

Overview of Nebius Group

About Nebius Group

Introduction

Amsterdam, the Netherlands, July 16, 2024 — Nebius Group is building one of the largest commercially available Artificial Intelligence (AI) infrastructure businesses, based in Europe. The group is led by Arkady Volozh, the entrepreneur and visionary co-founder of Yandex.

Our core business is an AI-centric cloud platform built for intensive AI workloads. We are building full-stack infrastructure to service the explosive growth of the global AI industry, including large-scale GPU clusters, cloud platforms and tools and services for developers.

One of our greatest assets is a team of more than 1,000 top-tier AI engineers who comprise the Group's R&D backbone and have a long track record of working together to develop world-class tech and AI infrastructure. By leveraging their skills and experience, we are developing a global AI business that integrates essential components for successful AI development: infrastructure, data and expertise.

Our headquarters and main R&D presence is in Amsterdam, with additional R&D hubs in Europe, North America and Israel. Our businesses already operate and serve customers worldwide.

In addition to our core AI infrastructure business, we are developing three other businesses that will operate under their own distinctive individual brands:

- [Toloka AI](#) – a data partner for all stages of AI development from training to evaluation;
- [TripleTen](#) – a leading edtech player in the US and certain other markets, re-skilling people for careers in tech;
- [Avride](#) – one of the most experienced teams developing autonomous driving technology for self-driving cars and delivery robots.

We will be flexible in our strategic approach to the future development of these businesses, including by seeking external investment or co-investment into individual businesses from strategic or financial investors where we believe this can increase value for our shareholders.

We also own a minority stake in ClickHouse, the creator of the popular open-source column-oriented database management system, which was spun out of the group in September 2021.

We believe that this is just the beginning. Our goal is to continue to innovate, experiment and invest in building additional businesses that tap into promising segments of the AI economy.

We see a significant opportunity to invest in building the leading AI infrastructure player in Europe, with a multi-billion annual revenue potential. This will require substantial capital investments, for which we expect to draw on a variety of funding sources including Nebius Group's balance sheet as well as public and private equity or debt.

Since we went public in 2011, we have developed and maintained a strong and trusted relationship with the capital markets, and the markets have always rewarded us in the past for being good stewards of capital. Maintaining that positive track record going forward will be an important priority for us as we look to seize the next big market opportunity.

Today we have total cash of approximately USD 2.5 billion and no debt. After determining appropriate reserves for tax and other potential liabilities in connection with the successful completion of the divestment, we will retain part of the net proceeds to finance the growth of our businesses, while returning a substantial proportion to our shareholders. We expect to provide updates on our intentions in this regard in due course.

We will remain an SEC reporting company and intend to retain our listing on the Nasdaq Global Select Market, and hope that trading in our shares will resume in due course. We believe that being a public company will give us long-term access to capital on favorable terms as we seek to scale our existing businesses and create new ones.

For investors, this will create an opportunity to invest in a pure-play AI infrastructure stock outside of traditional US “big tech”, while benefiting from the transparency and liquidity that public status offers.

Our core AI infrastructure business

Business model overview

Nebius is an AI-centric cloud platform building large, cost-efficient GPU clusters to service the explosive growth of the global AI industry. With proprietary cloud software architecture and hardware designed in-house (including servers, racks and data center design), Nebius gives AI and ML developers the compute, storage, managed services and tools they need for training, inference and data processing.

Nebius builds a full-stack infrastructure for AI developers - from owned or co-located data centers, where it deploys acquired GPUs, to a proprietary cloud platform to help clients train, develop and run their AI models. Thanks to its strong long-term relationship with NVIDIA as a cloud partner and also an OEM partner Nebius enjoys preferential access to the latest and most powerful GPUs. The Nebius team has more than 15 years of experience working with original design manufacturers (ODMs) to co-design and manufacture hyperscaler-type servers, reducing the total cost of ownership and increasing operational efficiency.

Nebius offers on-demand access to GPUs as well as reserved capacity plans, targeting two groups of clients:

- Developers of LLM models with diversified consumption profiles – who require clusters for training and inferencing;
- Developers of end-user AI applications – who use Nebius’s LLM-as-a-Service by tokens.

