

NEBIUS

# Nebius Group Investor Presentation

October 18, 2024

## Disclaimer

# Forward-looking statements

This presentation 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 presentation other than statements of historical facts, including, without limitation, statements regarding our future financial and business performance, our business and strategy, expected growth, planned investments and capital expenditure, 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: competitive pressures, technological developments, our ability to secure and retain clients, and our need to expend capital to accommodate the growth of the business, 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, 2023, filed with the Securities and Exchange Commission (“SEC”) on April 26, 2024, and “Risk Factors” in the Shareholder Circular filed as Exhibit 99.1 to our Current Report on Form 6-K dated February 8, 2024, which are available on our investor relations website at <https://group.nebius.com/sec-filings> and on the SEC website at <https://www.sec.gov/>.

All information and numbers in this presentation is as of October 18, 2024 (unless stated otherwise). The forward-looking statements made in this presentation relate only to events or information as of the date on which the statements are made in this presentation. 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 of this presentation, 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.

## Table of contents

<b>Introducing Nebius Group</b>	<u>4</u>	<b>Toloka</b>	<u>42</u>
<b>Nebius: our core AI infrastructure business</b>	<u>13</u>	<b>TripleTen</b>	<u>57</u>
Business overview	<u>14</u>	<b>Avride</b>	<u>71</u>
Competitive advantages	<u>26</u>	<b>Nebius Group Financials and Outlook</b>	<u>86</u>
Clients	<u>29</u>	<b>Appendices</b>	<u>90</u>
Growth plans	<u>35</u>		
Financials	<u>38</u>		

# Introducing Nebius Group

# Technology for Technologists

Strongly positioned to build one of the largest specialist AI infrastructure players

## Nebius Group at a glance

### Locations

Headquarters  
and main R&D hub:  
Amsterdam, NL

Additional R&D hubs in  
Europe, North America  
and Israel

## Global AI business

with ~850  
tech specialists<sup>1</sup>

## NASDAQ listed

with long-term access  
to capital

## \$2bn+ cash

on Balance Sheet<sup>2</sup>  
and no debt

(1) Refers to Nebius Group. As of September 30, 2024; (2) \$2.3bn total cash as of September 30, 2024, without giving effect to any potential return of capital to shareholders or withholding tax.

# Building a global AI business which integrates essential components for successful AI development: infrastructure, data, and expertise

## Our businesses and infrastructure

### NEBIUS

#### Nebius: our core business

AI-centric cloud platform built for intensive AI & ML workloads featuring both owned and collocation data center capacity, with ongoing expansion

### Toloka

Data partner for all stages of AI development from training to evaluation

### AVRIDE

Autonomous driving technology for self-driving cars and delivery robots

### tripleten

Leading edtech player, re-skilling people for successful careers in tech

### ClickHouse

Approximately 28%<sup>1</sup> stake in ClickHouse, creator of a popular open-source column-oriented DBMS

## Our data and expertise

**~850 Top-tier AI / ML / LLM, cloud engineers and other tech specialists<sup>2</sup>**

As our R&D backbone with a track record of developing world-class tech and AI infrastructure

(1) Approximately 28% stake on a fully diluted basis, upon exercise of penny warrants to acquire Common Stock and Series A and B stock; latest publicly available valuation of ClickHouse was \$2 billion following its Series B financing in October 2021; (2) Refers to Nebius Group. As of September 30, 2024.

## Illustrative value building blocks for Nebius Group

### Our core business: Nebius

State-of-the-art owned Finnish data center with ~14,000 GPUs paid and deployed<sup>1</sup>

Building specialized AI cloud (competing with Coreweave, Lambda, etc)

Full stack & efficient infrastructure platform

2025 ARR \$500-1,000m<sup>2</sup>

### Additional building blocks

Cash and access to financing

**\$2bn+**  
in cash on balance sheet<sup>3</sup>

**Access to capital markets**

Other group businesses

**Toloka**

**TripleTen**

**Avride**

Minority stakes

**Approximately 28% stake in ClickHouse<sup>4</sup>**

(1) The number of deployed GPUs is expected to exceed 20 thousand by the year end; (2) Depending on the amount of capital available to cover the company's investment program; (3) \$2.3bn total cash as of September 30, 2024, without giving effect to any potential return of capital to shareholders or withholding tax; (4) Approximately 28% stake on a fully diluted basis, upon exercise of penny warrants to acquire Common Stock and Series A and B stock; latest publicly available valuation of ClickHouse was \$2 billion following its Series B financing in October 2021.



Addressing a huge market opportunity across all of our businesses

## Over \$260bn total addressable market for Nebius

**The market for AI technologies is vast, and expected to grow up to \$800bn by 2030, with an average CAGR of 29% during 2023-2030<sup>1</sup>**

**NEBIUS**

**Total Addressable Market (TAM)**

**\$260bn+**

GPU-as-a-service and AI Cloud market in 2030<sup>2</sup>

**+35%**

2023-2030 CAGR<sup>2</sup>

Our three additional businesses offer an exposure to a combined 2030 TAM<sup>3</sup> of

**\$150bn+**

(1) Statista Market Insights, March 2024; (2) Nebius internal estimates based on a bottom-up analysis of publicly available disclosure on GPUs spending and GPU Cloud economics; the presented TAM includes estimates for two key markets: GPU-as-a-service and AI Cloud; (3) Internal estimates (Toloka TAM, based on our analysis on the client universe and assumptions on average spend by different client categories on the back of the company's data and industry checks); Persistence Market Research, 2024 (TripleTen TAM), Markets & Markets Forecast, 2023 (robo taxis TAM), and Grand View Research, 2023 (last mile delivery robots TAM).

Our R&D backbone

## A well-synced team with a track-record of developing world-class tech and AI infrastructure

**~1,300**

Total headcount as of Sep-24

Top-tier IT professionals  
in the most in-demand fields with an average of

**10+ years**

experience in AI / ML / LLM & other fields

Efficient intra- and cross-team  
collaboration enabled by

**Attuned teamwork**

and substantial  
previous experience working together

**~850 in-demand tech  
professionals<sup>1</sup>**

AI / ML / LLM engineers,  
Data scientists & analysts,  
Cloud engineers,  
Security architects, and other

**Big Tech mindset**

rooted in the background of tech team  
members, prior experience in leading  
tech companies in EMEA

**“Glocal” team**

to execute business growth  
including sales and marketing,  
leveraging **local** knowledge and  
**global** expertise

(1) Refers to Nebius Group. As of September 30, 2024.

## Significant opportunity to build a leading AI infrastructure player

The opportunity

# Multi-billion dollar annual revenue potential<sup>1</sup>

from pure-play in the fast-growing AI industry with access to our own AI infrastructure, the key bottleneck to the AI revolution

- **Transparency and liquidity access enabled by NASDAQ listing**
- **Flexible strategic approach to future development of our business**  
Potential to seek external investment or co-investment into individual businesses from strategic or financial investors where there is an opportunity to increase shareholder value

Funding needs

# Substantial capital investments

from the balance sheet, as well as public and private equity or debt

(1) Based on the planned capacity expansion and assumed utilization and pricing as per the company's medium-term financial plan (as of September 2024).

Why do we think we can be successful?

### Expertise

We have a strong team with extensive technical experience,

**having built 200MW+**

of cutting-edge infrastructure and

**cloud business from the ground up**

### Track record

We built one of the largest tech companies<sup>1</sup> in Europe with a

**\$30bn market cap**

and established a strong and trusted relationship with the capital markets

### Resources

We have

**over \$2bn in cash<sup>2</sup>,  
a big tech mindset,  
and the expertise**

to build the largest specialist AI infrastructure company, positioned to support the explosive growth of the AI industry

(1) Yandex N.V. reached ~\$30bn market cap in Q4 2021; (2) \$2.3bn total cash as of September 30, 2024, without giving effect to any potential return of capital to shareholders or withholding tax.

# **Nebius: our core AI infrastructure business**

**A leading AI-centric  
cloud platform built for  
intensive AI workloads**

# AI-centric cloud platform building large, cost-efficient GPU clusters to service the explosive growth of the global AI industry

Infrastructure with compute, storage, managed services and tools needed for AI model development:

**Training**

**Inference<sup>1</sup>**

... trusted by prominent clients, including:



## Our expertise & infrastructure

**~400**

Software and hardware engineers with full-stack expertise in AI cloud development and operations<sup>2</sup>

## Data center capacity

Energy-efficient own data center in Finland, colocation data center in Paris<sup>3</sup> and further expansion plans



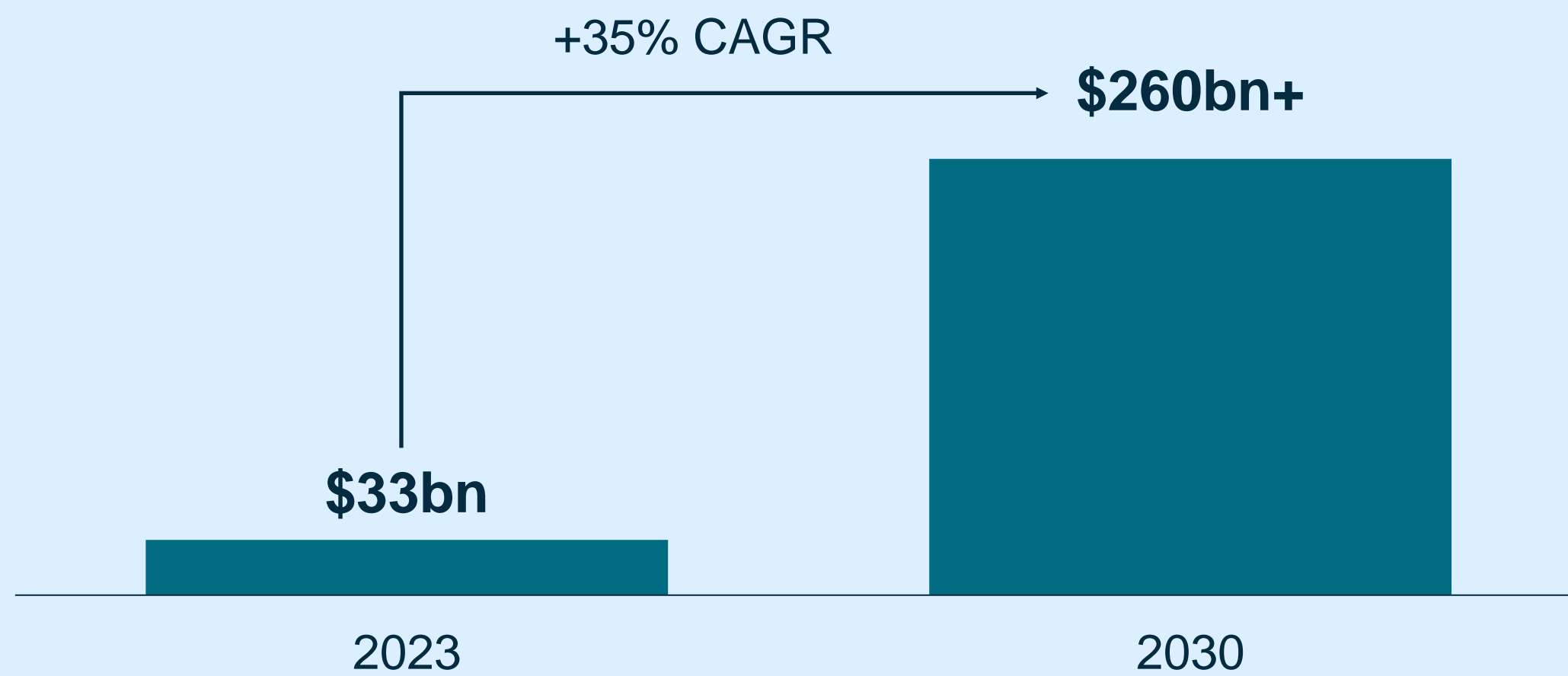
Long-standing collaboration across hardware and cloud as a Cloud and OEM partner in the NVIDIA Partner Network



(1) Primary driver in the AI industry, where inferencing workloads are expected to overtake training; (2) Refers to Nebius core business. Nebius Group has ~850 top-tier AI / ML / LLM, cloud engineers and other tech specialists. As of September 30, 2024; (3) Agreement signed in July 2024, new capacity in Paris, France to be put into operation by the end of November 2024.

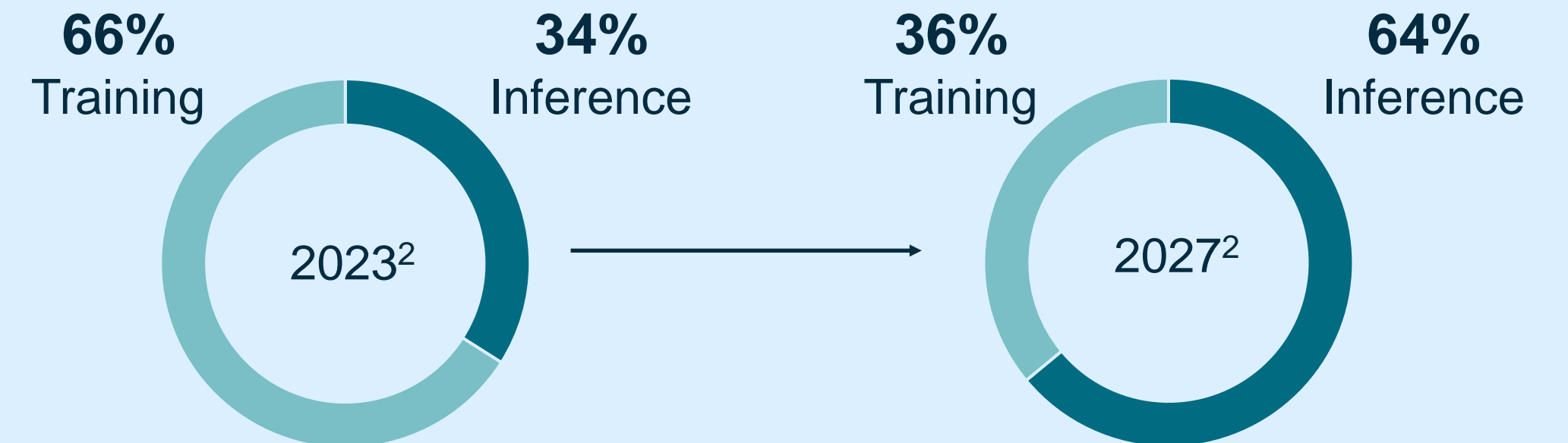
# Sustained demand growth in GenAI driving revenue growth

## Total addressable market<sup>1</sup>



## Growth will be mostly driven by inference workloads

### End-user spending for AI servers, by use



(1) Nebius internal estimates, Nebius internal estimates based on a bottom-up analysis of publicly available disclosure on GPU spending and GPU Cloud economics; the presented TAM includes estimates for two key markets: GPU-as-a-service and AI Cloud; (2) Gartner, 2023.



