

NEBIUS

# Nebius for physical AI

Cloud infrastructure built for the robotics era. Train perception models. Run digital twins. Deploy intelligent systems at scale — on AI infrastructure purpose-built for physical AI and robotics.

Try a PoC

## What you can do



### Generate and augment multimodal data

Scalable, high-performance storage for real-world sensor data. Synthetic data generation powered by NVIDIA Cosmos World Foundation Models augments data by 1,000x.



### Train robot and world foundation models

NVIDIA Blackwell clusters for vision-language-action (VLA) model training. Cut training time and focus on models — not infrastructure.



### Simulate before you deploy

Instances with NVIDIA RTX PRO 6000 Blackwell Server Edition for large-scale simulation built with NVIDIA Isaac Sim and NVIDIA Cosmos. De-risk robot training through virtual environments.



### Accelerate the full robotics workflow

NVIDIA OSMO Managed by Nebius connects synthetic data, training, simulation and deployment in one orchestrated pipeline — eliminating DevOps overhead.

## Why Nebius

### Purpose-built AI-infrastructure cloud

Vertically integrated cloud platform optimized for physical AI workloads

### Production-ready inference

Cost-effective platform to scale models from prototype to production



### Jointly engineered by Nebius and NVIDIA

Deep collaboration with NVIDIA on AI factories powering inference and agentic AI



## Physical AI innovators on Nebius

### RoboForce

Building robo-labor to elevate humans beyond dull, dirty and dangerous work.

Cut setup time by more than 70%.

Generates thousands of scenario variations with Nebius and NVIDIA.

Eliminated manual handoffs in physical AI workflows.

### milestone

Global leader in video management software and computer vision.

Cost-effectively transforms real-world footage into trained models.

Leverages weeks of sustained access to AI infrastructure, data pipelines and workflow orchestrators.

# Technical infrastructure

Built for physical AI workloads — from simulation to production.

## Challenge

## Nebius solution

VLA training stalls on network bottlenecks

- 3.2 Tbps InfiniBand
- Linear scaling to 512+ GPUs

Spiraling simulation costs

- Competitive pricing on instances with ray-tracing cores
- 60% savings on preemptible instances

Multimodal data pipelines choking GPUs

- 500 GB/s storage throughput
- Zero data starvation

Training runs interrupted by infrastructure failures

- Auto node health monitoring
- Intelligent restarts

Complex orchestration slowing dev

- Soperator by Nebius: native Slurm on Kubernetes
- Use both Slurm for training and Kubernetes for inference

## AI infrastructure capabilities

### Compute

- NVIDIA HGX H100, H200, B200, B300, and GB200 NVL72 clusters
- NVIDIA RTX PRO 6000 Blackwell Server Edition instances for simulation
- Bare-metal performance and cloud flexibility

### Storage

- Choice of SSDs for performance, reliability and price
- Enhanced Object Storage for multimodal ingestion
- WEKA and VAST integration for petabyte-scale data

### Networking

- 3.2 Tbps NVIDIA Quantum-2 InfiniBand
- Deterministic latency for distributed training and simulation
- Rail-optimized topology to eliminate jitter

### Orchestration

- Soperator (Slurm on Kubernetes) by Nebius
- Gang scheduling with Kubeflow and Volcano
- Anyscale, dstack, NVIDIA Run:ai and SkyPilot integration

## Trusted by hyperscalers and frontier AI labs



Make it your experience

[nebius.com/solutions/phy](https://nebius.com/solutions/phy)

Try a PoC