCOVERING OPPORTUNITIES

You can help your prospect be the catalyst that improves outcomes and drives breakthroughs on AI and HPC systems with outstanding performance, efficiency, TCO benefits and adaptability. Some questions to consider asking your prospect include:

- Can your data center currently meet the heavy demands of AI and HPC that the enterprise now requires?
- Do you need to deploy more AI-powered solutions without adding extra compatibility or operational complexity or requiring specialized designs?
- Do you find it challenging to support AI or HPC initiatives while keeping your budgets under control?
- Do you seek more options for AI than always having to utilize cloud services?
- Are you facing pressure to cut back on your data center's physical space and/or power consumption?
- Would you like to start testing models more quickly with simple building blocks and open access to code libraries?

KEY TAKEAWAYS

Leave every prospect with these four facts:

1

The AMD Instinct™ MI350 Series delivers leading AI performance, with up to 4x higher training and inference throughput compared to MI300X and 1.3x better inference throughput versus Nvidia B200.[MI350-004, MI350-038]

2

Each GPU offers 288 GB of HBM3E memory and 8 TB/s of bandwidth, enabling support for massive models, including up to 520 billion parameters on a single device.[MI350-012]

3

AMD Instinct MI350 Series platforms support both air-cooled and liquid-cooled configurations, to enable high-density deployments. The MI355X delivers up to 40% better tokens-per-dollar versus Nvidia.[MI350-049]

4

Built on the open-source ROCm 7 stack and trusted by industry leaders, MI350 Series GPUs offer developers a scalable, flexible path to enterprise AI, free from proprietary constraints.

AMD INSTINCT MI350 SERIES ARE READY TO DEPLOY.

LEARN MORE AT AT AMD.COM/INSTINCT AND AMD.COM/ROCm.

AMD
together we advance_AI