Jun 4, 2025 11:00 AM Eastern Daylight Time

MangoBoost Sets New Benchmark for Multi-Node LLM Training on AMD GPUs in MLPerf Training v5.0

Share

BELLEVUE, Wash.--(BUSINESS WIRE)--MangoBoost, a provider of cutting-edge system solutions for maximizing compute efficiency and scalability, has validated the scalability and efficiency of large-scale AI training on AMD Instinct™ MI300X GPUs through its MLPerf Training v5.0 submission. Tailored for enterprise data centers prioritizing performance, flexibility, and cost-efficiency, this milestone demonstrates that state-of-the-art LLM training is now viable beyond traditional vendor-locked GPU platforms.

> *This showcases how the AMD Instinct™ MI300X GPUs and ROCm™ software stack synergize with MangoBoost's LLMBoost™ AI Enterprise software and GPUBoost™ RoCEv2 NIC.*
>
> Share

scheduling, and advanced memory management.

- Mango GPUBoost™ RoCEv2 RDMA: Inter-GPU communication hardware optimized for low-latency, high-throughput node-to-node communication, sustaining line-rate performance across thousands of concurrent QPs.

These technologies together deliver predictable and efficient multi-node training, ideal for organizations operating their own AI infrastructure or deploying on public cloud.

Industry-First MLPerf Training on AMD MI300X GPUs

This is the first-ever MLPerf Training submission on AMD GPUs spanning multiple nodes. MangoBoost's platform demonstrated robust performance with a 4-node, 32-GPU cluster and confirmed compatibility with additional model sizes and structures—including Llama2-7B and Llama3.1-8B—in internal benchmarks. These results validate the generalizability of MangoBoost's platform beyond benchmarks to diverse production-scale use cases.

"*I'm excited to see MangoBoost's first MLPerf Training results, pairing their LLMBoost AI Enterprise MLOps software with their RoCEv2-based GPUBoost DPU hardware to unlock the full power of AMD GPUs, demonstrated by their scalable performance from a single-node MI300X to 2- and 4-node MI300X results on Llama2-70B LoRA. Their results underscore that a well-optimized software stack is critical to fully harness the capabilities of modern AI accelerators.*" — **David Kanter**, Founder, Head of MLPerf, MLCommons

Vendor-Neutral AI Infrastructure Enabled by AMD Collaboration

The achievement was made possible through deep collaboration with AMD and

"At MangoBoost, we've shown that software-hardware co-optimization enables scalable, efficient LLM training without vendor lock-in. Our MLPerf result is a key milestone proving our technology is ready for enterprise-scale AI training with superior efficiency and flexibility," said CEO Jangwoo Kim.

MangoBoost continues to develop innovations in communication optimization, hybrid parallelism, topology-aware scheduling, and domain-specific acceleration to further scale performance in distributed AI workloads.

About MangoBoost

MangoBoost is a provider of cutting-edge, full-stack system solutions for maximizing compute efficiency and scalability. At the heart of the solutions is the MangoBoost Data Processing Unit (DPU), which ensures full compatibility with general-purpose GPUs, accelerators, and storage devices, enabling cost-efficient, standardized AI infrastructure. Founded in 2022 on a decade of research, MangoBoost is rapidly expanding its operations in the U.S., Canada, and Korea.

Contacts

Minwoo Son
Strategy & Operations Manager
minwoo.son@mangoboost.io

Industry: Software Audio/Video Hardware Artificial Intelligence Data Management Consumer Electronics Technology Semiconductor