## Our core pillars provide full-stack infrastructure for AI developers, allowing us to maintain control over the whole value creation chain

### 1. Data centers

Improves unit economics through increased energy efficiency and lower data center cost, while enabling scalability

### 2. Hardware

In-house hardware R&D further improves business unit economics by providing resilient servers and reducing power consumption

### 3. Cloud platform

A proprietary cloud platform is essential for creating large GPU clusters optimized for extensive AI training and inferencing without performance bottlenecks

### 4. AI expertise

Dogfooding by in-house LLM team helps to adjust hardware and platform to the real needs of ML/AI practitioners

### 5. Experienced team

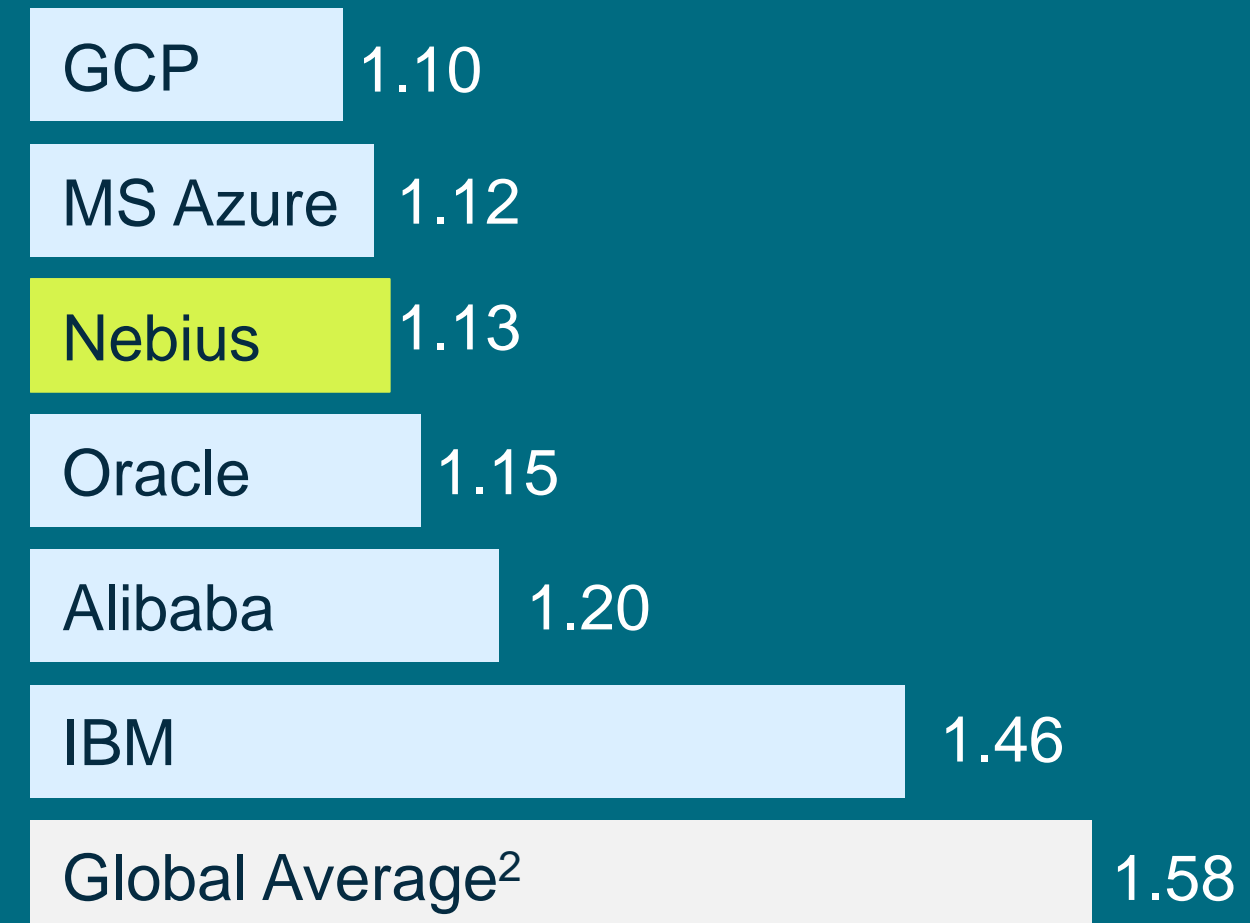
500+ employees, including ~400 engineers<sup>1</sup> experienced in building cloud businesses from the ground up, with expertise spanning code development, data center operations and business development

(1) Refers to Nebius core business. Nebius Group has ~850 top-tier AI / ML / LLM, cloud engineers and other tech specialists. As of September 30, 2024.

# 1. Data centers: efficiency gains from supercomputing

## Power usage effectiveness

(PUE)<sup>1</sup> (lower is better, ideal = 1.0)



## TOP 5%

energy efficient supercomputer among 500 most powerful supercomputers<sup>3</sup>

### #6

Europe's most powerful supercomputer<sup>4</sup>

### #19

Globally most powerful supercomputer<sup>4</sup>

Nebius supercomputer cluster **expanded 4x** since measurement<sup>5</sup>

## Energy-efficient supercomputer

Among the most potent commercially available supercomputers in Europe

### GPU server power consumption<sup>6</sup>, kW

Standard server<sup>7</sup>

10.0

Nebius server

8.2

**A**  
Up to ~-(20)%

**Nebius aims to replicate these best practices when expanding its computing capacity**

(1) Peer PUE as reported in the latest available full-year company reports as of October 2024, Nebius' figure represents PUE of Finnish data center under max. loads in 2020-2021; Due to corporate restructuring in 2022-2024, the data center has not operated at full capacity since 2022, so the measurement pertains to the period before that year; (2) Uptime Institute Global Data Center Survey Result 2023; (3) Green500 list, June 2024; (4) TOP500 list, June 2024; (5) As of October 2024 compared to November 2023 initial measurement; Company data; (6) Tested at temperatures of 25°C, 30°C, 35°C and 40°C; (7) Gigabyte GPU Server G593-SD0.

# 1. Data centers: pioneers in optimized cooling and heat recovery

## Optimized cooling

Wholly owned Finland data center employs **free cooling**<sup>1</sup> instead of traditional chillers, eliminating the need for water and refrigerants, which enhances **cost-effectiveness and environmental friendliness**

### Operational t° max

**~40 °C**

- ~10 °C higher than the upper limit typically set by off-the-shelf hardware
- Eliminates the need of inlet air subcooling
- Allows slower (less-energy intensive) airflows compared to similar DC designs

### Fully functional under 100% workload at

**18–40 °C**

while typical data centers worldwide would aim not to exceed the limit of 27 °C due to server architecture constraints<sup>2</sup>

### Additional energy saved

as data center's operating environment does not require subcooling



## Heat recovery

One of the first in the region to introduce **heat recovery tech** to warm local homes

**80k+ MWh**

of server heat was reused for municipal heating in 2020–2023. This is equivalent to energy consumed by **~2,500 Finnish households**<sup>3</sup> for heating their homes within the four-year period

**50%+**

of the annual heating needs of the town were historically<sup>4</sup> covered by the server heat from our data center

**Up to 12%**

of household annual heating costs (~30% of the data center electricity costs) were saved in recent years<sup>5</sup>

**Nebius aims to replicate these best practices when expanding its data center capacity**

(1) Leveraging external ambient air to reduce air temperature in the data rooms; (2) As per ASHRAE standards for inlet air temperatures; (3) Assuming that a household consists of three persons living in a standalone house; according to the data provided by a Finland-based energy producer, such a household may consume up to 8 MWh of electricity per year for heating; (4) Based on the latest pre-pandemic measurement. Annual heat sales to the municipality in 2020-2023 remained at and above the pre-pandemic levels; (5) Based on the 2019 (pre-pandemic) measurement. Figures are deemed representative for the following years, in line with the trends in heating sales and utility prices.

## 2. Hardware: why do we design, develop and produce our own servers?

Lower total cost of ownership via power efficient cooling and full control over the supply chain

Faster time-to-market with on demand development, quick prototyping and production, deployment and shipment

In-house engineering for full stack validation and quick feature implementation

Open-source infrastructure ensuring security, transparency & flexibility



## 2. Hardware: in-house expertise in R&D ...

### Key differentiating factors:

#### Tool-less design

allows for easy installation and replacement of broken components

#### Lower cost of servers

due to direct purchases of components and optimized design

#### Direct relationship with key suppliers

allows for easy installation and replacement of broken components

**20%**

More capacity within the same data center limits<sup>1</sup>



**Up to ~20%**

Lower electricity consumption decreasing power and data center costs<sup>1</sup>



### ... and a team covering the whole development chain



System architects



Electronic engineers



Embedded developers



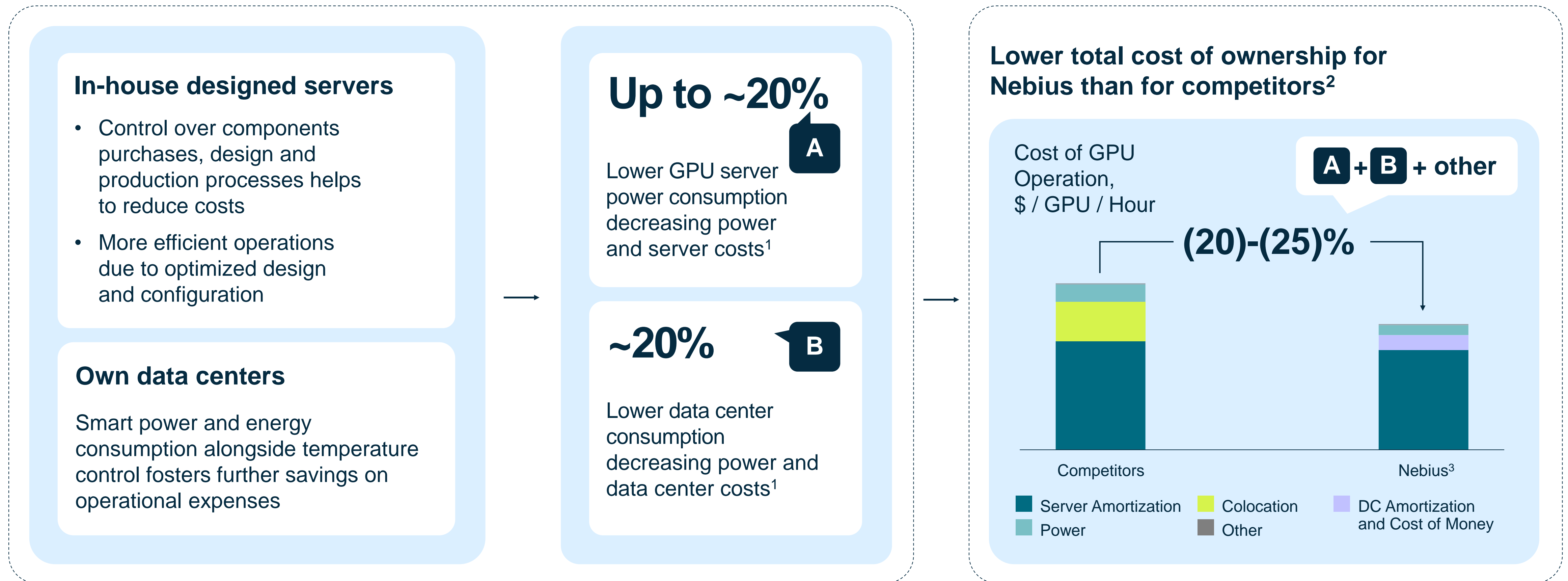
Mechanical engineers



Thermal engineers

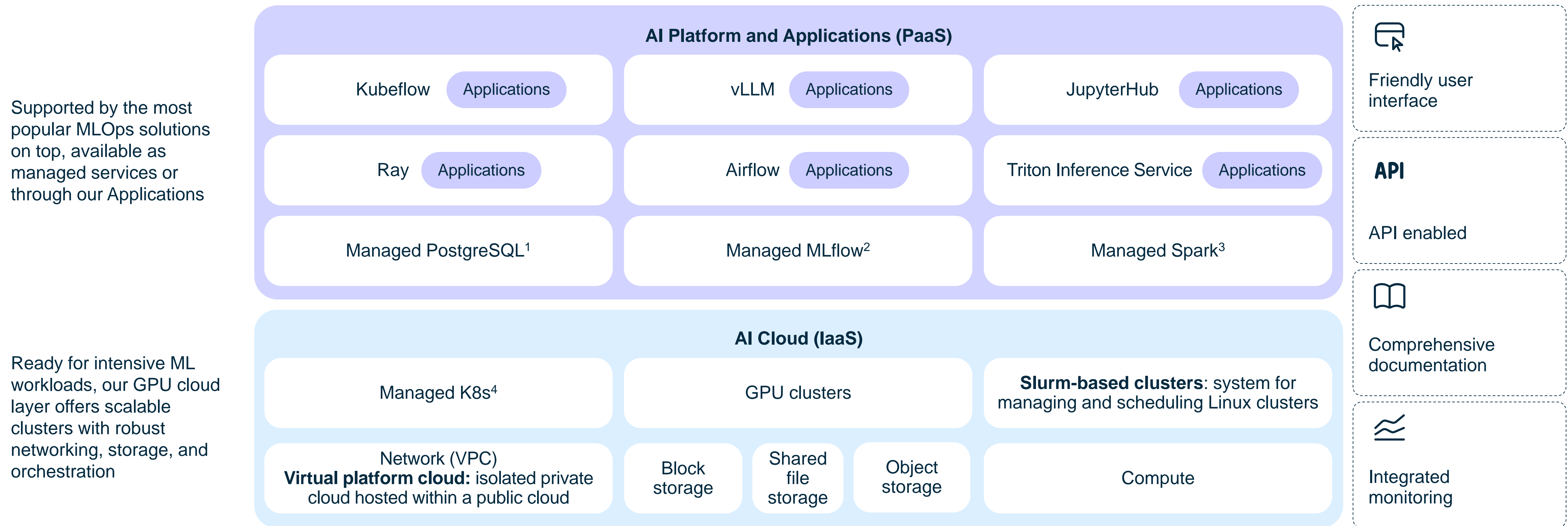
(1) Compares electricity consumption of Nebius' custom-designed servers to off-the-shelf alternatives, showing ~20% lower usage. This allows Nebius to deploy about 20% more servers per 1MW of data center power capacity.

# Bringing data center ownership together with in-house hardware development drives additional shareholder value through cost efficiencies



(1) Compares electricity consumption of Nebius' custom-designed servers to off-the-shelf alternatives, showing ~20% lower usage. This allows Nebius to deploy about 20% more servers per 1MW of data center power capacity; (2) Based on the internal calculation that take into account difference in PUE of Nebius data center in Finland vs publicly available market average as of 2023 and difference in electricity consumption of the Nebius in-house designed server vs. available off-the-shelf alternatives, among the key factors; (3) Nebius own data center.

### 3. Nebius Platform: more than 'bare metal', more flexible and wider product portfolio than GPU Cloud



(1) Open-source object-relational database system; (2) Open-source platform for managing the end-to-end machine learning lifecycle; (3) Open-source data tools for batch processing, querying, streaming and machine learning; (4) Kubernetes: portable, extensible, open-source platform for managing containerised workloads and services.

### 3. Nebius Studio: API SaaS for opensource models

**Nebius Studio** is a new product from Nebius designed to help foundation model users and app builders simplify the process of creating applications using these models with our first release, Inference Service, providing endpoints for the most popular large language models

**Benchmark-backed performance and cost efficiency**

- Up to 4.5x**  
faster time to first token in Europe than competitors<sup>1</sup>
- \$2.5**  
per million tokens for Meta-Llama-405B<sup>2</sup>

**Key benefits**

- 3x** saving on input tokens<sup>3</sup>
- Ability to choose **speed or economy**
- Achievement of **ultra-low latency**
- No MLOps experience** required
- Verified **model quality**
- Access to **expert support**

**Services**

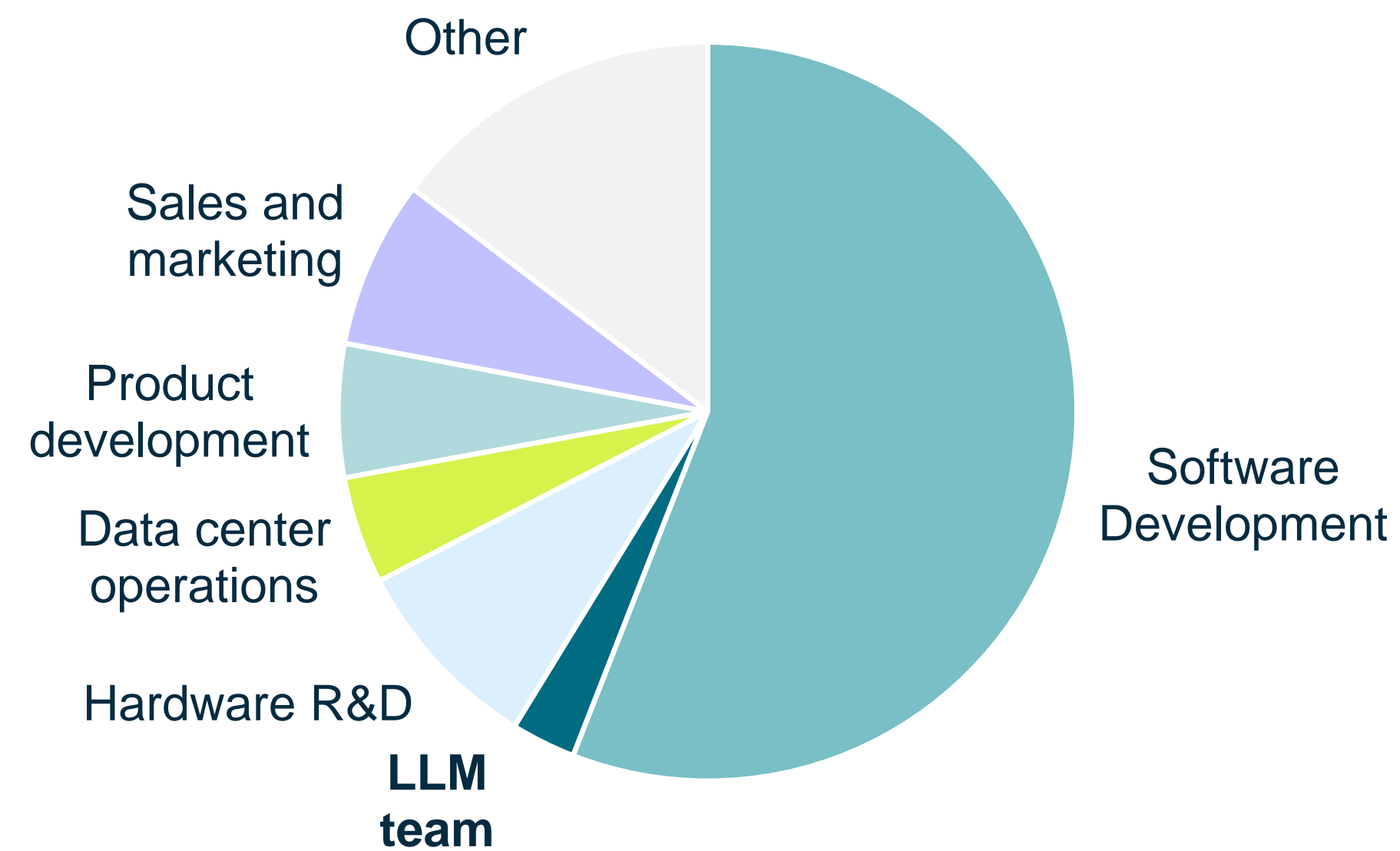
- Real-time inference
- Batch inference **Coming soon**
- Fine-tuning **Coming soon**
- Evaluation **Coming soon**
- RAG **Coming soon**

(1) Measured by Nebius in August 2024 on a group of selected competitors. (2) Company data; (3) The company's price on input tokens is 3x lower than the company's price on output tokens.



