

QNAP

Edge AI Storage Server × All-Flash × 25GbE

## QAI-h1290FX

Built for faster, simpler on-prem AI workloads



## GPU-Ready for Advanced AI

Supports NVIDIA RTX™ PRO Blackwell GPUs up to 96GB VRAM, with UI-based GPU assignment for instant acceleration—ideal for LLMs, generative AI, and complex 3D workloads.



## On-Prem LLMs with Private RAG

Run LLM inference and private RAG directly on enterprise data fully on-premises to keep control, privacy, and compliance without cloud dependency.



## All-Flash + 25GbE Connectivity

All-flash storage with 25GbE networking (and optional 100GbE expansion) removes I/O bottlenecks so GPU compute stays fully utilized as workloads scale.



## One-Click AI Deployment

Use Container Station with GPU acceleration to one-click deploy Ollama, AnythingLLM, Stable Diffusion, and more AI tools for faster time-to-value.

QNAP SYSTEMS, INC.

Copyright © 2026 QNAP Systems, Inc. All rights reserved.

www.qnap.com