

NEBIUS

Nebius for life science

Move from experience to validated result
faster with AI infrastructure and scientific
workflows built for research at scale

AICPA
SOC

SOC 2
Type II

AICPA
SOC

SOC 3



HIPAA



ISO/IEC
27001



ISO/IEC
27799



ISO
22301



ISO/IEC
27701



ISO/IEC
27018



ISO/IEC
27032



NIS 2



DORA

AI infrastructure for scientific AI

Scientific AI brings together the industries applying AI and accelerated computing to complex scientific problems, from drug discovery and genomics to molecular simulation and plant sciences. These organizations rely on scalable compute, reproducible workflows, and secure environments to turn research into validated results.

Science teams often manage the entire AI lifecycle, from data preparation and simulation to model training and inference. As projects mature, workflows evolve from single-GPU experiments to production deployments distributed across thousands of nodes. Scientific AI spans a broad range of industry and research applications, including the most common use cases listed below.

The compounding challenges

Scientific AI involves complex experimental workflows with highly unpredictable success rates, requiring researchers to run substantially more computational experiments to achieve life-changing breakthroughs. As a result, AI researchers must iterate across large datasets, complex models, and extensive simulation pipelines. With growing data volumes, access to GPU capacity becomes harder to secure, while provisioning delays and reliability issues make iterations slow and reproducibility difficult.

Most research teams lack dedicated DevOps/MLOps expertise, leading to significant frustration around infrastructure scaling, management, and troubleshooting. This creates a critical need: scientists and engineers require straightforward access to scalable AI infrastructure and integrated AI tooling paired with expert technical support, allowing them to focus on scientific discovery rather than system administration.

Stanford

“Nebius enabled rapid model screening and fine-tuning, which was very important for the CRISPR-GPT project.”



Professor Le Cong
Head of Cong Lab
Stanford University

Use cases

Biomarker identification

Cryo-EM reconstruction

Genomics/multi-omics
and genome editing design

Protein structure
prediction and design

Population-
scale studies

Next-Generation Sequencing
(NGS) and variant calling

Domain specific
language models

Single-cell and spatial
transcriptomics

Molecular dynamics
and QM/MM simulation

Plant phenotyping
and yield modeling

Clinical AI and trial
intelligence

Large-scale model pre-training
and fine-tuning

Generative
molecular design

Medical imaging
and diagnostics

Molecular docking
and virtual screening

Extreme-scale
virtual screening

How Nebius can help

Faster time to market. Start training your models faster and deliver AI results ahead of the competition. With Nebius, you can access large-scale, production-ready GPU clusters in days, not weeks.

Elasticity for unpredictable research pipelines. Launch and scale your experiments and batch inference workloads with pay-as-you-go pricing. Quickly scale up your cluster to handle burst workloads when needed.

Reproducible workflows, quick to deploy. Scale functional pipelines with fully supported industry-leading orchestration tools. Standardize environments, dependencies, and execution so workflows can be repeated easily.

Fully managed cluster environment. Nebius provides managed Kubernetes- and Slurm-based environments with integrated scheduling, monitoring, and observability. Researchers can submit, monitor, and troubleshoot jobs without maintaining the underlying cluster infrastructure.

Top-performing models, easy to find. Visualize all your experiments in one place: compare metrics, track experiment metadata and model versions, and select the best ones for production with managed MLflow

Smooth transition, seamless integration. Rely on our solution architects and engineers to guide you through the migration process. We'll help you connect Nebius AI Cloud to your existing infrastructure, minimizing disruption to current workflows.

Dedicated onboarding for every customer. Access our white-glove MLOps support to configure, troubleshoot, and scale workloads, helping science teams focus on discovery.

Co-engineered with NVIDIA. Run evolving life science AI workflows from a single platform. Accelerate drug discovery with NVIDIA BioNeMo and NVIDIA Parabricks frameworks in one click, and boost any life science workload with full support for custom model training and fine-tuning.

Proven across diverse research workflows



Population-scale multiomics foundation training

Train modality-agnostic foundation models from large cross-cohort datasets. Scale seamlessly from a few nodes to thousand-GPU clusters, and manage costs with autoscaling and preemptible capacity.



Extreme-scale molecular dynamics

Run microsecond-scale simulations for systems exceeding 100 million atoms with a single command, supported by resilient execution and checkpointing.



Ultra-high throughput docking

Screen the entire ZINC database of nearly 10 billion small molecules in just five days by using a few lines of code, featuring automated retry and error recovery for uninterrupted runs.

A purpose-built stack for Scientific AI

Drug discovery and biology platforms



NVIDIA BioNeMo and NVIDIA Parabricks



Biocipher

Scientific workflows



sequera

Outerbounds



Flyte



dstack



SkyPilot

Experiment tracking and MLOps



CLEARML



Weights & Biases

mlflow



Data and storage



snowflake



WEKA

Model and agent deployment

NEBIUS



TOKEN FACTORY



Serverless AI

Trusted by industry experts



"Basecamp Research trained the world's largest genomic foundation model (by token count) on Nebius's GPUs — 100% of the data is fully proprietary, sourced from five continents. The model will be used to tackle unsolved frontier challenges in the Life Science industry to design programmable medicines."



Phil Lorenz
CTO
Basecamp Research

PRIMA MENTE

"Nebius' GPU Cloud and AI-centric infrastructure give Prima Mente the elastic, cost-efficient HPC we need to train and fine-tune our 90M to 7B-parameter epigenetic foundation models at scale. Their life-sciences-ready platform lets us spin up secure clusters in minutes, compressing model-development cycles and accelerating delivery of next-generation brain-health diagnostics to researchers and clinicians worldwide."



Ravi Solanki
CEO and Co-Founder
Prima Mente

Ardigen

"Nebius AI Cloud provided the scalable compute we needed to run high-resolution DNA sequence embedding and regulatory variant scoring at genome-wide scale. It allowed us to focus on extracting meaningful biological insights from non-coding regions without being constrained by compute resources or workflow limitations."



Krzysztof Kotlarz
Bioinformatician
Ardigen



"Nebius gave us powerful GPUs, a smooth interface and competitive pricing to train large-scale models on sensitive biological data. We developed models that reveal spatial immune features in tumors — capabilities we haven't seen elsewhere. It's a great match for AI in biotech".



Dr. Roy Granit
Sr. Director & Head of Computational Discovery
Compugen



"Nebius gives us exactly what we need — powerful GPUs, fast provisioning and simple scaling without DevOps overhead. Paired with LynxKite: 2000MM's zero-code AI orchestration workflows, we help clients such as Hummingbird Bioscience go from one up to one thousand GPUs as needed, keeping utilization high and delivery fast."



Gyorgy Lajtai
CEO
Lynx Analytics



"With Nebius, we reduced our screening pipeline from days to hours. The scale and reliability of their infrastructure with Ray allowed us to push the limits of molecular docking and deliver faster results to our partners in pharma and consumer health."



Duy Trieu
CTO
Nanyang Biologics

**Get started with a free
proof of concept**

Contact us