Key infrastructure

Nebius has a highly energy-efficient data center in Finland that is home to the most powerful commercially available supercomputer in Europe (“ISEG”), which is among the 20 most powerful anywhere in the world. The data center matches the energy efficiency of top-performing data centers globally, with overhead energy consumption as low as 20% of the global average.

In addition to expanding the capacity of our Finnish data center, we plan to build greenfield data centers at new locations primarily in Europe, as well as deploying additional GPU capacity at colocation data centers with modified architecture to reduce installation time.

Market opportunity and competitive strengths

The global GPU-as-a-service market was estimated at USD 3.2 billion in 2023 and is projected to reach about USD 50 billion by 2032, a CAGR of more than 35% (based on Fortune Business Insights). This growth is being driven by increased penetration of ML and AI based applications in business and consumer contexts, leading in turn to increased demand for data center capacity and higher spending on public cloud. Gartner expects an increase of 20% in worldwide end-user spending on public cloud services to USD 675 billion in 2024, primarily as a result of adoption of generative AI.

Nebius primarily competes with other Cloud providers, including Coreweave, Lambda Labs, Together.ai, and DeepInfra, as well as hyperscalers (AWS, GCP, Azure, Oracle).

Nebius' key competitive advantages include:

- **Sole focus on AI**, allowing us to provide solutions precisely tailored to meet the specific needs of AI developers and businesses.
- **Full control over the whole value creation chain**, from in-house server design to proprietary cloud platform, ensuring maximum productivity and quick scaling, as well as higher cost efficiency and low downtime.
- **A team of AI/ML & cloud engineers** with a track record of building cloud services and infrastructure from scratch.
- **20-25% lower total cost of GPU ownership and operations** compared to an average GPU provider.
- **Long-standing collaboration with NVIDIA** across hardware and cloud. Nebius is a launch partner for NVIDIA's next-generation Blackwell platform, which will be available to Nebius clients as soon as it is launched
- **Long term relations** (more than 15 years) **and collaboration experience with leading server ODMs** for co-designing, manufacturing and optimizing cost structure on servers.
- **Expertise in running data centers with heavy power loads** (hundreds of megawatts) in a reliable and cost-efficient way.

Financial highlights

We see a large and rapidly growing addressable market and substantial demand, which together with our cost-efficient business model justifies significant additional and ongoing investments into capacity expansion. As we continue building out our AI infrastructure stack, we will focus on rapidly scaling up in two main areas:

- expansion of GPU capacity, with a focus on securing the next generation of NVIDIA chips; and
- increasing our data center capacity, including expanding our existing Finnish data center, deploying capacity at colocation data centers, as well as developing greenfield projects at new locations.

Over the next 6 to 9 months we believe Nebius can achieve ARR (annual recurring revenue) of about USD 200 million, and be in a position to cover all its cash operating expenses (before amortization and capital expenditures).

We also have started to expand our data center in Finland, and we plan to complete it over the next 12 months. In addition, we are securing additional capacity at colocation data centers. All these together will allow us to reach a total capacity of tens of thousands of GPUs. Full utilization of this capacity would allow Nebius to increase its ARR by several times.

Our people

Nebius' team is a key competitive advantage, with seasoned experts in hardware infrastructure, cloud software development and AI, as well as LLM and ML pioneers. The team brings experience from global big tech companies including Microsoft, Amazon, Yandex and others. The core of the team is based at Nebius' key operational hub in Amsterdam, with additional R&D hubs in Europe and Israel.

Along with CEO Arkady Volozh, Nebius is led by Chief Business Officer Roman Chernin, Chief Product and Infrastructure Officer Andrey Korolenko and Chief Operating Officer Ophir Nave.

Roman Chernin has 12 years of experience heading the development of digital services (including Search and Maps) at Yandex and is also an experienced startup mentor. Andrey Korolenko has nearly 20 years supervising IT infrastructure development at leading tech businesses and expertise in deploying supercomputers. Ophir Nave was a Lead Partner in the Corporate and M&A Practice at the Israeli firm Arnon, Tadmor-Levy, and previously worked for the U.S. law firm Wachtell, Lipton, Rosen & Katz.

