

Nebius AI Cloud 3.5 introduces serverless AI to give developers frictionless compute for real-world AI

- Latest “Aether” platform update enables teams to build, run and scale AI workloads without managing infrastructure
- Addition of NVIDIA RTX PRO™ 6000 Blackwell Server Edition GPU and Nebius’s Data Transfer Service simplify real-world AI deployment

Amsterdam, March 26, 2026 — Nebius today unveiled Nebius AI Cloud 3.5, adding significant new capabilities to its full-stack cloud platform that reduce operational friction and enable AI builders to prototype, test, and ship products faster.

The introduction of serverless features gives developers the ability to launch workloads almost instantly, eliminating the need for AI teams to spend significant time configuring infrastructure before they can run experiments, train or serve models in production. Infrastructure configuration and runtime management are handled by the Nebius platform, enabling developers to focus on building applications instead of managing environments.

Alongside serverless capabilities, Nebius is expanding its GPU offering with [NVIDIA RTX PRO 6000 Blackwell Server Edition](#) for a range of workloads including AI inference, industrial robotics, physical AI simulations, visual computing, and drug discovery.

Version 3.5 of Nebius AI Cloud “Aether” also introduces Nebius’s Data Transfer Service, which reduces data management overhead for teams working across environments by simplifying data migration and replication between external S3-compatible storage systems and Nebius cloud regions.

Configuration setup for Managed Soperator, Nebius's fully managed Slurm-on-Kubernetes solution, has also been overhauled to offer more options and granularity when creating a Slurm cluster for self-service users. Managed Kubernetes observability has also been updated to give teams additional cluster-level control.

The AI application marketplace has also been redesigned to help users access faster tools, models and applications required in their workflows.

Other updates in Aether 3.5 include improved user administration and role-based permissions, making it easier for organizations to manage access across teams. New public APIs for billing data streamline the export process for finance and operations teams.

All the new features that the Aether 3.5 release delivers are available now on the global Nebius AI Cloud infrastructure, with the serverless service available in public preview. NVIDIA RTX PRO 6000 Blackwell Server Edition is available today.

Nebius AI Cloud Aether 3.5 — at a glance

Serverless AI

- Elastic, pay-as-you-go compute accelerated by NVIDIA
- Simplified access to AI workloads without managing infrastructure
- Designed for prototyping, experimentation, and model inference evaluation

NVIDIA RTX PRO 6000 Blackwell Server Edition

- GPU option designed for a range of workloads including AI inference, industrial robotics, physical AI simulations, visual computing, engineering research, and drug discovery
- Enables cost-efficient AI inference and simulation-heavy workloads

Data Transfer Service

- User-friendly tool for data transfer and replication across Nebius regions and S3-compatible object storage services

Managed Soperator

- An updated cluster configuration wizard for Nebius' fully managed Slurm-on-Kubernetes solution

Platform enhancements

- Updated navigation for the AI/ML application marketplace
- Improved disk encryption, boot image management, and Kubernetes-level observability
- Expanded controls for user administration and role-based permissions
- Public API for exporting billing data in standardized formats

Additional resources

- [Blog post](#) from our Product Management team
- [Blog post](#) on NVIDIA RTX PRO
- [Blog post](#) on our new serverless AI
- [Webinar and live Q&A registration](#)

About Nebius

Nebius, the AI cloud company, is building the full-stack platform for developers and companies to take charge of their AI future — from data and model training to production deployment. Founded on deep in-house technological expertise and operating at scale with a rapidly expanding global footprint, Nebius serves startups and enterprises building AI products, agents and services worldwide.

Nebius is listed on Nasdaq (NASDAQ: NBIS) and headquartered in Amsterdam.

For more information please visit www.nebius.com

Contacts

Media relations: media@nebius.com

Investor relations: askIR@nebius.com

Disclaimer

Forward-looking statements

This press release contains forward-looking statements within the meaning of the Private Securities Litigation Reform Act of 1995, which involve risks and uncertainties. All statements contained in this press release other than statements of historical fact, including, without limitation, statements regarding our future financial and business performance, strategy, expected growth, planned investments and capital expenditures, capacity expansion plans, anticipated future financing transactions and expected financial results, are forward-looking statements. The words “anticipate,” “believe,” “continue,” “estimate,” “expect,” “guide,” “intend,” “likely,” “may,” “will” and similar expressions and their negatives are intended to identify forward-looking statements.

These forward-looking statements are subject to risks, uncertainties and assumptions, some of which are beyond our control. Actual results may differ materially from the results predicted or implied by such statements, and our reported results should not be considered as an indication of future performance. The potential risks and uncertainties that could cause actual results to differ from the results predicted or implied by such statements include, among others: market, macroeconomic and geopolitical conditions; our ability to build, operate and manage our businesses to the desired scale; competitive pressures; technological developments; our ability to secure and retain clients; our ability to secure additional capital to enable the growth of the business; unpredictable sales cycles; and potential pricing pressures; as well as those risks and uncertainties related to our continuing businesses included under the captions “Risk Factors” and “Operating and Financial Review and Prospects” in our Annual Report on Form 20-F for the year ended December 31, 2024, filed with the SEC on April 30, 2025, which is available on our investor relations website at <https://nebius.com/investor-hub> and on the SEC website at www.sec.gov.

All information in this press release is as of the date hereof (unless stated otherwise). Except as required by law, we undertake no obligation to update or revise publicly any forward-looking statements, whether as a result of new information, future events or otherwise, after the date on which the statements are made or to reflect the occurrence of unanticipated events.

In addition, statements that “we believe” and similar statements reflect our beliefs and opinions on the relevant subject. These statements are based upon information available to us as of the date hereof and, while we believe such information forms a reasonable basis for such statements, such information may be limited or incomplete, and our statements should not be read to indicate that we have conducted an exhaustive inquiry into, or review of, all potentially available relevant information. These statements are inherently uncertain, and investors are cautioned not to unduly rely upon these statements.