## 4–5. The most in-demand engineers are already on board

Team of 500+ professionals to focus on software, hardware and data center with the back-up of support departments<sup>1</sup>



### Key strengths

A team of software and hardware engineers with full-stack expertise in AI cloud development and operations, covering everything from server design to running cutting-edge large language models (LLMs)

In-house LLM team facilitate adjusting Nebius hardware and platform to the real needs of ML/AI practitioners

The right number of people with the necessary expertise in place to support scaling and future growth

(1) Refers to Nebius core business. As of September 30, 2024.

# Nebius competitive advantages

## Nebius key competitive advantages differentiate it from other players

### Benchmark-backed performance and cost efficiency

**Sole focus on AI infrastructure**, allowing us to provide solutions precisely tailored to meet the specific needs of AI developers and businesses

**Long-standing collaboration with NVIDIA** across hardware and cloud. Nebius to be one of the first providers in Europe to bring the state-of-the-art, energy-efficient NVIDIA Blackwell platform to customers in 2025

### Full stack with expertise across the whole AI infrastructure value chain

**Full control over the whole AI infrastructure value creation chain**, from in-house server design to proprietary cloud platform, ensuring maximum productivity and quick scaling, higher cost efficiency and low downtime

**20-25%<sup>1</sup> lower total cost of GPU ownership and operations** compared to an average GPU provider

**Long term relations and collaboration experience with leading server ODMs<sup>2</sup>** for co-designing, manufacturing and optimizing cost structure on servers

**Expertise in running data centers with heavy power loads (100s MW)** in a reliable and cost-efficient way










### Strong team with relevant experience for AI industry, track record, and skillset

**~400 AI / ML / software & cloud engineers<sup>3</sup>** with a track record of building cloud services and infrastructure from scratch

**Team sized to support future growth**

(1) Based on the internal calculation that take into account difference in PUE of Nebius data center in Finland vs publicly available market average as of 2023 and difference in electricity consumption of the Nebius in-house designed server vs. available off-the-shelf alternatives, among the key factors; (2) Original Design Manufacturers (ODMs); (3) Refers to Nebius core business. Nebius Group has ~850 top-tier AI / ML / LLM, cloud engineers and other tech specialists. As of September 30, 2024.

# Nebius proposition is differentiated vs. other players, allowing us to serve the broadest set of customer demands

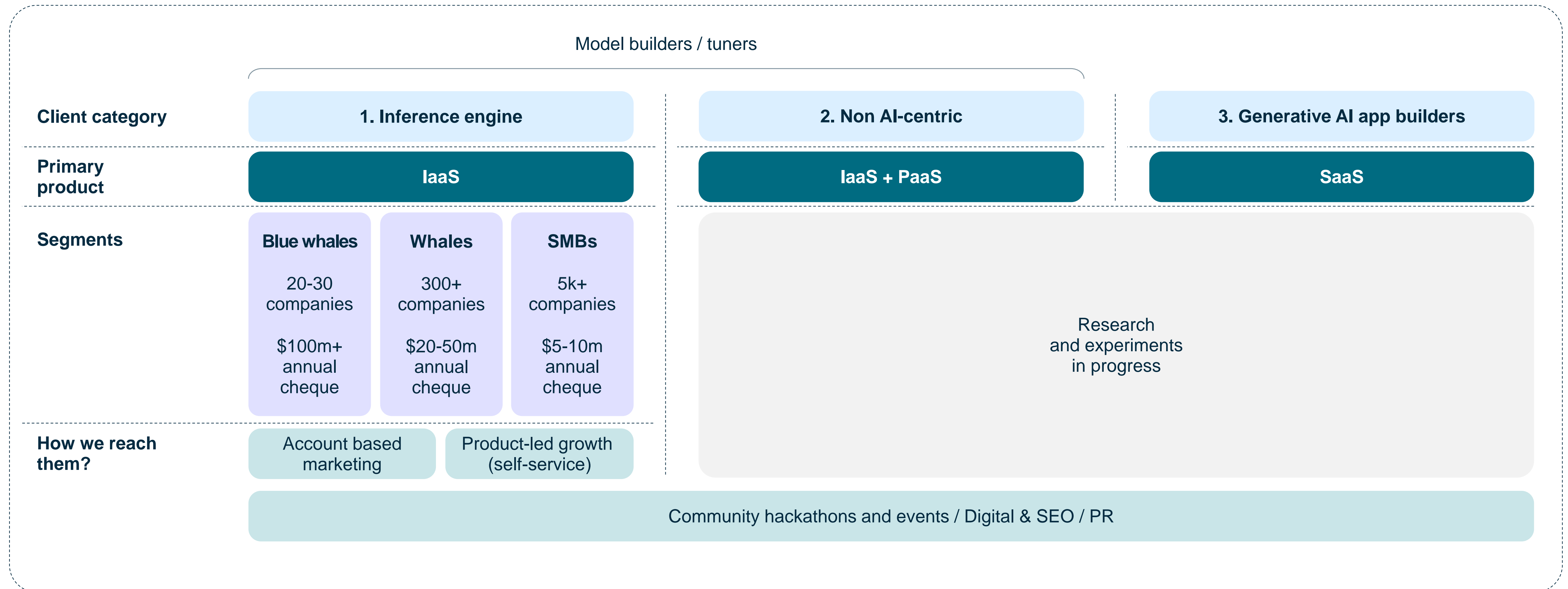
Our customer proposition that differentiates Nebius ...	NEBIUS	... with key advantages over each group of competitors	NEBIUS
Solutions tailored to meet the needs of AI developers and businesses	✓	<b>Bare metal</b>	<b>Nebius is more than just bare metal:</b> provides additional resources and services for AI workloads
Full control over the AI infrastructure value chain	✓	 Taiga Cloud  Voltage park	<b>Nebius offers more flexibility than GPU clouds</b> in regard to GPU consumption and has a broader service offering fuelled by a larger development team
Leading team of AI / ML & cloud engineers	✓	<b>GPU Clouds</b>	
GPU ownership and operations cost advantage	✓	 CoreWeave  Paperspace <sup>2</sup>	
Longstanding partnerships with NVIDIA & collaboration with leading server ODMs <sup>1</sup>	✓	 Lambda	<b>Nebius provides tailored approach</b> for AI developers and businesses (including small and medium size customers)
Experienced team of cloud solution architects to onboard and support clients	✓	<b>Hyperscalers</b>	
Expertise in running data centers with heavy power loads in a reliable and cost-efficient way	✓	 Oracle Cloud  Microsoft Azure  Amazon Cloud  Google Cloud	

(1) Original Design Manufacturers (ODMs); (2) Part of DigitalOcean.

Nebius

# Our clients

# Nebius is able to serve a large list of clients including model builders and Generative AI app builders



## Training scenarios

**What we offer:** Large resilient GPU clusters with Infiniband networking, orchestration tools (fully managed K8S service, Slurm operator based on K8S) and shared storage optimized for check pointing and data streaming

### Model types:

- LLMs
- Multimodal, including:

Text-to-Image

Text-to-speech

Speech-to-text

Other

### Large language models and multi-modal models

- Clusters of thousands GPUs
- Slurm deployment based on Nebius supported solution
- Shared file storage optimized for data streaming and check pointing (reading speed of 100GB per second)



### AI-powered drug discovery

- Training AI models for drug discovery
- Managed K8S cluster with advanced auto-scaling reduces costs by automatically releasing idle GPUs when not in use, maximizing resource efficiency



### LLM-powered copilot model

- Cluster of hundreds of H100 GPUs with Infiniband
- K8S cluster nodes provided by Nebius and managed from an external cloud
- Using Nebius API and Terraform support



## Inference scenarios

**What we offer:** Large resilient GPU clusters with Infiniband network, orchestration tools with fully managed K8S service with auto-scaling and load balancing, storage services optimized for inference (managed and object storage)

### Model types:

- LLMs
- Multimodal, including:

Text-to-video

Image-to-video

Text-to-Image

Other

### Generative AI design tool



- Nebius is the main supplier for the inference for their text-to-image 20B model
- Using managed K8S as core of their inference engine
- Storing generated images in Cloudflare R2 via direct 300 Gbit peering (with peak load 200Gbps) between Nebius and Cloudflare

### Multi-modal models for video generation



- Clusters of hundreds of L40s GPUs optimized for inference
- Several isolated cloud environments for production and pre-prod workloads
- Managed K8S service with GPU auto-scaling to handle consumption spikes



## What our customers say: testimonials and recognition



Recraft

“In Recraft we are serving image generation models to more than 1 million users. We have our own model trained from scratch and run training and data processing all the time. Recraft has been using Nebius for those purposes for about a year.

**Nebius stands out when compared to other clouds**, especially if you're looking for **flexibility** and quick support. With Nebius, you get your own managed Kubernetes cluster, which means you have full admin control to deploy whatever you need without restrictions. Plus, their **support is top-notch**—super fast, and you get to talk directly with cloud architects, which is a huge plus. On the scalability front, they **always have GPUs available**, so no frustrating delays waiting for quotas. The **storage speed** on Nebius is also **higher** in comparison to some other clouds.

Overall, **Nebius offers more control, better support, and reliable scalability compared to some other clouds.**

Apart from that, we are the first company to test experimental setups in Nebius. Most problems in those have been fixed in days. This shows the pace of development of new features and infra improvements in the company.”

Anna Veronika Dorogush, Founder and CEO at Recraft



Genesis  
Therapeutics

“We are grateful to be working with Nebius for our GPU infrastructure needs. Nebius' team is responsive, they clearly communicate capacity / availability, and technical set up has been straightforward. We consider them **a strategic partner who can accelerate our state-of-the-art AI research** at Genesis Therapeutics.”

Carl Tilbury, BD & Strategy at Genesis Therapeutics



MISTRAL  
AI\_

“Nebius compute and network **infrastructure** proved to be **in the top tiers** of what we tried in terms of **stability** and allowed us to train our models **efficiently.**”

Timothee Lacroix, CTO of Mistral.ai



Higgsfield

“Nebius is **one of the best cloud platforms**, providing excellent customer support and **flexible offerings**. Most other cloud providers don't provide a flexible offering of H100, which becomes a hurdle for early-stage startups. For Higgsfield, we faced the same issue, and we are grateful for Nebius for their on-demand H100 offering <...>. The consumer-targeted model was built with Nebius, and now hundreds of thousands of users are generating personalized videos in Diffuse app, which is pioneering AI-powered social media content creation on mobile.”

Alex Mashrabov, Founder of Higgsfield.ai

## User-centric approach

- Web console, CLI, API, Terraform
- Detailed documentation
- Built-in monitoring
- 24/7 technical support

## State-of-the-art hardware

GPU clusters up to 4,000 H100 SXM5, interconnected via 3.2 Tbit/s Infiniband

## Resilient infrastructure

All core infrastructure layers of the platform are designed by the Nebius team with a focus on speed and resilience

- Hardware & network
- Hypervisor
- Compute engine
- Storage
- Managed K8s and Slurm

**99.5% SLA for VM<sup>1</sup>**

## Flexible capacity allocation and pricing

On-demand and multiple term reserve consumption modes available

## Fast onboarding

Optimized software stack for fast onboarding

Dedicated team of engineers will help to set up infrastructure and implement training, inference, RAG and other solutions

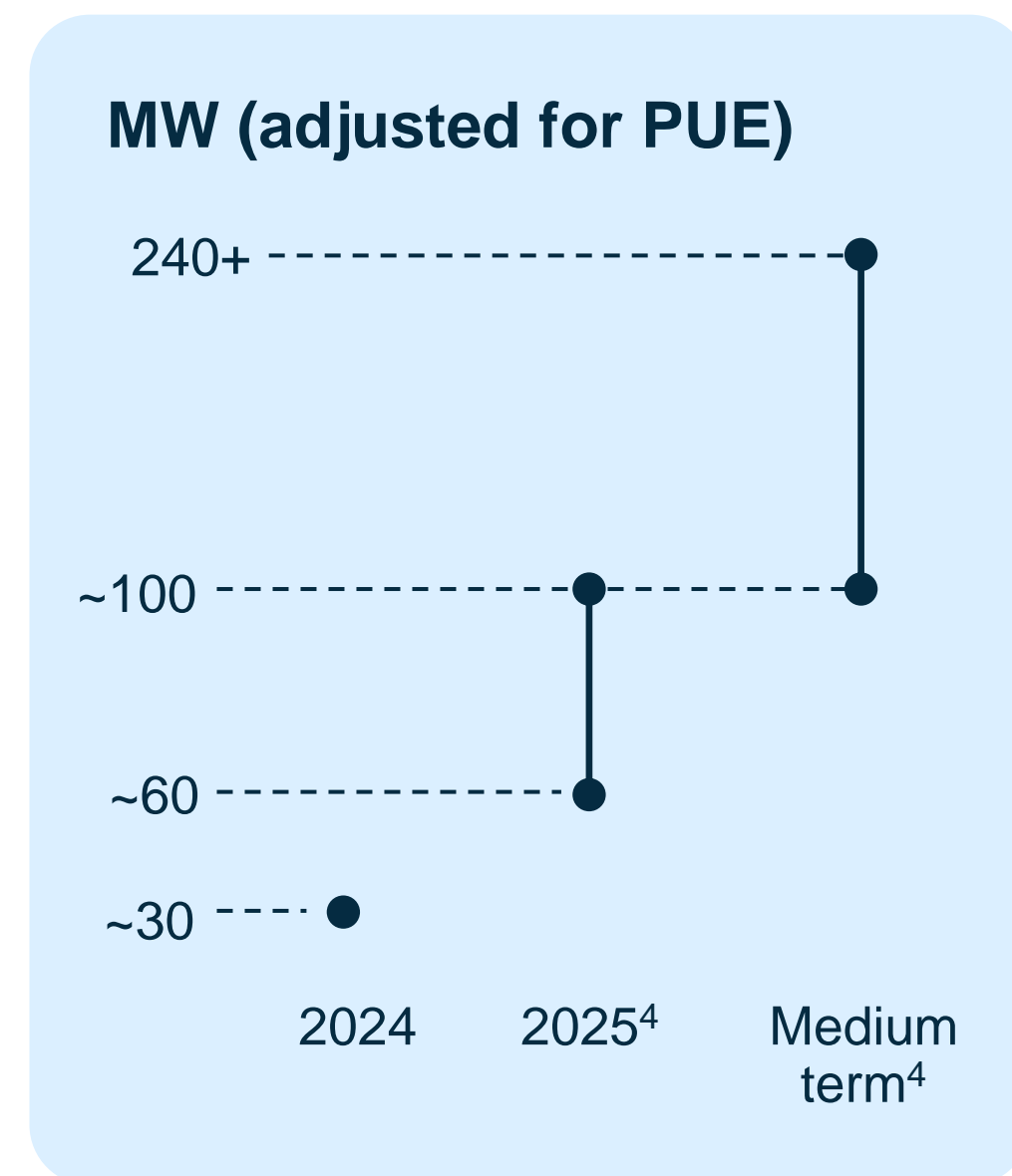
(1) Nebius guarantees 99.5% connectivity for a single VM (Virtual Machine) as per SLA (Service Level Agreement) with clients.

