

NEBIUS

Nebius Group shareholder letter Q1 2025

Introduction



Arkady Volozh,
Founder and CEO

Fellow shareholders,

Nebius Group has had a very strong start to the year. Our core infrastructure business delivered annualized run-rate revenue (ARR)* as of the end of Q1 of \$249 million, up 684% year-over-year. We are continuing to see strong dynamics in Q2, with April ARR of approximately \$310 million, and have maintained this strong momentum into May. This demonstrates the strength of our business model and our ability to scale up rapidly.

We believe our value proposition is unique. The depth of our technology expertise is a significant differentiator against other neoclouds. Nebius builds full-stack, vertically integrated AI infrastructure — going far beyond bare-metal services — with a broad set of solutions to meet our customers' AI needs. This allows us to serve a diverse range of customers, from AI-native tech startups today to growing number of large enterprise customers in the future.

We are rapidly expanding our capacity footprint. In just three quarters, we've gone from one location in Finland to five locations across Europe, the U.S., and now the Middle East. We are actively exploring new sites in the U.S. and around the world, and we expect to provide more news on this soon.

We are on-track to achieve \$750 million to \$1 billion in annualized run-rate revenue by the end of this year, and in the second half of this year, we expect Nebius Group to turn adjusted EBITDA positive. In the medium term, our base case expectations are to achieve billions of dollars in revenue with adjusted EBIT margins in the 20-30% range, assuming a conservative depreciation schedule of four years.

Delivering on our rapid growth plans will continue to require considerable investment. In this respect, we have been in a uniquely favorable position to fund our growth since we formally launched Nebius last July. We started with a significant amount of capital —

Group revenue

**\$55.3
million**

+385% YoY growth

Annualized run-rate
revenue as of the
end of March 31 for
core infrastructure

**\$249
million**

+684% YoY

Cash at the end of Q1

**\$1.44
billion**

Q1 CapEx

**\$544
million**

* ARR is defined as annualized run-rate revenue by the end of the period (revenue for last month of the period multiplied by twelve).

\$2.5 billion from our divestment — and raised an additional \$700 million from top tier investors in December, giving us more than \$3 billion in cash. This has enabled us to quickly build and scale our AI cloud business.

Looking ahead, we plan to fund our growth in part by tapping into our non-core businesses and equity stakes. We are not aware of any other company in our sector that can raise potentially billions of dollars in this way.

- The value of our minority stake in ClickHouse is substantial, and we believe it is continuing to grow. Should ClickHouse have a liquidity event in the near-to-mid term, this would provide us with significant capital to invest into our core business.
- We believe Toloka has the potential for strong future growth given its recent funding round with high caliber investors, including Bezos Expeditions and Mikhail Parakhin, the CTO of Shopify. We retain a significant majority economic stake in this business.
- Additionally, Avride is building autonomous technologies that can be strategically valuable to many companies — we are actively exploring strategic investments and partnerships to help accelerate its growth.

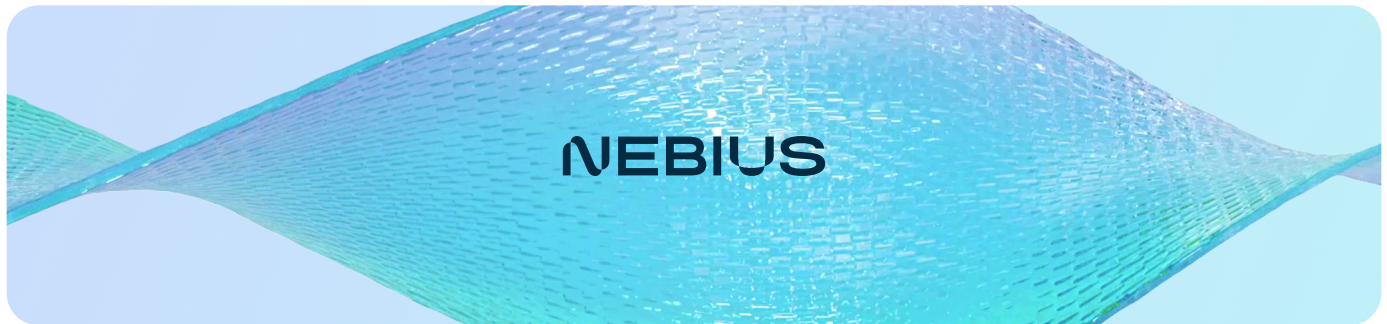
As we have minimal debt, we will be opportunistic in tapping capital markets for more traditional types of financing, while seeking to minimize dilution to our shareholders and to be prudent with respect to any future debt load. Given our mid-term growth prospects, fueled by our lower cost-of-capital funding opportunities, and coupled with our relatively stable cost structure, we see a clear path to building a high-growth business with a strong margin profile that can generate solid, sustainable free cash flows.