IP and technology

Nebius is developing a new proprietary platform to replace existing legacy IP. Migration to the new platform and the phase-out of the legacy stack is expected by Q4 2024. The new platform has a redesigned architecture and software stack that enables efficient scaling across multiple data centers, and aims to dramatically reduce delivery timelines from quarters to weeks. Built on open-source technologies, it leverages the cloud and infrastructure expertise of our core engineering team, and our ability to balance in-house development with acquiring solutions from the ecosystem, ensuring that the resulting product meets the ever-growing requirements of AI engineers.

Toloka AI

Business model overview

Toloka AI is a data partner for all stages of AI development from training to evaluation.

In 2023, Toloka successfully pivoted its business model to providing high-quality training data for Generative AI, focusing on three key stages of LLM production – fine-tuning, alignment and evaluation – which require large volumes of human-annotated data. The pivot is based on using highly skilled contractors instead of drawing from a mass audience and is based on a hybrid approach combining a human backed by LLM instead of purely human input. These contractors are experts in their fields who help to enhance the accuracy and reliability of AI models.

As a result of this strategic pivot, the Generative AI segment is already Toloka's primary source of revenue.

Toloka builds on strong technological expertise developed over more than 10 years in the data-labeling space. The technology that underpins Toloka's business allows the creation of automated pipelines of human input in a way that is measurable, scalable and manageable.

To support the pivot to the Generative AI-focused business model, Toloka's strong R&D function has acquired crucial expertise in LLM production and associated data requirements and has organized a series of research events at leading AI conferences.

Toloka's current operational capacity has significant potential for further rapid expansion. Toloka can scale supply fast, and unlock new domains and languages, while keeping healthy unit

economics. The company currently has more than 5,000 experts already onboarded and is able to grow capacity at a pace of approximately 1,500 new experts per month.

Market opportunity and competitive strengths

We estimate the market for high-quality, human-annotated data for Generative AI is currently worth around USD 2 billion, which will grow at more than 40% CAGR for next 5 years.

For successful model development, Generative AI needs huge amounts of high-quality data, which has to be accurate, diverse, legally compliant and produced at scale. This is a complex task and one of the bottlenecks of modern AI development.

Toloka's customers include prominent names (big tech, Fortune-500 companies and A-class startups backed by top investors) across the entire spectrum of Generative AI companies, from the largest foundation model producers in global big tech to emerging AI-powered start-ups, and enterprises who are starting to integrate Generative AI solutions.

There are few companies with **both the expertise and technological infrastructure to build robust and high-quality human data operations at scale**. This is the market opportunity that Toloka is positioned to capture. To scale fast and efficiently, Toloka invests in technological solutions like marketplace, quality control systems and expert copilots.

Toloka also anticipates increasing demand for data evaluation services for GenAI development, as models are deployed in production in increasing numbers. This is a whole new market that provides additional opportunities for significant growth and expansion of the total addressable market.

Financial highlights

Toloka's Generative AI offering is forecasted to grow multiple times this year, becoming Toloka's major source of revenue.

Following initial investments in the GenAI pivot, gross margins have shown steady month-over-month growth over the last quarter, reaching an expected range of 45-50% in the GenAI business stream. With continued focus on enhancing production efficiency, automation, and the use of synthetic data and auto-labeling, we see an opportunity to further improve our gross margins.

Our people

The Toloka team currently includes around 200 professionals with deep expertise in ML, engineering, and marketplaces. The core management team – including the CEO, CTO and CPO – is based in Amsterdam and includes alumni of Google, Microsoft, SAP, Yandex, McKinsey, Bain and Y Combinator-backed startups. Toloka's founder and CEO is Olga Megorskaya, who has over a decade of experience in AI and data labeling, including extensive expertise managing distributed tech- and non-tech teams.

IP and technology

Toloka is transitioning from its legacy crowdsourcing platform, which is set to be shut down by the end of 2024, towards a new focus on Generative AI solutions.

The pivot from crowdsourcing to Generative AI data partner requires a radically different technology platform, which Toloka is currently building from scratch. The new platform entails

creating data for Generative AI models instead of just labeling data and designed with enhanced safety, cybersecurity and privacy features.

While the new platform is expected to be fully up and running by the end of this year, Toloka is launching certain products and parts of the new platform as they are ready. In April 2024, Toloka also launched Mindrift, a platform for efficient contracting and communicating with highly-skilled experts to support the development of Generative AI projects.