# Growth plans

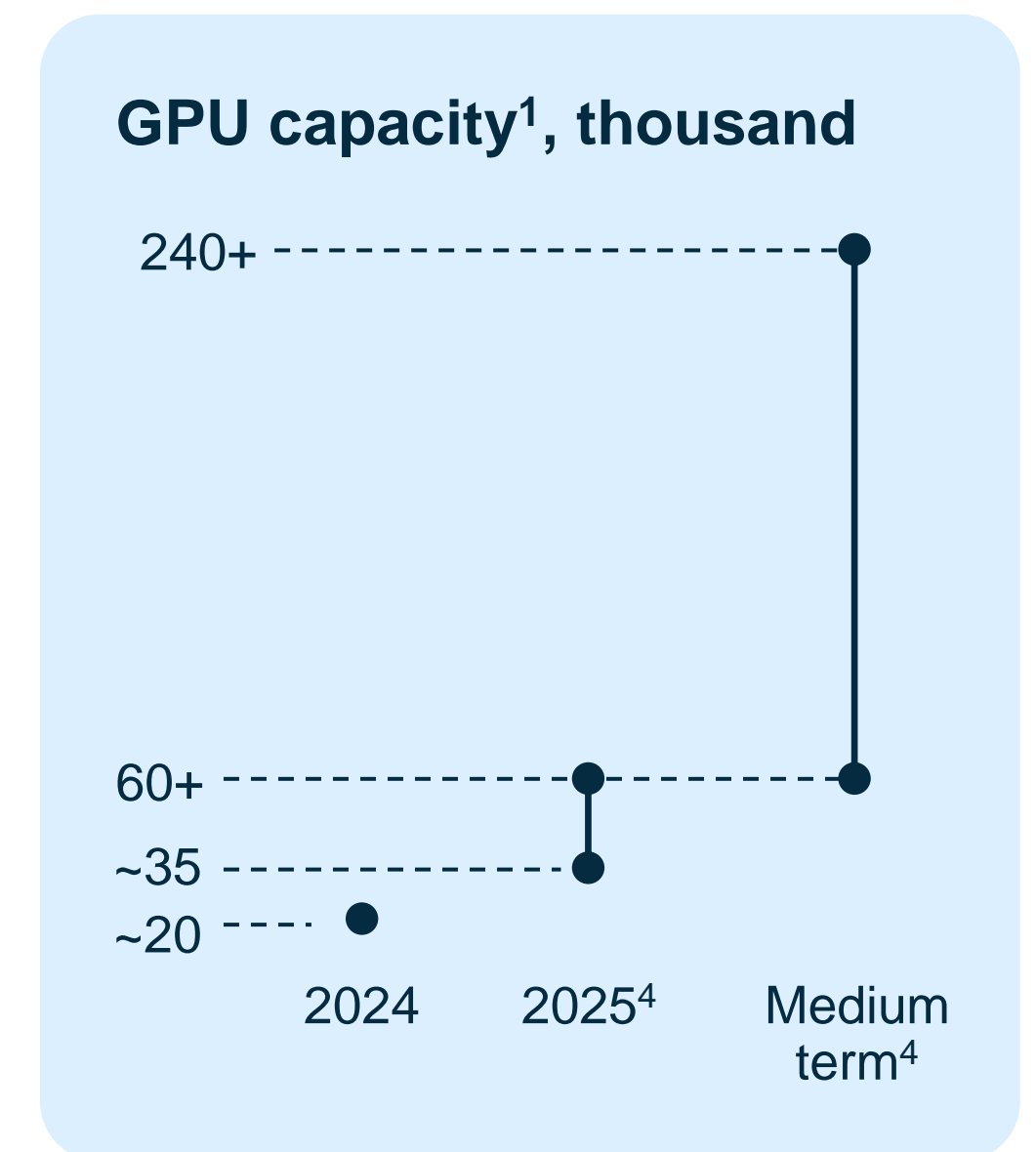
# Nebius has ambitious medium-term plans on capacity expansion of both data center and GPU capacity

## Scaling is backed by robust infrastructure

- **Finnish data center expansion:** expected to triple power capacity to 75MW by the end of 2025 or early 2026<sup>2</sup>
- **Co-location data centers:** signed first agreement in July 2024 with a facility in Paris, France<sup>3</sup>. More locations to be added (including in US)
- **Greenfield data centers**



## Expansion of GPU capacity



(1) In Nvidia H100 equivalent; (2) Finnish data center expansion Phase I commenced in Q2'24; (3) To be put into operation by the end of November 2024; (4) Depending on the amount of capital available to cover the company's investment program during the respective periods.

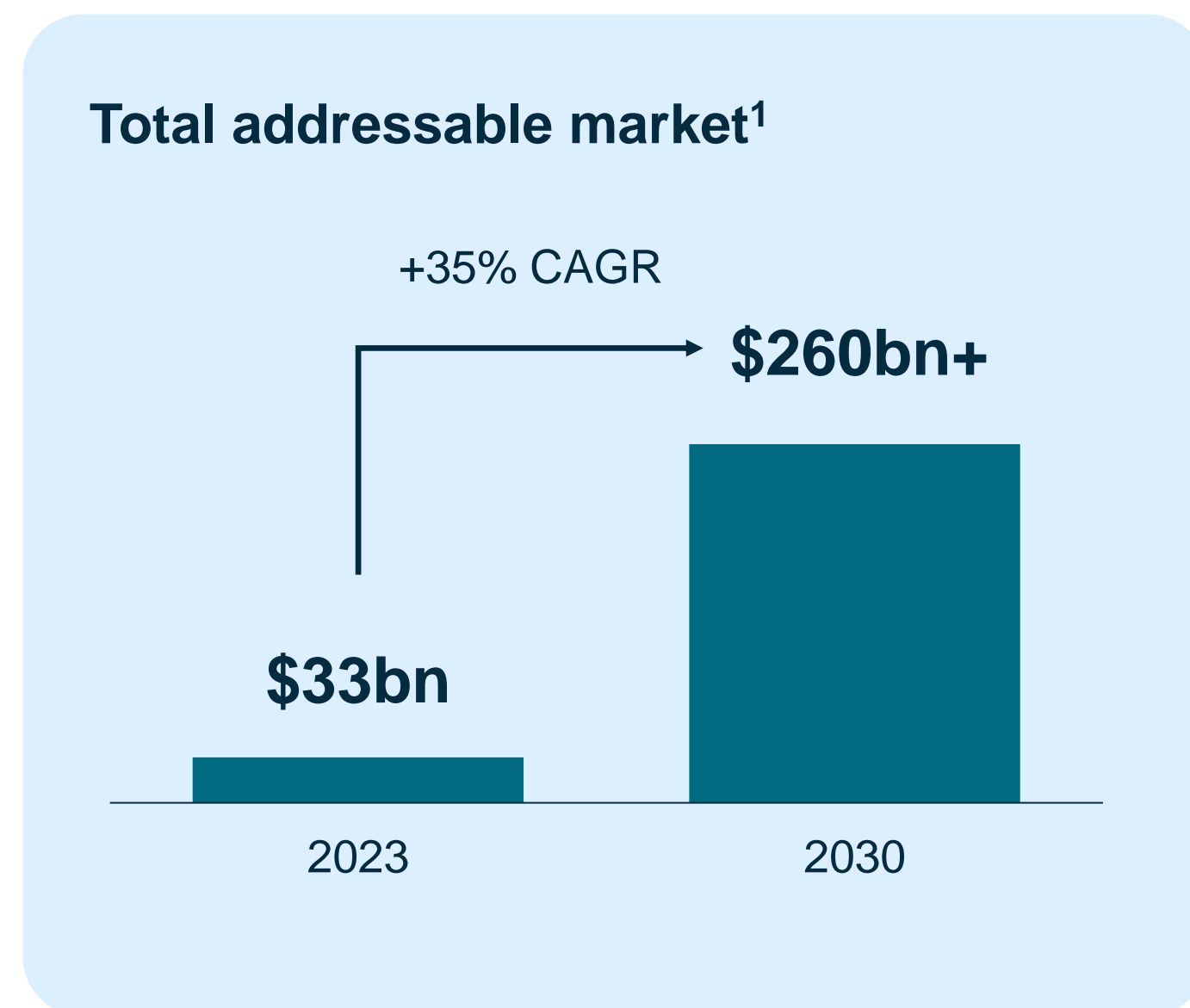
# This is just the beginning: huge potential for growth across multiple levers ...

## External

### Tremendous market growth

Leverage **GPUaaS** and **AI Cloud** markets growth – expected to be 8x over the next 7 years

Nebius to continue expanding **market share** in a growing market



## Internal

### Client base

- Secure **long-term contracts** with existing customers, acquire additional GenAI labs and **expand into new customer segments**
- Continue supporting existing customers and providing **additional volume** as they evolve, starting small and **continuously scaling**; complemented by a unique team of cloud solution architects to onboard and support new clients

### Value-add services

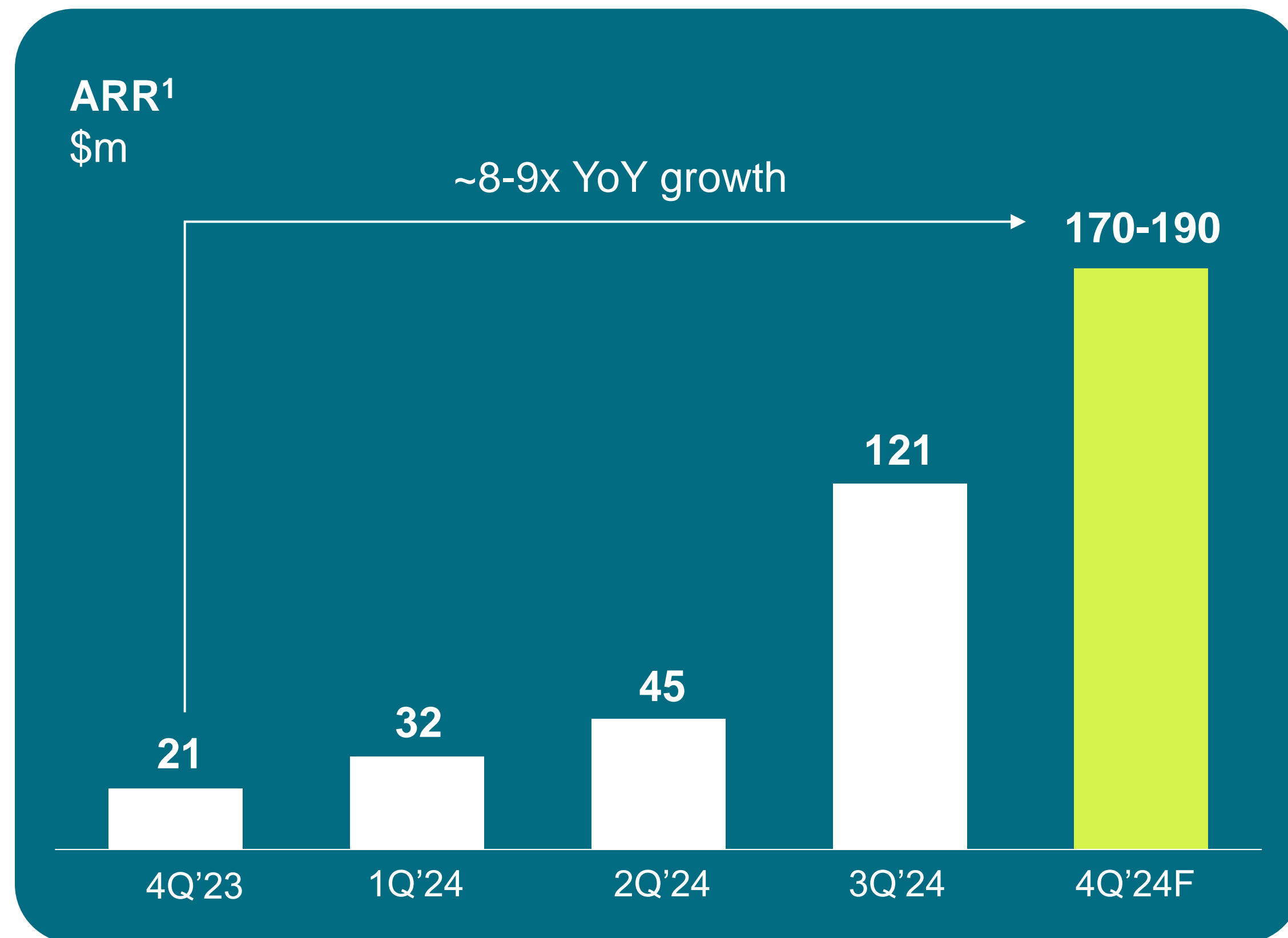
Continue **expanding product range** offered to customers leveraging the existing expertise, data and infrastructure (i.e., Launching AI Studio: API SaaS for open-source models with per token pricing)

... enabled through significant data center and GPU capacity expansion

(1) Nebius internal estimates based on a bottom-up analysis of publicly available disclosure on GPUs spending and GPU Cloud economic; the presented TAM includes estimates for two key markets: GPU-as-a-service and AI Cloud.

# Financials

## Historical and current trading



### Key drivers

**Expansion of our client base** from 10 clients at the end of 2023 to 40+ managed clients as of now and we expect further growth by the year-end

Driven by **capacity expansion** from ~2k in 4Q'23 to 20k+ GPUs expected by the end of 2024

**Growing consumption** by existing clients

(1) Annualized run-rate revenue by the end of the period (revenue for last month of the period multiplied by twelve); ARR for 3Q'24 is based on preliminary revenue for September 2024 as per management accounts.

## Nebius expects the revenue to increase at least 5x and become Adj EBITDA positive in 2025

### Current trading and 2024 guidance

	Q3'2024 <sup>3</sup>	2024
<b>ARR<sup>1</sup></b>	\$121m	\$170 – 190m
<b>Revenue</b>	\$27m	\$70 – 80m
<b>Adj. EBITDA<sup>2</sup></b> Cash Opex of \$15-17m / month expected to be fully covered by revenue by the end of 2024 or early 2025	(\$7m)	(\$60 – 70m)
<b>Capex</b>	\$164m	\$0.9 – 1.0bn

### 2025 Outlook

Depending on the amount of capital available in 2025 to cover the company's investment program we intend to:

- spend between **\$600m and \$1.5bn on CAPEX** (most of it will be invested in NVIDIA GB200 GPUs, as well as expansion of our DC capacity)
- generate **\$500-1,000m ARR** by the end of 2025 and corresponding \$400-600mn revenue for the full year based on assumed capacity expansion
- become **Adj EBITDA positive for the full year 2025**

(1) Annualized run-rate revenue by the end of the period (December revenue multiplied by twelve); ARR for 3Q'24 is based on preliminary revenue for September 2024 as per management accounts; (2) Adjusted EBITDA/(loss) for business segments is segment revenue minus all cash operating expenses, including certain cash Stock Based Compensation (SBC) costs before one-off restructuring and other expenses; (3) Preliminary Q3 2024 financial results based on the management accounts.