The opportunity ahead of us is immense, and we have the technology, talent and funding flexibility to successfully capture it.

Sincerely,

Arkady Volozh

Group business update



Core infrastructure

We continued to make solid progress in building out our global data center footprint in Q1.

As of the end of Q1, we operated mostly NVIDIA H200s, with the remainder being NVIDIA H100s. As we start rolling out NVIDIA Blackwells in Q2'25 and NVIDIA Grace Blackwells and NVIDIA Blackwell Ultras in 2H'25, we expect to significantly increase our capacity in terms of H100-power-equivalent GPUs in service by the end of this year.

We have significantly expanded our data-center network, adding a number of new locations in addition to existing facilities in Finland and France:

- **Iceland.** This colocation opened in March 2025.
- **Kansas City, US.** This colocation came online in April 2025 and is now fully operational. We will start rolling out NVIDIA B200 GPUs in Q2'25.
- **New Jersey, US.** Announced in March 2025, this build-to-suit facility is being constructed to our custom design specifications. Construction is well underway, and we plan to make the site's first capacity available in late summer 2025.

We expect that our global data center footprint will amount to approximately 100 MW of contracted capacity by the end of the year, and we plan to significantly grow our capacity in 2026.

Our robust cloud software offering is solving real needs for customers and is used by nearly all of our managed customers.

Nebius AI Cloud — the core of our software layer — enables us to serve a diverse range of customers, from small startups to large enterprises, and to quickly deploy new offerings at scale. We believe it can both drive greater customer demand for our business and generate sustainable higher margins over time.

As well as reliable access to infrastructure, our customers demand efficient, convenient tools to manage and optimize it and extract the most value from every unit of compute. This is where Nebius AI Cloud plays a central role.

Our infrastructure-as-a-service layer sits on top of our hardware, and enables us to serve the varied needs of our customers. For example, our proprietary orchestration algorithms enable us to rapidly deploy thousands of GPUs for a customer quickly so they can promptly start their workloads.

Nearly all of our managed customers use Nebius AI Cloud solutions including Kubernetes and managed Slurm — both of which automate and organize how computing jobs are scheduled and run across clusters — as well as networking, storage and performance monitoring tools.

We also offer a comprehensive MLOps suite designed to simplify machine learning operations, which is especially valuable for customers who do not maintain dedicated in-house MLOps teams. Nebius AI Cloud also integrates advanced monitoring tools that continuously track infrastructure performance, ensuring seamless and uninterrupted service.

In Q1, we introduced approximately 50 new software capabilities to improve usability, scalability and performance across our platform. Some of our notable launches were:

- **Container Registry update.** This tool, which allows teams to store and distribute containers, now supports parallel requests at higher scale and integrates more easily with continuous integration tools.
- **Slurm-based cluster upgrades.** Slurm is used to manage workloads that run across large GPU clusters. We introduced automatic recovery for failed nodes, proactive system health checks to detect issues before jobs fail, and topology-aware training that optimizes for cluster topology. These changes reduce downtime and increase job efficiency.
- **Enhanced object storage.** Now supporting read speeds of up to 10 gigabytes per second and write speeds of 5 gigabytes per second per compute node. This ensures that large datasets — often running to hundreds of terabytes — can be accessed and saved quickly during model training runs, reducing time to results.
- **New integrations.** We expanded integrations with external AI platforms such as Metaflow (a popular tool that helps manage end-to-end machine learning workflows) and SkyPilot (which helps teams run AI jobs across multiple clouds). These integrations allow customers to bring their existing tools into our ecosystem with minimal friction.
- **Prebuilt tools.** We launched software development kits (SDKs) for developers who use Go or Python. These SDKs simplify the process of interacting with Nebius AI Cloud without needing to write low-level code or work directly with complex APIs.

We also introduced a cross-functional certification roadmap, developed in close collaboration with product, engineering and customer teams. Based on identified needs, we've begun the process of validating our operations against key standards such as ISO/IEC

27000 series, SOC 2, HIPAA and others tied to AI management and business continuity. These efforts will continue throughout 2025 and will position Nebius to serve a broader set of enterprise and compliance-driven customers.

AI Studio continues to gain traction.

AI Studio, our inference-as-a-service platform, continues to gain solid customer traction with over 60,000 registered users as of quarter end. In Q1, we launched several new products and updates:

- We made fine-tuning generally available for all users.
- We expanded the catalog of open-source models available to include six new models, including Google's Gemma 3 27B; and DeepSeek R1, V3—0324, and R1-Distill-Llama-70B.
- We increased rate limits for scaling and added integrations with developers tools like Hugging Face, Helicone, LlamaIndex, and Postman.

While still early from a revenue perspective, we believe AI Studio could become a solid, high-margin contributor to our revenue over time.

We delivered annualized run-rate revenue of \$249 million, up 175% quarter-over-quarter, driven by increased capacity and strong execution.

ARR reached \$249 million in March, driven by two factors: increased capacity; and strong sales execution, which resulted in a record number of managed customers who came to Nebius to help them accelerate their AI workloads.

Our investments in sales and go-to-market teams are translating to customer success.

Our investments in our sales force, as well as pre-sales and customer support teams, are starting to drive solid business outcomes. Our solutions architects team works with prospective customers to help them set up large-scale proofs of concept so that they can test our network. And our customer support team, which includes engineers in the US and Europe, enables us to provide assistance to customers around the clock.

Customers continued to choose Nebius in Q1 for our powerful software and services that deliver:

- fast, efficient GPU cluster orchestration enabling production deployment in days;
- advanced observability tools for optimized performance;
- reliability with guaranteed uptime as high as 99.98%, supported by seamless and automated equipment health checks; and
- comprehensive support from pre-sales to post-sales, including direct access to engineering when needed.

Our newly hired sales team is also ramping up rapidly and bringing in more diverse clients across a wide range of industries, from technology providers to media and entertainment, to healthcare and life sciences, and more.

We also see substantial opportunities outside of our current customer base of mostly venture-backed AI-native tech startups. We are working hard to add customers among larger tech firms, including frontier AI labs, as well as enterprise customers that are interested in AI inference workloads to drive improvements and performance in their businesses.

Our brand building and partnership efforts are starting to pay off.

In under a year, the Nebius brand has gained strong traction with industry analysts and key partners. Independent research firm SemiAnalysis recently named Nebius in its Gold tier of cloud providers — alongside Azure, and above other hyperscalers. The ranking cited Nebius's deep engineering expertise and lower total cost of ownership as major strengths.

We also saw continued strength in our long-standing relationship with NVIDIA:

- NVIDIA announced that Nebius will be among the first AI clouds to offer the new NVIDIA Blackwell Ultra AI factory platform. By the end of 2025, our customers will have access to GB300 NVL72-powered instances, featuring 72 NVIDIA Blackwell Ultra GPUs for unmatched AI acceleration.
- Nebius joined the NVIDIA Dynamo ecosystem, an open-source inference serving framework designed to deploy generative AI across large-scale, distributed environments. NVIDIA Dynamo delivers one of the most efficient solutions for scaling compute during inference.
- Nebius was named as one of five Reference Platform NVIDIA Cloud Partners (NCPs) that specialize in delivering AI-accelerated services built on the NCP reference architecture.