TripleTen

Business model overview

TripleTen is a leading edtech platform developed by a team with over 15 years of experience, specializing in reskilling and upskilling individuals for successful careers in tech.

TripleTen offers affordable, high-quality training through a proprietary learning platform with course content developed in-house. It currently offers four immersive study tracks – Software Engineering, Data Science, Business Intelligence Analytics and Quality Assurance – with two more planned for launch in 2024: Cyber Analytics and UX/UI Design.

The business model leverages a blend of bootcamp and massive open online course (MOOC) formats. This means that TripleTen can offer highly sought-after personalized learning while also ensuring course scalability by being able to enroll thousands of participants simultaneously.

TripleTen also provides career services to its graduates, and ranks among the top US edtech providers in terms of graduates' success in finding employment (based on Fortune and market data). The company currently has more than 40 partners that offer job opportunities to our graduates, and will continue to invest in growing this network.

TripleTen is primarily a B2C business targeting individuals from diverse backgrounds who want to break into tech roles; over 1,000 people enroll in TripleTen's programs monthly.

Market opportunity and competitive strengths

TripleTen capitalizes on growing demand for reskilling and online education driven by the widening tech skills gap created by rapid digitalization. The global digital education market is estimated to reach USD 33.2 billion by 2025 ([JP Morgan, 2023](#)).

TripleTen's primary markets include the US, Latin America and Israel, where demand for skilled tech talent is high and growing. Programs are accessible remotely across these locations.

TripleTen has also recently launched a B2B offering targeting companies that are responding to technological disruption by seeking to close IT skills gaps, embrace AI and retain talent by upskilling staff in high-demand areas. Some companies are allocating as much as 1.5% of their total budgets to reskilling – comparable to what many firms spend on IT deployment ([BCG, 2023](#)). TripleTen offers B2B clients AI tech courses for beginners (AI Literacy) as well as for professionals who are already working in AI.

TripleTen competes primarily with other bootcamps (such as General Assembly, Springboard and Thinkful) and, to an extent, with MOOC platforms. The company ranks among the best edtech providers in the US based on employment rates, median post-graduation salaries, and student feedback, according to the market data. In June 2024, TripleTen was named best overall provider in [Fortune magazine's ranking](#) of software engineering bootcamps.

TripleTen's key competitive advantages include:

- **Recognition and customer loyalty.** Referrals from graduates contribute significantly to reducing per customer marketing spend and investing in expanding B2C and B2B course offerings.
- **Track record of successful employment.** TripleTen is a top-rated edtech company in the US by employment rate and median salary after graduation. 87% of graduates find a job within six months of graduation, ahead of the 77-85% reported by most key peers.
- **Affordable pricing.** Enabled by optimization of training costs through diversified sourcing of trainers (hiring qualified English-speaking talent globally and using part-time educators for collaborative teaching), automation of e-learning workflows (from automated task dispatch based on student skill level and availability to built-in machine translation for code reviews), and AI-enabled student support and guidance.
- **Flexible e-learning environment** that enables seamless launching, scaling and localization of courses at minimal additional cost.

Financial highlights

TripleTen's revenue is primarily driven by cash inflows from enrolled students. Student numbers tripled in Q1 2024 compared to Q1 2023. Bookings (total value of anticipated payments from students who have made at least one initial payment for a course) year-to-date in 2024 have grown five-fold year-on year, with the run-rate Bookings value estimated to exceed USD 60 million by year-end.

TripleTen's biggest expense is customer acquisition costs, which have decreased by 80% year-on-year on a per-customer basis. The company achieved an important milestone in Q3 2023, with customer acquisition costs across all markets fully covered by initial customer payments, driven by referrals from graduates, higher upfront payments and streamlined sales processes.

Our people

TripleTen's team numbers nearly 150 professionals specializing in education, development, product management and marketing. R&D is a significant focus, accounting for a quarter of employees. Most educational specialists on the team have more than 15 years of experience as well as deep local expertise. The company's CEO is Ilya Zaleskiy, who has extensive experience building edtech ventures, and strong expertise in startup consulting and go-to-market strategies.

Avride

Business model overview

Avride develops autonomous driving technology for self-driving cars and delivery robots. The Avride team builds on a successful track record of delivering and commercializing projects across North America and Asia over the past seven years. Avride is headquartered in Austin, Texas, with additional offices in Tel Aviv, Belgrade and Seoul.