Recap

## Nebius key investment highlights

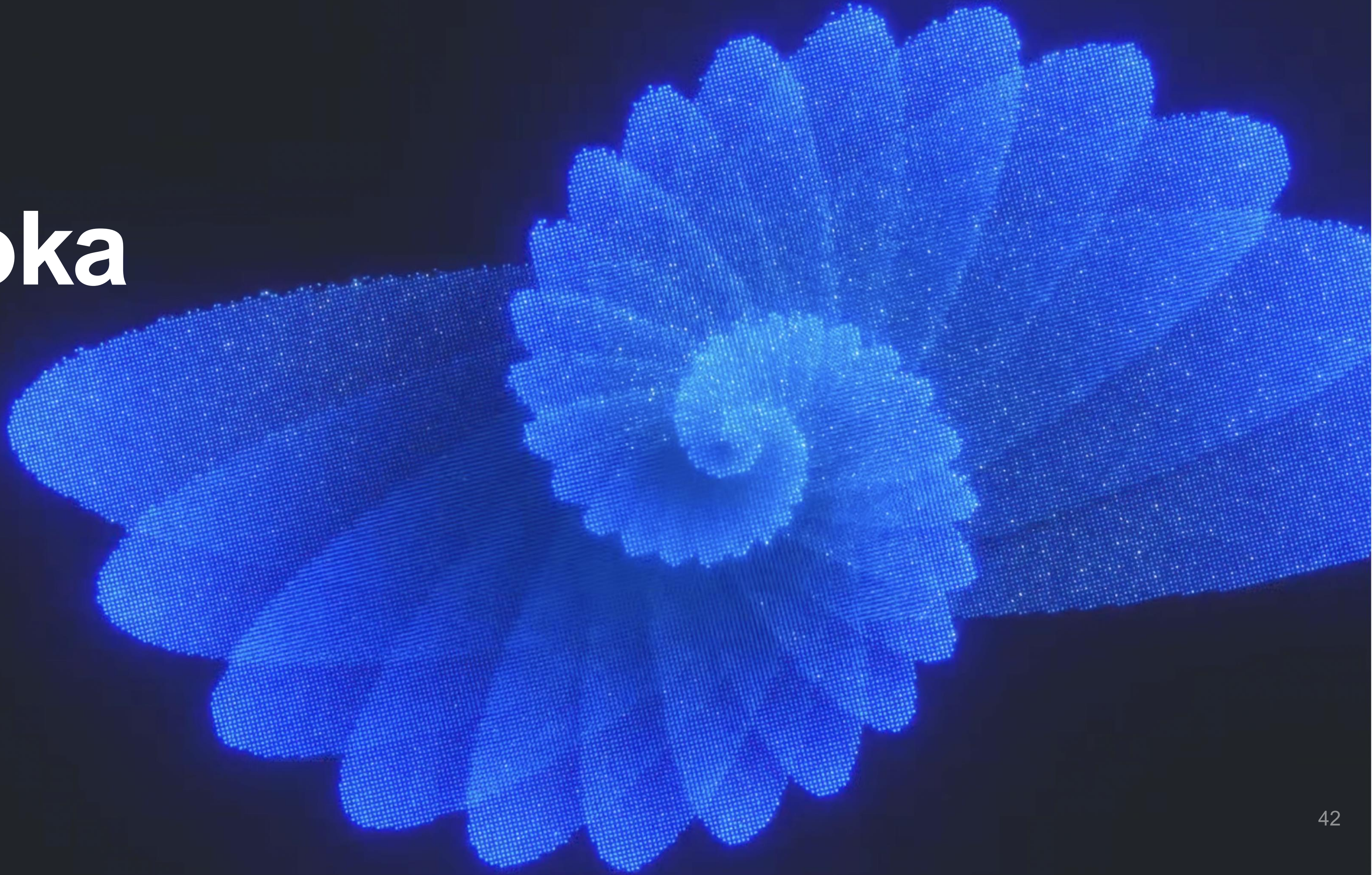
Pure play in attractive AI market in its most in-demand segment: infrastructure

Full stack with expertise across the whole AI infrastructure value chain; from data centers and in-house server design to AI cloud platform and expertise to run LLMs

Strong team with relevant experience for AI industry, track record, and skillset, prepared to support future growth

[Visit website](#)

# Toloka



**A data partner for all  
stages of AI development  
from training to  
evaluation**

## Toloka AI offers data-for-generative-AI solutions by leveraging both AI-powered labeling and human expert input

**Pivoted to a new business model** in 2023, focused on providing **high-quality data for GenAI**, now the primary source of revenue

Leveraged past experience in data labeling and crowdsourcing to build **proprietary tech platform** focused on highly-skilled contributors and providing **GenAI expert data at scale**

**Fast scaling platform with significant potential for further expansion in the GenAI market:**

**8,000+ experts<sup>1</sup>**

~50% of experts hold Master's degrees or above

**25+ knowledge areas**

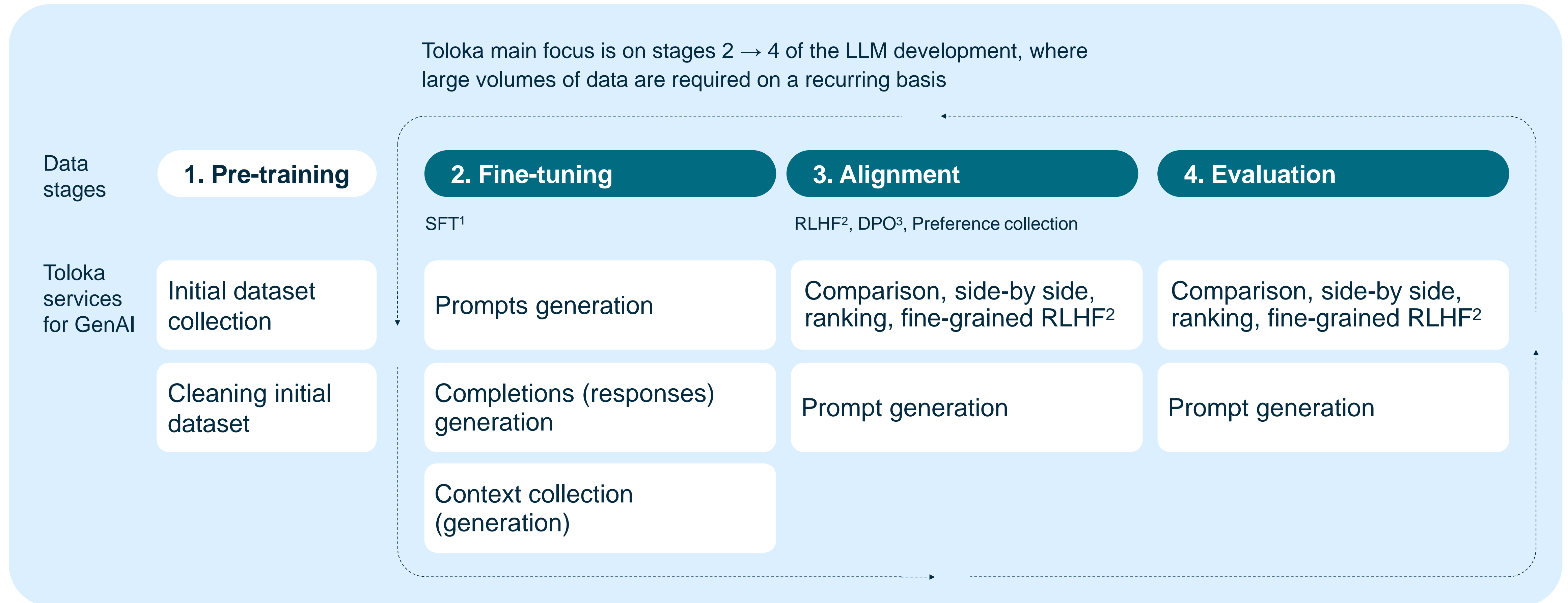
60+ coding languages, and 40+ languages supported

**Up to 250k+ MAU<sup>2</sup>**

Sourced via crowdsourcing platform

(1) Cumulative amount of onboarded experts (domain experts and expert annotators) as of end of Q3 2024. Experts comprise highly skilled contributors from various industries; (2) Monthly active users, for the period between January 1 and September 30, 2024.

# Toloka powers AI production with human intelligence



(1) Supervised Fine-Tuning; (2) Reinforcement Learning from Human Feedback; (3) Direct Preference Optimisation.

# Growth of Gen AI raises the demand for large volumes of data

## Gen AI Data User Profiles

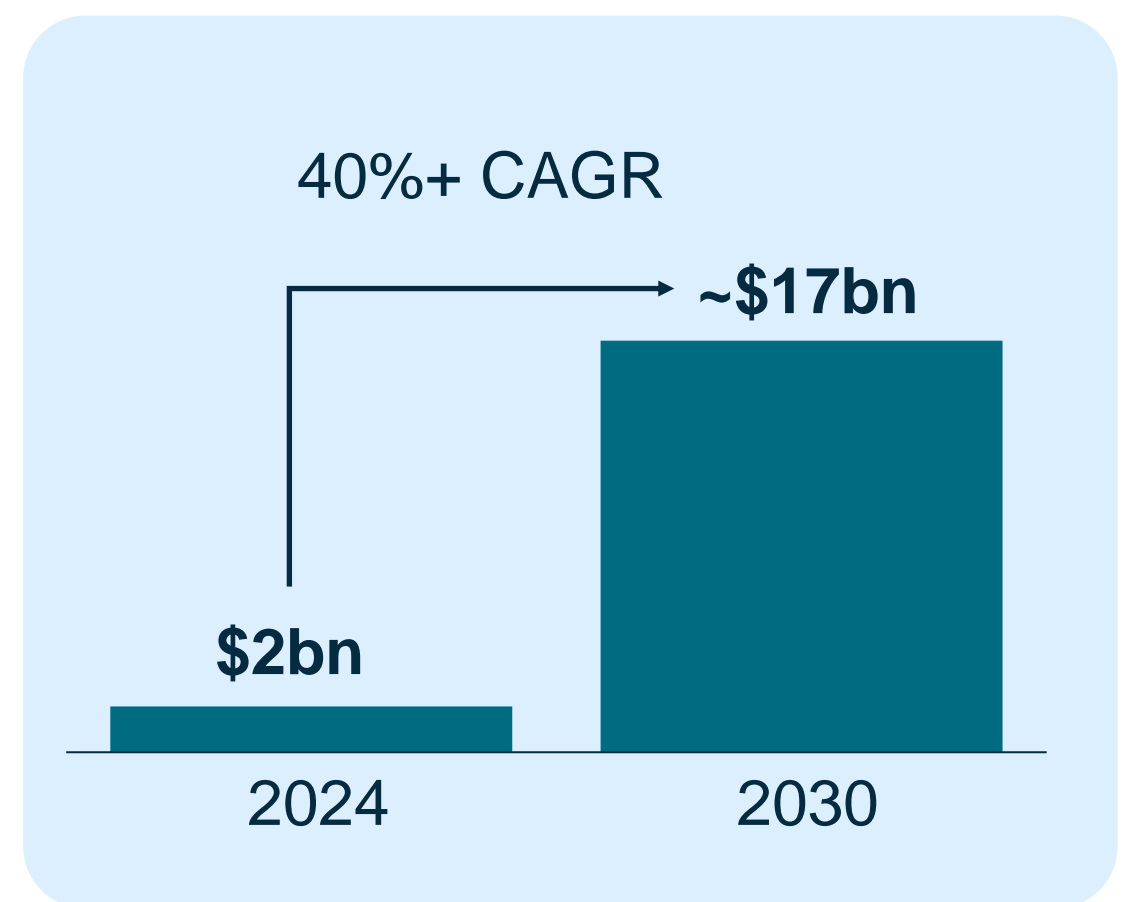
Cluster	Specialization	Spending for data	SFT	RLHF	Eval.
1	Top foundational models producers Examples: Anthropic, OpenAI, Gemini	~20 in the world Spending 10s-100s MM	✓	✓	✓
2	Smaller-scale, specialised, national models producers Examples: Mosaic, Aleph Alpha	~100 in the world Spending 1s-10s MM	✓	✓	✓
3	Heavy finetuners Examples: Perplexity, Rabbit, Character.ai	~100s in the world Spending 1s MM		✓	✓
4	Enterprise end users of GenAI	~1,000s in the world Lower spend			✓

Companies tend to diversify sourcing from multiple providers

↑ Increased spend per company  
↓ Increased # of companies

## Expert data for generative AI market size<sup>1</sup>

Growth of AI raises the demand for large volumes of accurate, diverse, legally compliant, human-annotated data providing a vast opportunity for TAM expansion



(1) Toloka's internal estimates based on our analysis on the client universe and assumptions on average spend by different categories on the back of the company's data and industry checks.

# Toloka AI has both the expertise and technological infrastructure to build high quality human data operations at scale

## Deep & diversified expertise at different levels

### Domain experts

Industry professionals (coding, finance, law, STEM, ...)

### AI Tutors

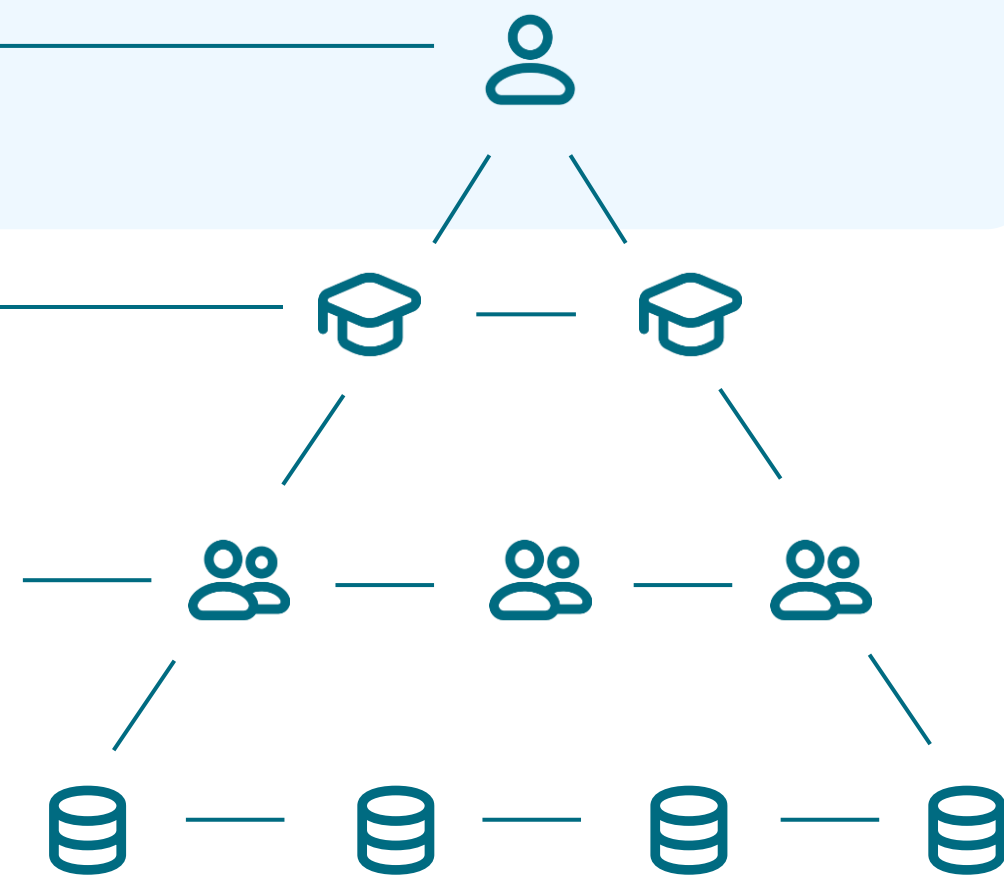
Experienced writers and fact checkers who craft texts

### Global Crowdsourcing

Annotators for data labeling, collection, and feedback

### Synthetic Data

LLM-as-a-judge, synthetic datasets generation



## Coupled with project decomposition into ~40 modular elements to ensure quality & optimized cost

### Human & synthetic data

for GenAI model training and inference, both requiring large amounts of high-quality data

### Evaluation models

capturing the right metrics to determine whether an LLM is accomplishing its goals

## Backed by unique efficiency tools

Automated quality control & antifraud

Marketplace mechanics to find best suited experts

Expert Copilot for automating expert routines, boosting time efficiency<sup>1</sup>

Mindrift, a portal for cost-efficient expert onboarding



**This makes Toloka one of only a few companies globally able to generate high-quality data at scale**

(1) Overall effect of fact-check, intent-check, and knowledge search automation, based on a user study involving leading experts.

## Toloka's competitive highlights

	Quality	Breadth	Scalability	Customer experience / satisfaction	Experts experience / satisfaction
How	Technological tooling, methodology, and expertise to control quality of human efforts at scale	Access to global talent pool of a high variety of domains and languages using proprietary approaches developed with crowdsourcing expertise	Ability to rapidly scale within any domain / language / modality combination using proprietary technology and acquisition + talent verification methodology	A true Data Partner approach providing expertise and acting as an extension of clients' Data Science teams	Intuitive onboarding process, user-friendly product, community-driven approach
Measure	<p><b>100%</b> renewal rate of multi-year contracts with BigTech customers since 2020</p> <p>Proven by <b>continued retention</b> inside Clusters 1-2 accounts</p> <p><b>Win rate of up to 75%</b> of models tuned with Toloka data vs leading open-source dataset<sup>1</sup></p>	<p><b>25+</b> domains, <b>40+</b> languages</p>	<p>Prior experience <b>scaling the number of expert engagements 10x</b> (0.9k &gt;&gt; <b>9.5k<sup>2</sup></b>) within 2 months for a client project with no quality compromise</p>	<p>Customer <b>NPS<sup>3</sup> of 52</b> supported by a <b>high client retention</b></p>	<p>Experts <b>NPS<sup>3</sup> of 54</b> with a <b>retention per project 75%+</b></p>

(1) Internal Toloka assessment, based on several use cases and customer research experiments; (2) Data refers to February-April 2024; (3) The number of experts engaged varies from 1,000s to 10,000s at different points throughout the year, depending on the evolving needs of ongoing projects; (3) Data as of Q2 2024. Net Promoter Score (NPS) = Percentage of Promoters – Percentage of Detractors, with a range of scores from -100 to 100, where a score above 30 implies a good customer relationship. NPS is measured semi-annually.



