



AI Factory Infrastructure at Any Scale

AMAX Engineered on NVIDIA DGX™ and NVIDIA Vera Rubin Platforms

AI infrastructure has progressed beyond proof-of-concept environments into production-ready AI factories supporting large scale training, inference, reasoning, and agentic workloads across enterprise and hyperscale deployments.

AMAX Engineering delivers turnkey AI factory systems based on NVIDIA DGX platforms using the Vera Rubin architecture. These systems are engineered for predictable performance, controlled scale-out growth, and operational consistency under sustained production workloads.

Through high density system engineering, liquid cooled thermal design, and end-to-end deployment services, AMAX accelerates the transition from AI evaluation to AI production while reducing deployment risk, accelerating time-to-value, and establishing AI infrastructure as a core operational capability.

AMAX ENGINEERING

AMAX delivers fully validated rack units with configuration tuning, performance profiling, deployment readiness testing and predictable resource utilization.

NVIDIA DGX Systems for AI Factories

NVIDIA DGX Vera Rubin NVL72

Rack-scale AI compute platform designed for sustained throughput on large, demanding workloads

Workload Focus

Large model training, reasoning workloads, large-scale distributed AI

Architecture Highlights

NVIDIA Rubin GPUs and NVIDIA Vera CPUs integrated with high-speed NVIDIA NVLink™ interconnects and advanced networking to support dense parallelism and high-volume data exchange across nodes

NVIDIA DGX Rubin NVL8

Purpose-built AI infrastructure unit optimized for flexible deployment within AI factory environments

Workload Focus

Agentic AI, inference, post-training optimization, continuous model refinement

Architecture Highlights

Eight Rubin GPUs with high-bandwidth memory and sixth-generation NVLink, optimized for parallel compute and memory-intensive workloads





AMAX Validated Scale Out Architecture and Network Fabric

DGX Vera Rubin NVL72 and DGX Rubin NVL8 systems support expansion from single-rack deployments to multi-rack AI factory clusters. High-bandwidth interconnects and low-latency networking enable efficient communication and data movement across nodes.

Cluster network fabrics are implemented using NVIDIA ConnectX® SuperNICs™ and NVIDIA Spectrum-X™ Ethernet, supporting RDMA traffic, congestion control, and multi-node data flow. Network configurations are validated to maintain throughput and latency characteristics as system scale increases.

AI Factory Management and Orchestration

NVIDIA Mission Control™ streamlines every aspect of the AI factory—from workload scheduling and orchestration to monitoring and autonomous recovery— while empowering platform teams to operate efficiently and scale confidently with validated software.

It combines real time visibility, precise control over power and cooling, and always-on resilience for accelerated AI token production. Management functions including software control, workload orchestration, firmware configuration, and network policy enforcement support efficient GPU operation while enabling consistent automation and predictable scaling of production workloads.

Liquid Cooling and Thermal Engineering

DGX Vera Rubin NVL72 and DGX Rubin NVL8 platforms operate at high power density and require precise thermal control to sustain continuous workloads. Liquid-cooled system configurations are used to support stable operation without thermal throttling.

Thermal performance is validated through system-level testing, including sustained load operation and fluid flow verification, to confirm temperature stability under production conditions.

Deployment and Lifecycle Services

AMAX Engineering provides the following services throughout the AI factory lifecycle:

- System configuration and engineering optimization
- Factory assembly, integration, and wiring
- Burn-in testing and workload stress validation
- Firmware, BIOS, and network alignment
- Rack integration and pre-deployment certification
- Ongoing maintenance support and capacity planning

By embedding these services into the delivery process, AMAX Engineering helps organizations reduce deployment complexity and ensure long-term operational consistency.

Technical Foundation for Production AI

AI factories built on NVIDIA DGX systems powered by Vera Rubin and delivered by AMAX Engineering provide a scalable, validated infrastructure foundation for continuous AI workflows. By combining rack-scale compute platforms with thermal engineering, high-performance networking, and lifecycle services, AMAX Engineering enables organizations to operate AI infrastructure with predictable performance, efficient scaling, and long-term reliability.