In Q1, we also extended our core storage capabilities beyond our purpose-built file and object offerings through strategic partnerships with DDN, VAST Data, and WEKA. Our goal is to provide a full range of data solutions for all workloads, from the highest levels of performance to exabyte scale.



Avride

In Q1, Avride — which develops and operates delivery robots and robotaxis — continued to scale, with a number of key milestones and partnerships.

Avride entered into a partnership with Hyundai to co-develop a full autonomous vehicle platform.

In March, Avride announced a partnership with Hyundai Motors for the co-development of a full autonomous vehicle platform, including redundant safety systems such as steering and braking. Hyundai will handle vehicle manufacturing and assembly, enabling Avride to focus on software development and systems integration.

The partnership also provides a pathway for Avride to scale its AV fleet, starting with the addition of 100 Ioniq vehicles in 2025 and further expansion planned in the years ahead. Avride and Hyundai plan to begin testing a robotaxi service in partnership with Uber by the end of this year in Dallas.

Food delivery with Avride robots continued to scale across US university campuses and urban centers.

In January, Avride launched delivery operations at Ohio State University in partnership with Grubhub. The fleet now includes over 100 robots completing more than 1,200 deliveries per day. University campuses remain a strong fit for the service, and Avride is preparing to launch additional campuses in Q3 to align with the start of the academic semester.

In Jersey City, Avride robots are delivering Uber Eats orders from local restaurants, with coverage gradually expanding to larger diners. Avride is also growing its footprint in Austin and Dallas, continuing to expand service areas and onboard new restaurant partners.

International interest is growing, with Tokyo marking Avride's first entry into Japan.

Avride deployed its initial fleet of delivery robots in Tokyo, through a partnership with Rakuten. There is strong interest from local authorities and other autonomous tech stakeholders. This reinforces Japan's long-term potential as a high-quality, attractive market.



Toloka

Toloka is a data partner for LLM and generative AI developers, delivering scalable, high-quality solutions for all stages of AI development, including training, fine-tuning, alignment, and evaluation.

Strategic investment from Bezos Expedition positions Toloka for accelerated growth.

In May, Nebius announced a strategic investment in Toloka, led by Bezos Expeditions with participation from Mikhail Parakhin, CTO of Shopify. The investment will enable Toloka to scale rapidly and capitalize on the substantial market opportunity in AI data, while positioning Toloka among the top tier of global AI data companies.

Nebius will retain a significant economic stake in Toloka, securing long-term upside for shareholders from its continued growth. At the same time, Nebius will relinquish majority voting control, giving Toloka greater governance independence and flexibility while focusing on its core AI infrastructure business.

Toloka demonstrated consistent growth throughout Q1.

Toloka delivered strong and consistent performance in Q1 2025, with revenue more than doubling year-over-year. This growth was driven by a combination of market demand and disciplined execution. All major frontier lab contracts were renewed, reaffirming Toloka's position as a trusted partner for high-impact AI data solutions.

Toloka deepened engagement with frontier labs and diversified its revenue base.

Toloka added leading AI developers and enterprise clients to its pool. Its current customer base includes top technology companies such as Anthropic, Amazon, Microsoft, Shopify, Recraft, and others. As AI agent-based computing accelerates, Toloka is differentiated by its ability to solve some of the most technologically complex data challenges in the industry, setting it apart from other players in the AI data space.

The TripleTen logo is displayed in white lowercase text on an orange background. The background features a faint, stylized graphic of concentric circles and overlapping shapes.

TripleTen

TripleTen is an edtech platform focused on reskilling individuals for careers in technology.

TripleTen doubled its year-over-year revenue, supported by recent program launches.