# Trusted to provide expert data at scale

## Scaling is backed by robust infrastructure

MS Azure as base infrastructure, with storage options in the US and EU, as well as private and on-premises data storage options

Certified to & compliant with:

### ISO 27701

secure collection, handling, storage and erasure of personal identifiable information

### HIPAA

protection and confidential handling of protected health information (PHI)

### GDPR

privacy of users & data protection

### ISO 27001

information risk management

### SOC2

security, availability, processing integrity, confidentiality, privacy

## Deep roots in the scientific community ...

... help Toloka to quickly acquire the knowledge necessary to address industry needs and land clients as well as expertise in LLM production and associated data requirements

### Gartner

Toloka AI recognized as one of key reinforcement learning from human feedback (RLHF) providers<sup>1</sup>

### Forbes



Featured as one of the key data providers for GenAI<sup>2</sup> and recognized in AI Infrastructure roadmap<sup>3</sup>

Top conferences on AI

### NeurIPS

... and more

### ICML

### AAAI

Joint research with top universities in the US and Europe



... and more



(1) Hype Cycle for Generative AI Report, Gartner, 2023; (2) Forbes, July 2024; (3) Bessemer Venture Partners, June 2024.

# Toloka AI is trusted by Big Tech, Fortune-500 companies, and A-class startups given strong ML experience of the team and research partnerships

## Select customer feedback



“Thanks a lot to the Toloka team. Wonderful collaboration in the past two years.”

Data Scientist, Microsoft Bing



“Our collaboration in both business and research areas has always been extremely fruitful and pleasant. Toloka has established itself as a highly professional and knowledgeable team with a strong focus on responsible AI.”

Principal Responsible AI Architect, ServiceNow



“We were excited with Toloka’s speed – 14 categories of personal sensitive information were annotated from 12,000 code files within just 4 days.”

ML Engineer, Hugging Face

“Communication and iteration towards QA goals are fantastic. Team members are always available to answer questions and bring forward questions of their own that improve the results of the project.”

Data Scientist, BigTech Company

## Select research collaborations



to be published soon

**MIT:** using human effort to improve AI-generated content detection technology



ongoing

**Cohere for AI:** evaluation benchmark for assessing end-to-end real-world code generation task



**Meta / Microsoft / UCL:** large-scale machine translation evaluation for African Languages

## Positive feedback and research experience leading to industry recognition

**Gartner:** Recognized in Hype Cycle for Generative AI 2023

**Bessemer Venture Partners:** Recognized in AI Infrastructure roadmap 2024

**Cohere:** Aligning LLMs to Low-Resource Languages, tutorial at AAAI 2024

**Hugging Face:** Reinforcement Learning from Human Feedback: A Tutorial at ICML 2023

**SIGIR:** Best Prompts for Text-to-Image Models and How to Find Them 2023

Customer case study

## Multiturn Dialog Evaluation: How we scaled the process to evaluate 3,000 dialogs per day

**Goal:** Agent model for IT service management ticketing evaluation and fine-tuning

**Solution:** customized pipelines for evaluating agent and user responses, ensuring criteria such as responsiveness, transparency, accuracy, groundedness, and helpfulness were met

**Result<sup>1</sup>:** increased throughput **250 times** compared with the other evaluation provider. The project scaled from **400 tasks per week to 3,000 tasks per day** that ServiceNow processes currently

 Toloka × **servicenow**

“One of the most remarkable aspects of working with Toloka is the **invaluable support provided by their team during project setup, which we greatly appreciate. Their efficiency and organization in communicating instructions to experts significantly contribute to the overall success of the project.**”

(1) The results are based on the performance comparison between Toloka and ServiceNow's previous provider conducted in December 2023.

# Toloka AI key competitive advantages differentiate it from other players

## Beyond traditional data labeling

- 1** Best-in-class technology driven by **robust scientific cooperation** and R&D capabilities
- 2** Robust **quality controls** and **infrastructure security** meeting the most rigorous standards
- 3** **Unmatched quality data at scale** to meet the growing demand for fine-tuning, alignment, and evaluation in LLM development
- 4** Marketplace mechanics supporting **delivery and efficient scaling: several thousands<sup>1</sup> of domain experts** working across **25+ knowledge areas** in 40+ languages
- 5** **Efficiency tools to ensure quality, scalability and optimized cost** translating into competitive pricing for clients
- 6** Engaging in **scientific research and R&D collaboration** with renowned partners in **ML / AI**

(1) The number of experts engaged varies from 1,000s to 10,000s at different points throughout the year, depending on evolving needs of ongoing projects. 8,000 experts cumulatively onboarded as of the end of Q3 2024.

# This is just the beginning: significant potential for growth across multiple levers

2014-2023

## Crowdsourcing for classic ML

- Democratized access to unqualified global crowd
- Gained expertise in building crowdsourcing solutions
- Opened tools for external requesters

One of the world's largest open crowdsourcing platforms

2024+

## Data partner for GenAI

- Democratized access to high skilled experts
- Gained expertise in building data pipelines for GenAI (RLHF etc)
- Opening pipelines for external requesters

Intelligence under the hood of world's major AI solutions

2025+

## Ultimate marketplace for AI+human actions

- Democratized access to unqualified global crowd
- Gained expertise in building crowdsourcing solutions
- Opened tools for external requesters

“Uber” for cognitive work

2026+

## Data partner for GenAI

- From being Data Provider for major GenAI producers
- To orchestrating efforts of “uberised” cognitive workers whose labor was partly automated by AI
- To developing AI Agents collaborating with humans for the long tail of activities

Powering the spectrum of automation with human insights

# Financial Highlights

# Overview of key revenue drivers

Revenue categories	<b>Classic GenAI Solutions</b> <div data-bbox="483 712 849 825" style="background-color: #00728f; color: white; padding: 5px; border-radius: 10px; text-align: center;">80%+ of Toloka's revenue in 2024<sup>1</sup></div>	<b>Evolved GenAI</b>	<b>Self-serve platform</b>	<b>Non-core non-Gen AI services</b> <div data-bbox="2482 712 2848 825" style="background-color: #e0f2f1; padding: 5px; border-radius: 10px; text-align: center;">&lt;20% of Toloka's revenue in 2024<sup>1</sup></div>
Core growth drive	<p><b>Securing new Cluster 1 and Cluster 2 customers</b> (at least 1-2 new Cluster 1 customers and 3-4 new Cluster 2 customers by the end of 2024)</p> <p><b>Expand wallet share with existing clients</b> by consistently delivering uncompromised quality to meet increasing data needs across more domains, modalities, and specializations</p>	<p>Growing demand for <b>security and evaluation solutions</b> from both Cluster 1 and Cluster 2-3 ML teams</p> <p><b>Re-launch of the Toloka platform</b> will enhance ability to deliver <b>responsible AI solutions</b></p>	<p>Launch of Toloka <b>self-service solution</b> in 2025</p> <p>Appealing solution to <b>lower-cluster companies</b> or those seeking <b>direct access to experts and annotators</b></p>	<p><b>Hybrid solutions</b>, including <b>ML use cases</b> not effectively covered by GenAI or other state-of-the-art models at reasonable costs</p>

(1) Based on the expected full year 2024 revenue, as per September 2024.

Recap

## Toloka AI key investment highlights

Delivering unmatched quality data at scale to satisfy broad needs of LLM development

Large team of domain experts, operating a marketplace model to enable rapid and efficient scaling

Premium technology backed by scientific research collaborations and extensive previous R&D

High-quality client base providing strong positive feedback



[Visit website](#)

# TripleTen



**A leading edtech platform  
specializing in reskilling  
individuals for successful  
careers in tech**

# Addressing the widening skills gap created by rapid digitalization

## Top-rated edtech

5 immersive B2C study tracks:

Software engineering

Quality assurance

BI analytics

Data science

Cybersecurity

## B2B offering

AI tech courses for beginners and professionals already working with AI

## Expertise and results in STEM

**10+ years**

of experience in edtech for Triple Ten management team

**1,000+ people**

enrolling monthly, and the number is growing

**87% employment rate**

within six months of graduation<sup>1</sup>

**FORTUNE**

**Best overall provider<sup>2</sup>**

of software engineering bootcamps



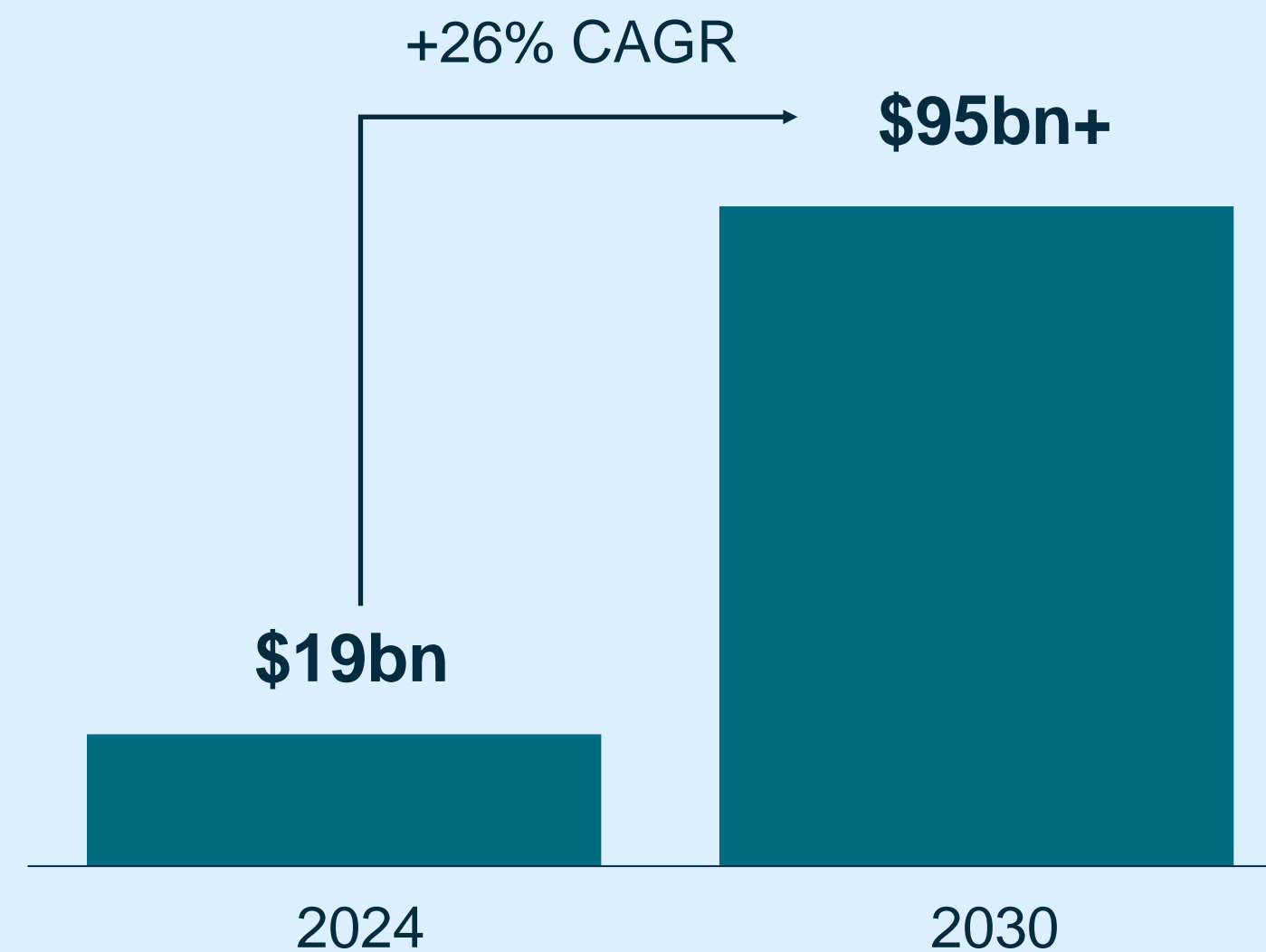
(1) In the US market. See 2023 Outcomes Report; (2) Fortune Magazine's ranking, 2024.

# Capitalizing on growing demand for IT reskilling

**Demand for tech talent is high and growing,**

including across TripleTen's primary markets — **US and LATAM**

Global digital education market<sup>1</sup>



## Growing B2B market

as companies respond to tech disruption by seeking to close IT skills gaps, embrace AI and retain talent by upskilling staff in high-demand areas; allocating as much as

**up to 1.5% of total budget to reskilling<sup>2</sup>**

# TripleTen offers high-quality and immersive training

## Immersive training & expert guidance ...

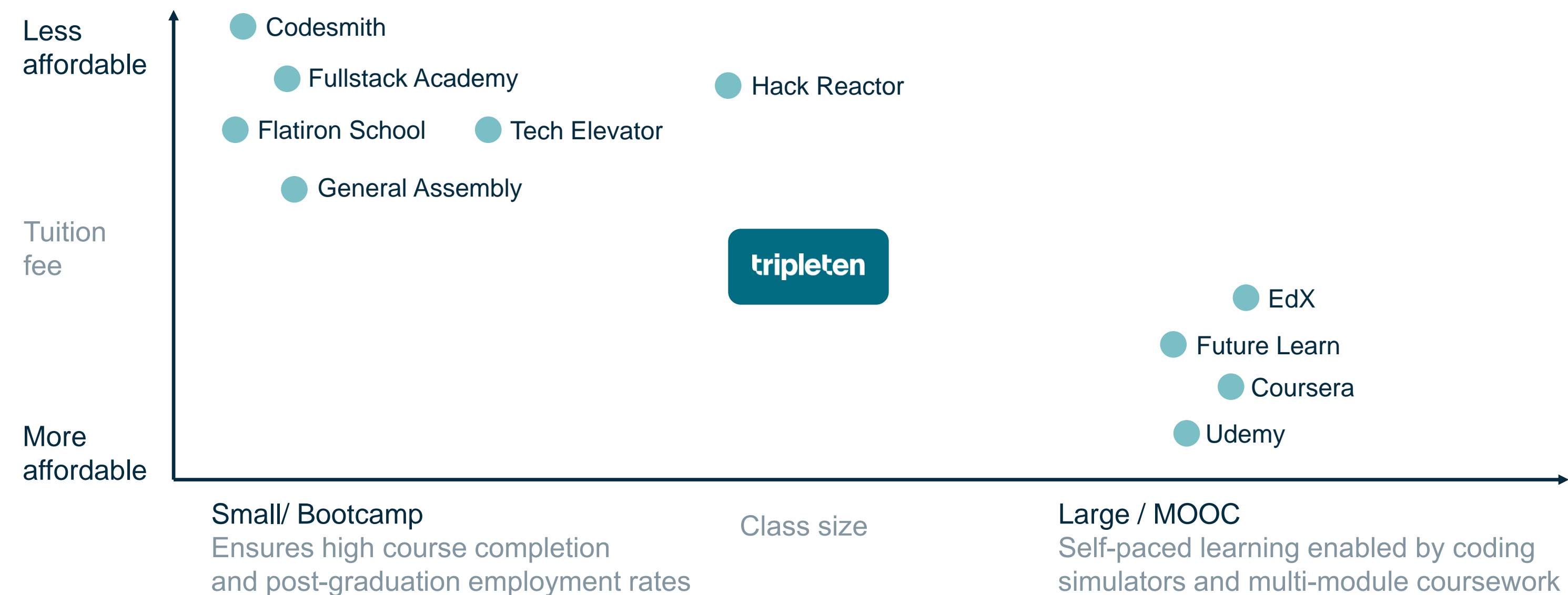
Expert guidance and detailed code reviews from IT practitioners

Work-simulated environment featuring real projects and career services

Flexible & personalized learning paths tailored to a student's knowledge level

... combining the best of bootcamp and MOOC<sup>1</sup> formats ensures personalized learning and easy scalability for broader reach

