XPerf Inc. Emerges from Stealth with Multi-Million Pre-Seed to Tackle GPU Utilization Challenges

XPerf Inc. unveils AI-native platform to boost GPU efficiency by 20% and cut infrastructure costs for next-gen AI data centers

ROUND ROCK, TX, UNITED STATES, November 17, 2025 /EINPresswire.com/ -- XPerf Inc., an AI infrastructure startup specializing in GPU and ASIC utilization optimization, today emerged from stealth mode and announced that it has raised several million dollars in pre-seed round to accelerate the development and commercialization of the full-stack GPU utilization optimization software platform.

> With our unique AI-native method, XPerf Inc.'s technology can boost the AI data centers' GPU utilization by at least 20% while meeting the dynamic workload demands,"
>
> *Dr. Molly Guo*

As AI infrastructure evolves from training-dominated workloads to a mix of pre-training, post-training, inference, and emerging applications, its complexity continues to grow. The deployment of multi-generation, multi-vendor GPUs, ASICs, and diverse networking fabrics makes GPU utilization optimization increasingly challenging, requiring technologies that dynamically balance hardware capabilities and workloads. According to Uptime Intelligence, GPUs are operational only 80% of the time and even well-optimized models typically only reach 35 - 45% of the GPU's compute performance. Coreweave reported the highest Model FLOPS Utilization (MFU) on Nvidia H100 GPUs is only 52% in large-scale training benchmarks. This inefficiency erodes the return on investment, increases infrastructure OPEX, delays innovation and applications time-to-market, and inflates the carbon footprint of AI operations.

Co-founded by veteran engineers with extensive experience in deploying AI clusters with thousands of accelerators and hundreds of deployment cases, the XPerf team has been working on novel AI-native approach with advanced algorithms to address the multi-dimensional (hardware and network inhomogeneity, energy consumption, and workloads) challenges.

"With our unique AI-native method, XPerf Inc.'s technology can boost the AI data centers' GPU utilization by at least 20% while meeting the dynamic workload demands," said Dr. Molly Guo, CTO and Co-Founder of XPerf Inc.

“As AI data center deployments are designed towards inferencing, optimizing utilization and cost becomes critical. The challenge that XPerf is addressing and the solution it is developing will be extremely valuable for our customers who are either deploying large scale GPU clusters or developing custom ASICs for AI applications,” commented Dr. Rakesh Sambaraju, Executive Vice President of Infraeo, an early investor of XPerf Inc.

XPerf Inc. will be showcasing its technology with partner, Infraeo Inc., a cutting-edge AI cluster copper and optical connectivity interconnect solutions company, at SC25 (booth #5412).

About Xperf Inc.

XPerf Inc. is an AI infrastructure startup specializing in GPU and ASIC utilization optimization for large-scale AI clusters. Co-founded by ex-Intel Gaudi engineers with extensive experience deploying clusters with thousands of accelerators, the company takes an AI-native approach to optimize workload performance, energy efficiency, and operational effectiveness across multi-generation, multi-vendor hardware environments. XPerf’s full-stack optimization platform is built on a hardware monitoring and management foundation, helping organizations address the industry-wide challenge of GPU underutilization. XPerf Inc. is a member of the NVIDIA Inception program, a member of CNCF, and a participant in the gBETA accelerator program.

About Infraeo Inc.

Infraeo Inc. designs and manufactures high-speed copper and optical interconnects for AI clusters, data centers, and high-performance computing. With global production and millions of units shipped, Infraeo delivers next-generation performance at scale.

Molly Guo
XPerf Inc
press@xperf.ai