TripleTen's Q1 2025 revenue grew 144% year-over-year driven by a nearly 50% year-over-year increase in student enrollment across key markets in the US and Latin America. Growth was boosted by the successful roll-out of tech-focused bootcamp programs launched in late 2024, including Cybersecurity and UX/UI Design. Customer acquisition costs remained stable quarter-over-quarter even as the company continued to invest in brand and reach.

TripleTen focused on enhancing study experience and productivity and recorded first revenue in the B2B segment.

In Q1, TripleTen recorded its first revenue from B2B training products, a new strategic pillar that leverages TripleTen's platform capabilities. The business also enhanced its offerings to include AI-enhanced learning tools aimed at improving engagement and outcomes. Together, these developments position TripleTen's expertise to address a broader learner base and unlock new channels for growth through 2025 and beyond.

The ClickHouse logo is displayed in black text on a yellow background. The logo consists of four vertical bars of increasing height followed by the word "ClickHouse".

ClickHouse

ClickHouse is a database software provider that we believe is growing quickly, and competes with firms like ElasticSearch, MongoDB and Snowflake. We own approximately 28% economic stake in this business.

Management team update



Maria del Dado Alonso Sanchez, Chief Financial Officer

We are pleased to announce the appointment of Dado Alonso as Chief Financial Officer of Nebius Group, effective 1 June. Dado brings over 25 years of international finance leadership across high-growth tech and consumer companies including Amazon, Booking.com, Naspers/OLX, and BBG. She brings extensive experience in capital markets, operational finance, and digital business models, and will play a key role in supporting Nebius as we scale our global AI infrastructure platform. We would also like to extend our thanks to Ron Jacobs, who is stepping down as CFO after playing a pivotal role in building our finance function during the company's restructuring. We are grateful for Ron's dedication and contribution during an important defining phase of the company's formation.

Financial Update

Revenue

	Three months ending March 31		
	2024	2025	Change
Revenue	11.4	55.3	385%

Group revenue was \$55.3 million, up 385% year over year.

- Our core AI infrastructure business represented almost three quarters of total group revenue in Q1'25. Annualized run-rate revenue was \$249 million at the end of the quarter.
- As a large part of our GPU capacity additions came online towards the end of Q4'24, we entered Q1'25 with substantially more available capacity to sell. In March 2025 we also made available the first capacity at our newly launched Iceland cluster.
- In addition, the sales and go-to-market team that we have been building out over recent quarters also made solid contributions, resulting in selling out most of our capacity in March 2025.
- The overall demand landscape has been positive as AI use cases continue to be proven out and adoption is growing, and our client base is expanding.
- Advancements such as the release of the DeepSeek R1 open weights model in January provided an immediate demand boost for our available H200 capacity. More broadly, we believe that the continued pace of such innovations and efficiency gains on models and applications are supportive of faster and larger scale AI adoption and fueling infrastructure demand.
- Other businesses accounted for just over a quarter of group revenue.
 - TripleTen continued to deliver strong year over year revenue growth in Q1 2025 supported by the addition of over 4,000 new students across the US and Latin America and the launch of new programs.
 - Toloka's revenue grew over 100% year-over-year, supported by positive market demand and expansion of the customer base into new highly promising accounts from among the leading global AI labs. Following the recently announced funding round, we will be de-consolidating Toloka's revenue from our financials going forward.
 - Avride's revenue contribution in Q1 was not material.

Operating expense

In USD \$ millions

Three months ending March 31

Expense category	2024	2025	Change
Cost of revenues	8.9	29.5	231%
as a percentage of revenues	78%	53%	
Product Development	25.2	40.0	59%
as a percentage of revenues	221%	72%	
Sales, general and administrative	51.3	66.1	29%
as a percentage of revenues	450%	120%	
Depreciation and amortization	8.9	49.2	453%
as a percentage of revenues	78%	89%	
Total operating costs and expenses	94.3	184.8	96%
as a percentage of revenues	827%	334%	
Total share-based compensation expense	5.9	17.6	198%
as a percentage of operating expenses	6%	10%	

Cost of revenue for the group in 1Q 2025 was \$29.5 million, up 231% year over year, equivalent to 53% of revenue for the quarter, down from 78% of revenue in 1Q 2024. An important driver for cost of revenue has been the scaling of our core AI infrastructure business with the launch of new clusters in colocation facilities in France in 4Q 2024, along with Iceland and the U.S. in 1Q 2025, with lease payments contributing to cost of revenue.