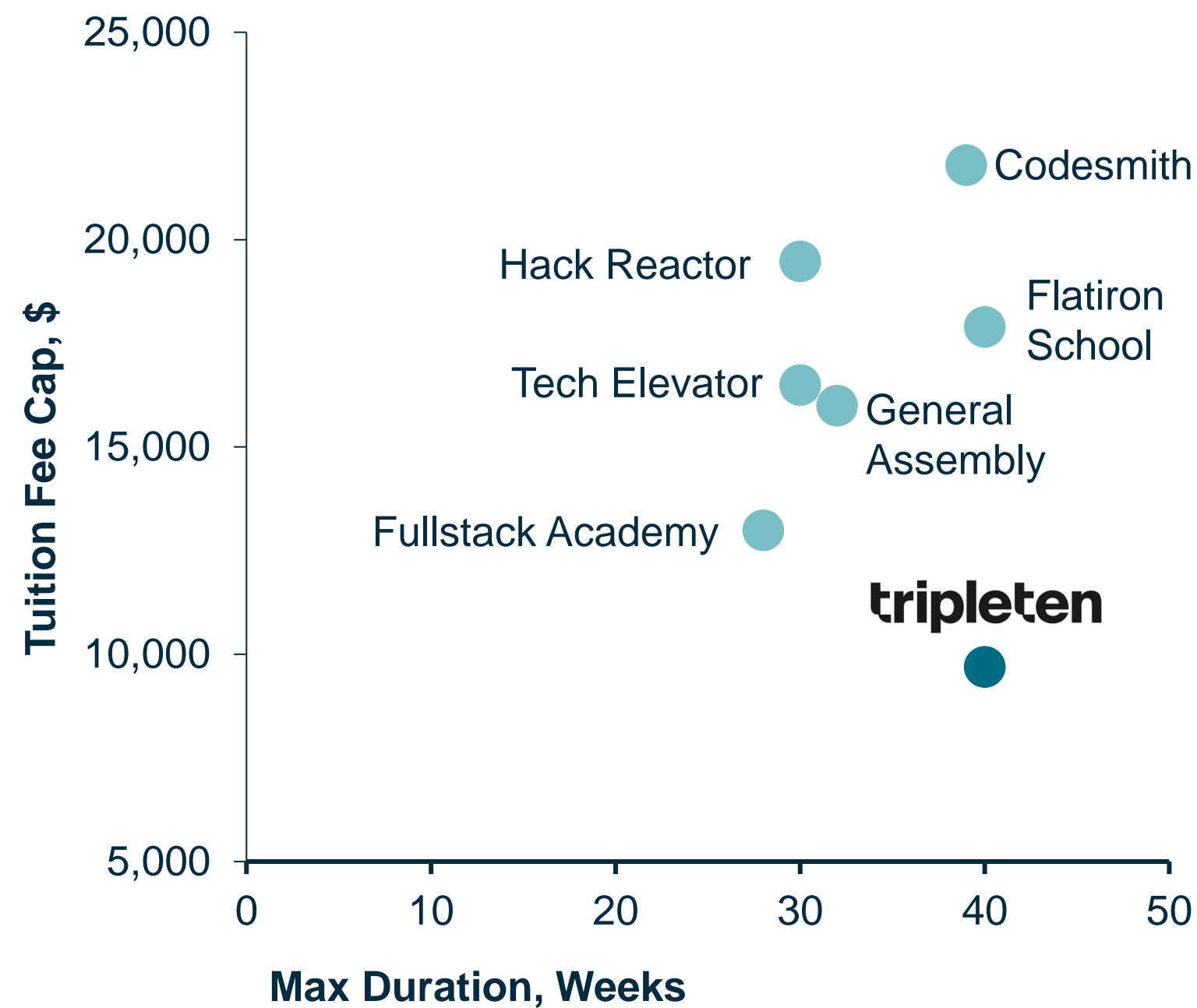
### Market positioning



(1) Massive Open Online Course (MOOC).

# TripleTen offers affordable prices backed by optimized costs and automation

## Affordable offering vs peers<sup>1</sup> ...



## ... enabled by optimized costs of training and processes automation

### Diversified sourcing of trainers and facilitators

Qualified English-speaking talent globally / part time educators for collaborative teaching

### Automation of e-learning workflow

- AI Tutor integrated in all study tracks, facilitating learning progress for students while reducing service expenses
- Automated task dispatch based on student skill level and availability
- Built-in machine translation for code reviewers

(1) Based on publicly available data on tuition fees for the comparable courses, as of October 2024.

# TripleTen is a leading edtech platform in the US

Higher ● — ● — ○ Lower ● — Not comparable

	Employment Rate Within 6 Months	Median Salary After Graduation	Switchup Rating <sup>1</sup>	Course Report Rating <sup>1</sup>
<b>tripleTen</b>	87%	\$76,600	4.88	4.88
Tech Elevator	83%	\$65,505	4.91	4.90
Codesmith	70%	\$124,800	4.92	4.89
Flatiron	68% <sup>3</sup>	\$72,000	4.59	4.46
Hack Reactor	69%*	\$80,000	4.75	4.62
Fullstack Academy	72% <sup>4</sup>	\$88,561	4.84	4.80
General Assembly	69% <sup>5</sup>	N/A	4.27	4.31

**FORTUNE**

**Best overall provider** of software engineering bootcamps<sup>2</sup>

---

**80+ employers**

among partners that offer job opportunities to graduates

Note: Bootcamps listed in the table are those that TripleTen regards as its peers. Source of data: latest data reported by companies, e.g. to the Council on Integrity in Results Reporting (CIRR), and available in the public domain in October 2024. For bootcamps that calculate employment rates and salary statistics separately for different campuses and / or study tracks, an overall average is taken. If the employment rate for 6 months is not reported, the one-year rate is provided instead for illustrative purposes and marked with (\*); (1) SwitchUp and Course Report are major review platforms for US bootcamps, ratings are based on aggregated student reviews; (2) Fortune Magazine's ranking, 2024; (3) Data is not reported anymore – latest available used; (4) As of 2021 Outcomes Report; (5) As of 2022 Outcomes Report.

## What our customers say: testimonials and recognition

4.9 ★★★★★

Across 800+ reviews on Career Karma, SwitchUp, and Course Report<sup>1</sup>



“AC’s strong data skills and positive mindset have made him a standout at OMD USA. Completion of a TripleTen bootcamp and ability to optimize processes have positioned him as a valuable contributor in driving business decisions and leveraging media data.”

Ryan Mu,  
Manager at OMD USA



“We hire TripleTen students because of their ability to work well, own the project, care about it and solve new and complex problems. Any software engineer, even a junior, should demonstrate these qualities on their level”

Ivan Shcheklein,  
Co-owner at Iterative

### Student reviews

“Because of TripleTen’s sprint based project schedule, I was able to study and build on my own time, but still practice the pace of delivery that I use in my new career.”

“They prepare you for a real career. They give you a taste of what your professional tasks will look like.”

**FORTUNE** Education

### Best overall<sup>2</sup>: TripleTen

“TripleTen’s software engineering bootcamp aims to teach total beginners the languages and tools needed to become a full-stack developer.”

### Why we picked it

“The bootcamp strikes the best balance of costs, curriculum, brand prestige, and student experience. The program also merges theoretical study with hands-on, practical application through multiple 2-3 week sprints, while teachers provide line-by-line code reviews, portfolio-building opportunities, daily office hours, and access to experienced mentors.”

(1) As of October 2024; (2) Fortune Magazine's ranking, 2024.



## TripleTen key competitive advantages differentiate it from other players

- 1 Affordable pricing** enabled by optimization of training costs through diversified sourcing of trainers, automation of e-learning workflows, and AI-enabled student support and guidance
- 2 Top-rated edtech in the US** by employment rate and student feedback with a solid median salary after graduation. 87% of graduates find a job within six months of graduation, ahead of the 60-82% reported by most key peers<sup>1</sup>
- 3 Proprietary tech stack and automated platform** that enable seamless launching, scaling and localization of courses at minimal additional cost
- 4 Extensive assistance and guidance** provided by tutors and experts as well as **flexible e-learning environment** allow customers to tailor the study journey to personal needs and knowledge level and to receive significant value-add from solving complex problems provided by real companies

(1) Based on the Outcome Reports analysis in October 2024.

## This is just the beginning: significant potential for TripleTen growth across multiple levers

### Grow market share in existing markets

- Continue to **expand** in the US & **gain share from competitors**
- Growth in the number of alumni to help **enrol new students and optimize CAC**

### Continue to expand B2C offering

- Launch new programs:**
- Cyber Analyst (recently launched)
  - UI/UX designer (near future launch)

### Geo expansion to LATAM

- **Deep localization** to better address market needs
- **Fast scaling** of solutions that have already been tested
- **Expansion of courses** available in LATAM, **increasing penetration** across existing LATAM markets

### B2B product expansion

- Launch **B2B product for upskilling**
- Corporate bootcamps in **English and Spanish**
- **Role assessment** for data and development specialists
- Potential and **motivation assessment** service

### AI-enabled proprietary platform

Further development and integration of AI across the platform to **fuel further efficiency and enable further scaling**

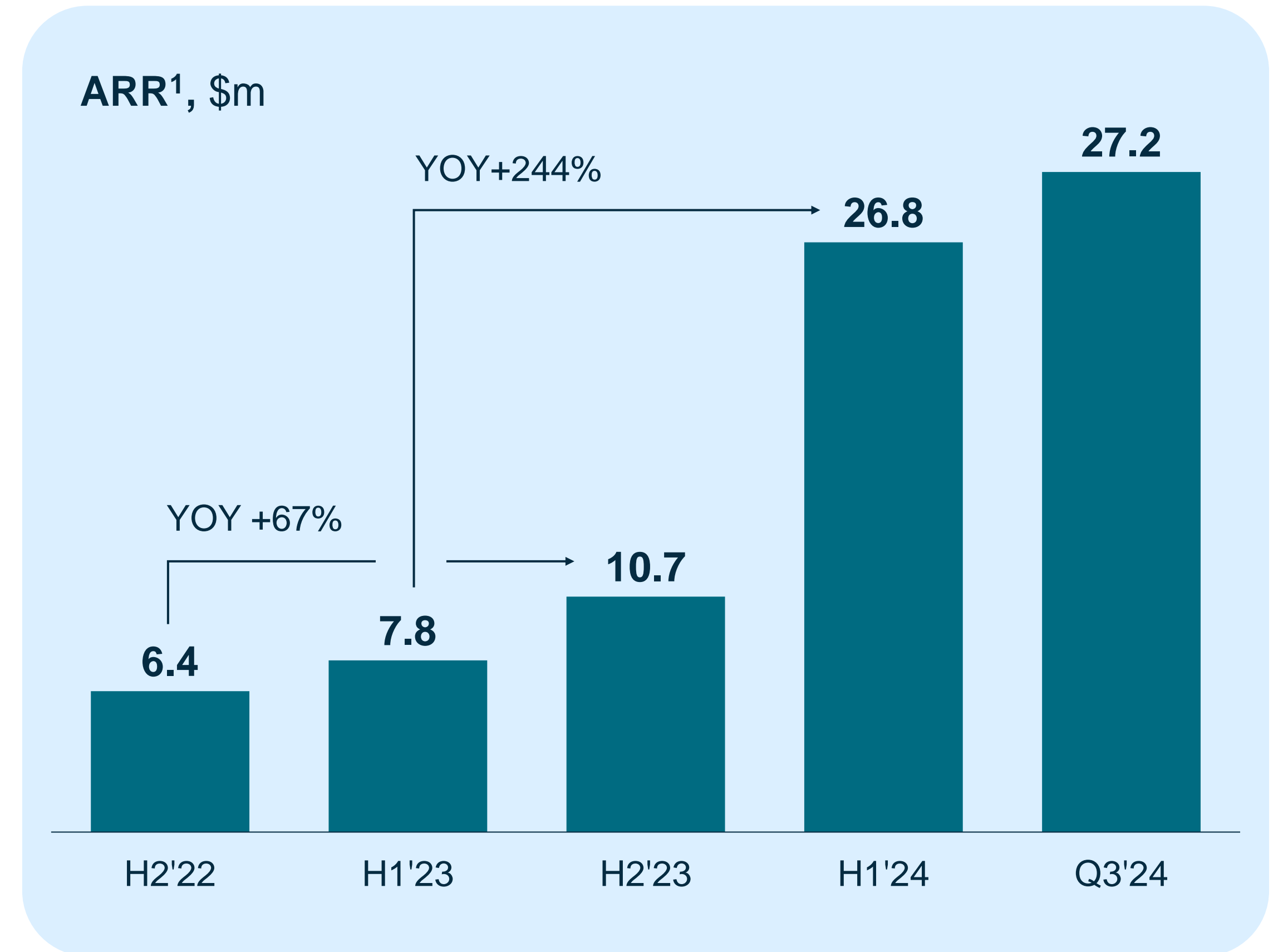
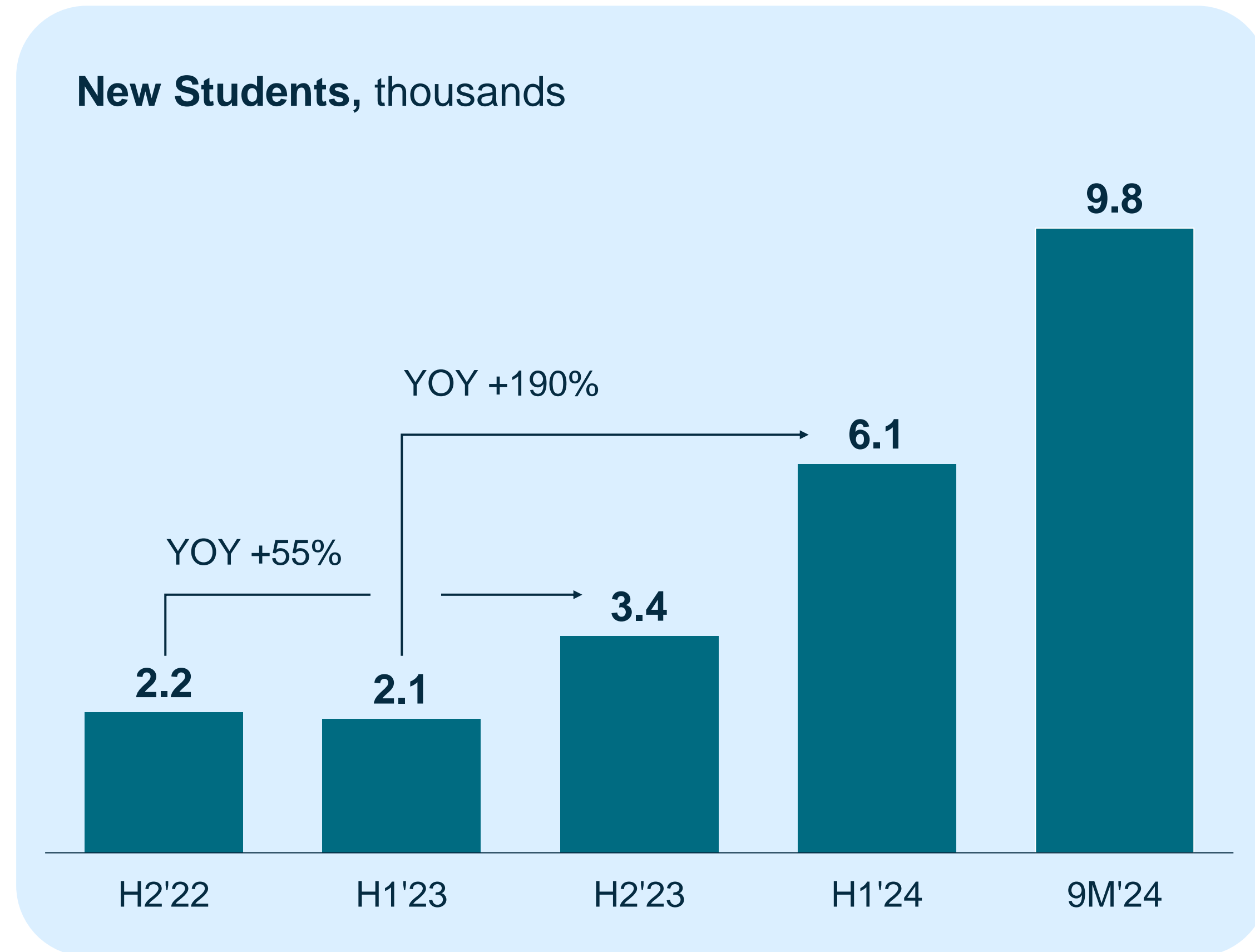
First **AI-related courses** for B2B clients on practical aspects of implementation of AI tools and apps for managers and technical staff

**AI Co-pilot** to assist B2B clients in assessing hard skills during interviews with skilled specialists

AI-driven identification of students struggling during the course to **provide help and guidance in advance**

# Financial highlights

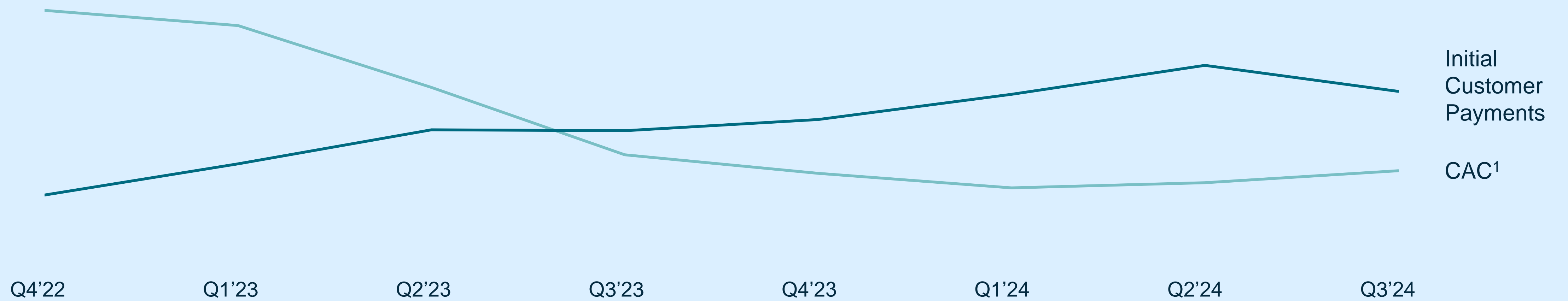
## Historical and current trading



(1) Annualized run-rate revenue as of the end of the period (calculated as revenue as of the last month of the period multiplied by twelve); Q3 2024 ARR is based on preliminary revenue of September 2024 as per management accounts.

