NVIDIA Elite Partner

AMAX AI Factory for SaaS ISVs

Validated enterprise-grade infrastructure for SaaS AI workloads

Self-Host AI workloads with Validated On-Premises GPU-Accelerated Compute to Offset Rising LLM API and Cloud Costs

Cloud/SaaS ISV's face mounting LLM API costs as AI becomes integral to both internal operations and product offering features. AMAX AI Factory solutions enable ISVs to bring GPU-accelerated compute on-premises, offsetting cloud and LLM API expenses while maintaining control over data and models. Whether deploying for AI-assisted development, content and documentation generation, or integrating AI-powered features into product offerings, AMAX AI Factory solutions and services enable ISVs to self-host GPU-accelerated compute—reducing operational costs, limiting dependency on third-party APIs, and gaining flexibility to customize models and control deployment timelines.

NVIDIA AI Enterprise Software for Accelerated Time-to-Value

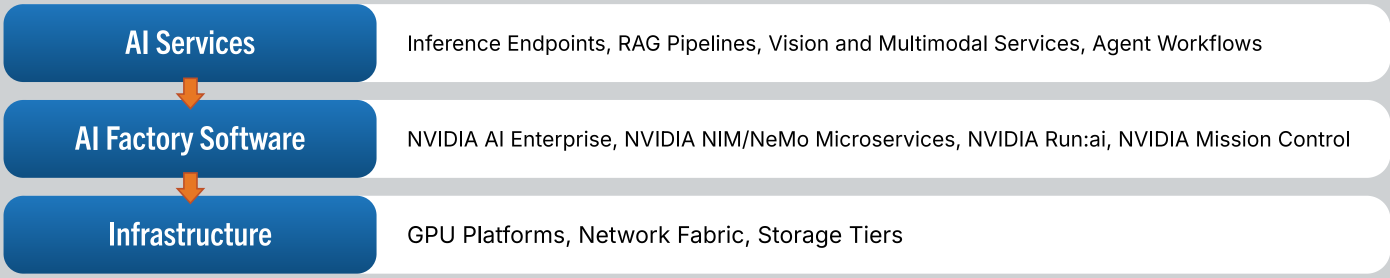
NVIDIA AI Enterprise software accelerates ISV deployment through pre-packaged AI building blocks and integrated infrastructure management. NIM/NeMo containerized microservices provide optimized inference runtimes and fine-tuning capabilities, enabling development teams to build AI features rapidly without having to start from scratch. NVIDIA Mission Control delivers centralized orchestration: Base Command Manager for enterprise IT concerns, such as provisioning, monitoring, and lifecycle management, while Run:ai empowers AI practitioners with workload scheduling, resource optimization, and on-demand AI workspaces. Built on cloud-native architectures and familiar deployment patterns, the integrated software stack enables ISVs to accelerate from hardware deployment to production AI workloads in weeks rather than months, while reducing integration friction, enabling faster ROI on AI infrastructure.

AMAX ENGINEERING

AMAX AI Factory solutions enable ISVs to deploy validated GPU-accelerated compute infrastructure with consistent performance and predictable scaling. Whether for internal development workloads or AI-enabled product features, self-hosting infrastructure delivers significant cost savings while reducing dependency on third-party LLM APIs. Additionally, organizations gain control over data, models, and deployment timelines—transforming AI from a variable operational expense into a strategic infrastructure asset.

Minimize Integration Risk. Maximize Production Readiness.

AMAX simplifies integration to accelerate production readiness. Our professional services and deployment expertise reduce complexity and mitigate risk. Standardized, fully integrated architectures across compute, networking, storage, liquid cooling, and management enable reliable, scalable deployments ready to support AI growth anywhere in the world.



AI Services — Inference Endpoints, RAG Pipelines, Vision and Multimodal Services, Agent Workflows

AI Factory Software — NVIDIA AI Enterprise, NVIDIA NIM/NeMo Microservices, NVIDIA Run:ai, NVIDIA Mission Control

Infrastructure — GPU Platforms, Network Fabric, Storage Tiers

Reference Architecture & Appliance Build

Sizing for throughput & latency targets, standardized deployment patterns, appliance-ready images, and validated reference configurations

Manufacturing, Lifecycle & Global Fulfillment

NPI and lifecycle planning, ISO-backed QA, ATA test coverage, PLM controls, configure-to-order fulfillment, warrant, and streamlined RMA

Business & Operational Outcomes

Higher GPU utilization, predictable inference performance, faster model rollout, clear path to usage-based AI features

At AMAX, engineering drives everything we do.

AMAX brings deep engineering expertise to design and deploy AI infrastructure that performs from day one and scales with you over time. We are the right partner to help you build, scale, and operationalize your AI Factory with confidence.