Product development expenses for the group were \$40.0 million, up 59% year over year, or 72% of revenue, down from 221% of revenue in 1Q 2024. The growth in product development expenses was primarily driven by expansion of the engineering teams.

Sales, general and administrative expenses for the group were \$66.1 million, up 29% year over year, equivalent to 120% of revenue, down from 450% of revenue in 1Q 2024. In 1Q 2025, share based compensation ("SBC") expenses accounted for \$11.0 million of our sales, general and administrative ("SBC") expenses. Excluding SBC costs, growth year over year was just 13%, reflecting ongoing investments in the growth of the business, including

expansion of the sales force, go-to-market activities and build out of the corporate platform to support our status as a publicly traded company.

Depreciation and amortization expenses for the group were \$49.2 million, or 89% of revenue for the period, compared to 78% of revenue in 1Q 2024. The key driver of the increase in D&A expenses was the substantial ramp up in investments in GPU capex and related data center hardware through 4Q 2024 and 1Q 2025 for the core AI infrastructure business. The depreciation of these fixed assets begins after delivery, whereas revenue generation from newly deployed GPUs comes with a lag. Furthermore, we depreciate our hardware over a four-year period which we believe is conservative when compared to some other cloud providers.

Adjusted EBITDA

In USD \$ millions

Three months ending March 31

	2024	2025	Change
Adj. EBITDA (loss)	-70.9	-62.6	-12%
as a percentage of revenues	-622%	-113%	

Adjusted EBITDA loss for the group was \$62.6 million, an improvement of \$8.3 million compared to 1Q 2024. This was driven primarily by the growth of our core AI infrastructure business as well as Toloka. Our investments in Avride, TripleTen and Toloka contributed more than half of the adjusted EBITDA loss for the quarter.

Guidance

The company will share its guidance on its Q1 2025 earnings call.

Earnings webcast

Nebius Group (Nasdaq:NBIS) will host a conference call and earnings webcast at **5:00 a.m. Pacific time/8:00 a.m. Eastern time/2:00 p.m. Central European Time on May 20, 2025** to discuss these financial results. To register to participate in the conference call, or to listen to the live audio webcast, please visit Nebius's Investor Relations website at group.nebius.com/investor-hub.

A replay will be available on the same website following the call.

Forward-looking statements

This document contains forward-looking statements that involve risks and uncertainties. All statements contained or implied other than statements of historical facts, including, without limitation, statements regarding our review of strategic options to accelerate growth, business plans, market opportunities, capital expenditure requirements, financing requirements and projected financial performance, are forward-looking statements. In some cases, these forward-looking statements can be identified by words or phrases such as “may,” “will,” “expect,” “anticipate,” “aim,” “estimate,” “intend,” “plan,” “believe,” “potential,” “continue,” “is/are likely to” or other similar expressions. In addition, these forward-looking statements reflect our current views with respect to future events and are not a guarantee 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, our ability to successfully operate and develop a fundamentally different, early-stage group following the divestment of a significant portion of our historical operations; to implement our business plans; to continue to successfully capture customers; to continue to successfully obtain required supplies of hardware on acceptable terms; and to obtain any further debt or equity financing that may be necessary to achieve our objectives. Many of these risks and uncertainties depend on the actions of third parties and are largely outside of our control. We also continue to be subject to many of the risks and uncertainties 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 U.S. Securities and Exchange Commission (“SEC”) on April 30, 2025, which are available on our investor relations website at <https://group.nebius.com> and on the SEC website at www.sec.gov. All information in this document is as of May 20, 2025, and the Company undertakes no duty to update this information unless required by law.

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 of this document, 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.

Contact investor relations:
askIR@nebius.com

NEBIUS