Category Award

AI-Stack

INFINITIX Inc.
數位無限軟體股份有限公司

[Company Website](#)

Winning Reason

This product focuses on AI infrastructure management, offering rich software and integration capabilities. It provides global customers with effective AI hardware and software integration solutions, significantly enhancing the scheduling and integration of hardware resources. It is suitable for various AI model training applications.

The platform distributes AI tasks across multiple computing nodes and makes use of distributed technology to process large data sets, which boosts the efficiency of AI model computations and improves resource use. In the current era of prioritized cost-effectiveness in hardware performance and operational cost management, this product highlights the critical role of a software management platform in AI infrastructure and underscores the value of software in AI applications.

Product Feature

1.Efficient GPU Management: Partitioning and aggregation elevate usage to 90%, allowing multi-user, multi-task operations. 2.High-Performance Distributed Computing: Runs training across multiple nodes in parallel, significantly reducing runtime. 3.Open-Source & Multi-Cloud Integration:Combines TensorFlow/PyTorch with multi-cloud, streamlining AI development and deployment. 4.Intelligent Workload Scheduling: Leverages Kubernetes-based multi-queue, overcommitment, and auto pause-resume for optimal resource usage. 5.User-Friendly MLOps:Provides a containerized environment for quick GPU and framework selection, enhancing the entire AI lifecycle. 6.Industry Applications:Deployed in healthcare, manufacturing, data centers, and research to optimize resources and enable data-driven decisions.