Avride is uniquely positioned in the industry by simultaneously developing and operating two products – autonomous cars and delivery robots – which leverage shared and mutually enhancing technologies. As a commercially deployed project, delivery robots enable Avride to accumulate valuable commercial expertise for the future deployment of its self-driving cars.

Avride's technologies target autonomous transportation within sectors such as ride-hailing, logistics, e-commerce and food and grocery delivery, with use cases including passenger rides, hub-to-warehouse deliveries and package deliveries to end customers.

Avride has deployed its delivery robots commercially for restaurant deliveries, grocery deliveries and small-scale logistics, in partnership with a number of food delivery companies.

Market opportunity and comprehensive strengths

The development of autonomous technologies presents significant opportunities for improving transportation safety and efficiency, and also reduces operational costs by optimizing labor expenses in ride-hailing and logistics. The market size for robotaxis is expected to reach USD 45.7 billion in 2030 (Markets&Markets Forecast, 2023), while Avride's total addressable market for last-mile delivery with robots is forecast to grow to around USD 2.2 billion in 2030 (IndustryArc, 2024).

Avride plans to expand its operations in the USA, Asia and globally. The company is currently testing its autonomous vehicles and has commercial delivery projects in the USA and South Korea.

Avride's key competitive advantages include:

- **One of the most experienced teams in the autonomous vehicles industry, standing behind the launch of Europe's first robotaxi**, which has given 65k+ passenger rides since 2018. The fleet designed by the Avride team was among the first to start giving passenger rides. It has been tested in a wide range of weather and road conditions to ensure safety, and has achieved a milestone of 22 million autonomous kilometers driven on public roads.
- **Extensive experience of commercial deployment**, with successful projects in North America and Asia. Delivery robots developed by the team have completed over 200,000 food deliveries.
- **Track record of partnerships** with leading automakers as well as Tier 1 OEMs and foodtech platforms.
- **Outstanding safety record**. Safety is a core focus, and the fleet developed by the Avride team has driven over 22 million kilometers with no serious accidents recorded.
- **Broad range of potential applications**. Avride's delivery robots have been tested in snowy, frosty and rainy weather conditions as well as on busy streets with heavy vehicle and pedestrian traffic, and are ready for complex road conditions.

Our people

Avride's team of approximately 270 employees, including about 200 talented engineers, has extensive experience developing autonomous driving technologies from scratch in diverse regulatory, cultural and operational contexts, including road testing in comprehensive weather and road conditions. Avride's CEO is Dmitry Polishchuk, who has 13 years of experience leading mobile services, including mapping, navigation, and browser and over 7 years of experience heading the development of autonomous driving tech.

Nebius Group corporate information

We will give notice in the coming days of an Annual General Meeting of Shareholders to be held in August, at which we will ask shareholders to approve the change of the legal name of the

company to Nebius Group, and to consider and approve the nominations of several additional executive and non-executive directors, as well as other matters.

The Board of Directors currently consist of John Boynton (Chairman), Charlie Ryan and Rogier Rijnja. We will be proposing five additional directors at our upcoming AGM; details on the director nominees will be provided in the AGM materials.

We intend to hold an investor day in September to discuss our strategy, plans and growth opportunities. We will provide more details on this in due course.

Information on Nebius Group can be found at <https://nebius.group/>

Contacts

Investor Relations
askir@nebius.com

Media Relations
media@nebius.com

Disclaimer

Forward Looking Statements

This press release contains forward-looking statements that involve risks and uncertainties. All statements contained or implied in this press release other than statements of historical facts, including, without limitation, statements regarding our business plans, market opportunities, capital expenditure requirements, financing requirements and projected financial performance, are forward-looking statements. We can provide no assurance that we will be successful developing our retained or new businesses or achieving our goals in line with our plans. 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; 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. Notwithstanding the sale of our Russian businesses, 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, 2023 and "Risk Factors" in a shareholder circular filed as Exhibit 99.2 to a Report on Form 6-K filed with the U.S. Securities and Exchange Commission (SEC) on February 8, 2024, which are available on our investor relations website at <https://nebius.group> and on the SEC website at www.sec.gov. All information in this release is as of July 16, 2024, and the company undertakes no duty to update this information unless required by law.