## Efficient business model with optimized costs

CAC fully covered by initial customer payments as of Q3'23



(1) Customer acquisition costs comprised principally of marketing and advertising expenses; per one student.

Recap

## TripleTen's key investment highlights

Top-rated US bootcamp characterized by high employment rate, salaries and student feedback

Automated flexible e-learning platform simplifies launching and scaling on new markets and enables affordable pricing for customers and cost savings for the business

Growing presence in both B2B and B2C opening massive growth opportunities across key focus markets (US and LATAM)

[Visit website](#)

# Avride



**Developer of autonomous driving technology for self-driving cars and delivery robots**



## Who we are?

**200+**

self-driving tech engineers & software developers<sup>1</sup>

**7+ years**

experience of the team in developing autonomous solutions

Headquarters



**Austin, TX**

US-based executive team with additional offices in Tel Aviv, Belgrade, Seoul

Testing and operation



**USA & South Korea**

R&D hubs and offices



**USA, Israel, Serbia and South Korea**

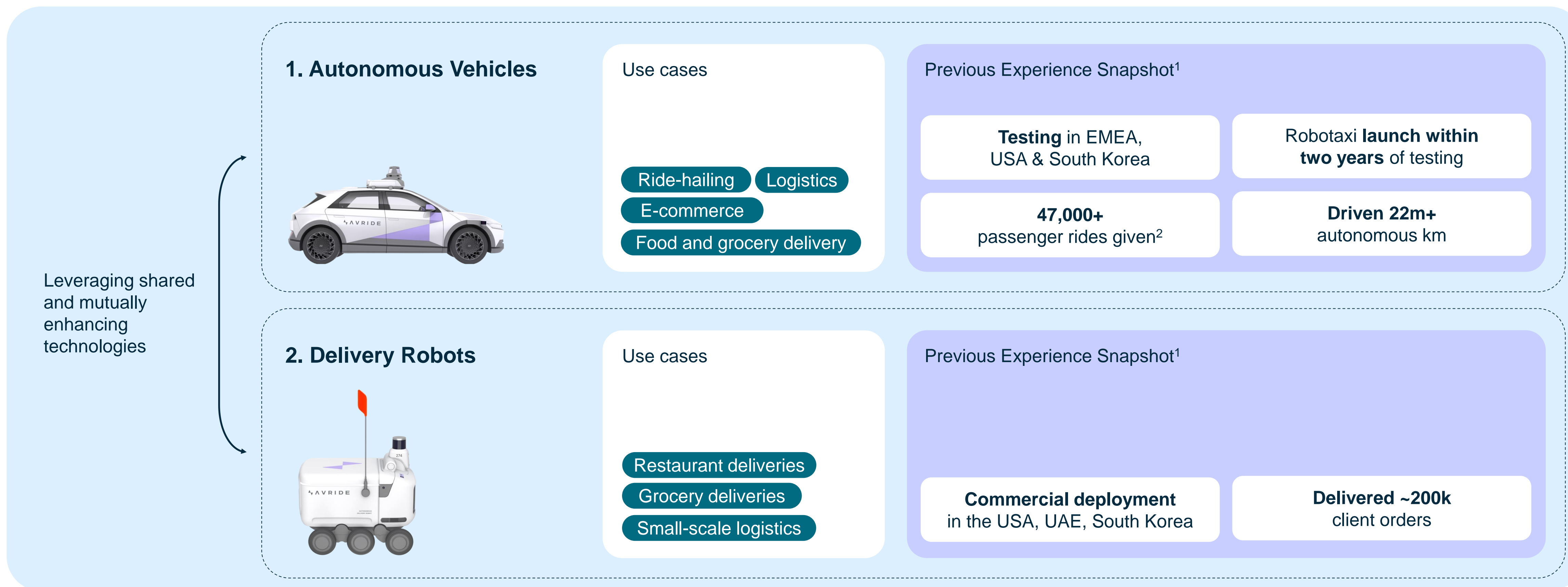
Strategic partnership with Uber on autonomous delivery and mobility

**Uber**



(1) As of September 30, 2024.

# Avride is uniquely positioned in the industry by simultaneously developing autonomous driving technology for cars and delivery robots



(1) The number of passenger rides, driven kilometers and delivered orders is cumulative for the period since the launch of the service until September 30, 2024, including previous operations as Yandex Self-Driving Group in 2017-2021; (2) Non-commercial rides.

# Autonomous vehicles and delivery robots leverage shared and mutually enhancing technologies



**Autonomous vehicles** advancements drive the development of delivery robots by applying new car-related solutions to robots

**Delivery robots'** faster deployment and monetization support future cars development and deployment



## Common elements

Models training system

Code deployment and data collection

Simulator, offline algorithms and more

Localization

Perception

# Scalable business model focused on large and attractive end markets

## Core segments under development

**\$46bn TAM in 2030<sup>1</sup>**

### Robo taxis (B2B)

- Focus on fleet management software, tech maintenance and local operations
- Cooperation with e-hailing partners (i.e. Uber, Lyft, etc.)

### Intra-city delivery (vans/ robo taxis)

- Delivery from hub to small warehouses
- Delivery of medium-size packages to end customers (doorstep)

**\$6bn TAM in 2030<sup>2</sup>**

### Last mile delivery robots

- Restaurant deliveries
- Groceries
- (Small) e-commerce deliveries
- Subscription services (e.g., logistics companies, hotels, etc.)

## Potential future upsides

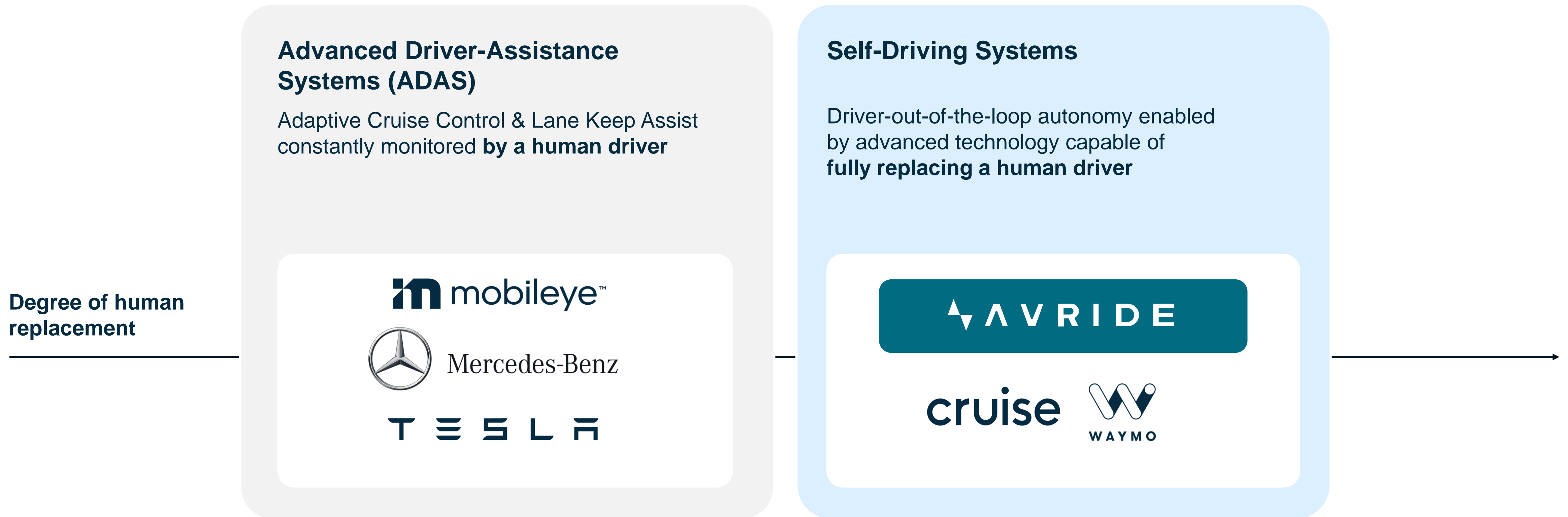
**~\$14bn TAM in 2030<sup>3</sup>**

### Inter-city delivery

Highway and non-urban transportation

(1) Markets&Markets Forecast, 2023; (2) Grand View Research, 2023; (3) Management calculations based on third party data; calculated as an average of forecasts ranging from \$4bn to \$23bn+; Precedence Research, 2023; Allied Market Research, 2023.

# Avride is working on the highest level of automation, developing fully autonomous vehicles



# Rapid fleet progression, reaching first passenger rides in under two years

Enabled by the team's expertise across multiple domains, including **ML, computer vision, hardware engineering, sensor technologies**

## Zero

serious accidents<sup>1</sup> over 22 million kilometers driven autonomously on public roads since the launch of the service in 2017

Testing

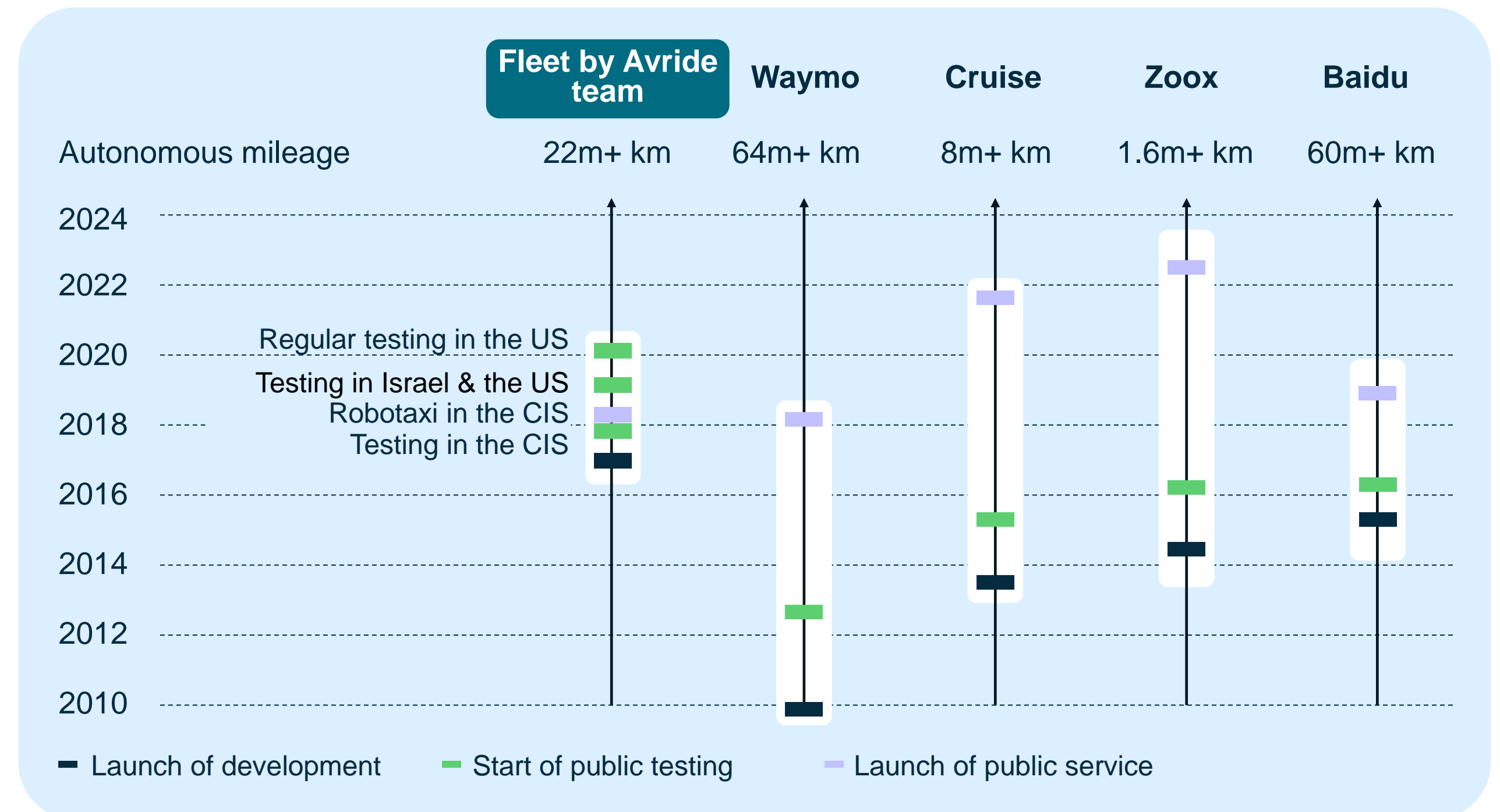
### Different countries

Since 2017, we have conducted tests in four countries, each with own unique road infrastructure and driving culture

### Diverse conditions

from wide & congested to narrow & parked-up streets; in rain, snow, and heat

One of the fastest AV fleets to progress to public testing and one of the first to offer passenger rides despite being one of the latest players to enter the AV market



(1) 0 incidents with human injuries or significant damage to the vehicle.

# Avride delivery robots have been successfully tested and can operate in diverse conditions

## Snapshot: Delivery robot

### Ready for various road conditions

Successful testing in the most diverse conditions including city streets with heavy vehicle and foot traffic to ensure safe operations

Segment	Sidewalk, Indoor
Max speed	8 kmh <sup>1</sup>
Driving range	55 km
Max sidewalk height to climb/descend	100 mm
Wireless charging option	Yes
Battery replacement option <sup>2</sup>	Yes

## Delivery robots production

**In-house design of robots** allows company to quickly introduce any upgrades

**Factory partner in Taiwan** provides access to high-quality and rapidly scalable production and assembly of robots

**Potential for optimization** of cost per robot at scale



(1) Software-enforced speed limit to ensure pedestrian safety; the robots' design allows to reach a maximum speed of 24 kmh; (2) Delivery robots can charge wirelessly or have the battery replaced in ~1 min instead of being connected to the power source.

## Successful and quick rollout will enable further commercial partnerships globally

**Suitable for complex road conditions**

**~200,000**

orders delivered by robots designed by the team since the launch of the service

**2–3 weeks**

for robo-delivery roll-out in a new location: from mapping & testing to integration into partner service<sup>1</sup>

### Commercial partners

Signed a **multiyear strategic partnership with Uber**

Delivery partnership with **sidewalk robots** on Uber Eats in **Austin in the coming weeks** before expanding to Dallas and Jersey City, New Jersey, later this year

**Mobility partnership** is expected to launch for riders in **Dallas later next year**

**Uber**

Negotiating with other **food- & ride-tech companies** to broaden commercial reach

Plan to **integrate hundreds additional robots** into commercial delivery services in 2025

(1) Based on Avride's launch process in South Korea in Jun-23.





## Human-less delivery robots operations

**Goal:** Enable fully autonomous operations with minimal human intervention, and maintaining operations on site

**Solution:** Implement wireless charging stations, manage data via cellular connection, and ensure hardware redundancy for continuous operation

**Result:**

- Over 10 robots operating autonomously on campus
- Engineers only required on-site once a month for routine maintenance



## High-load delivery robots operations<sup>2</sup>

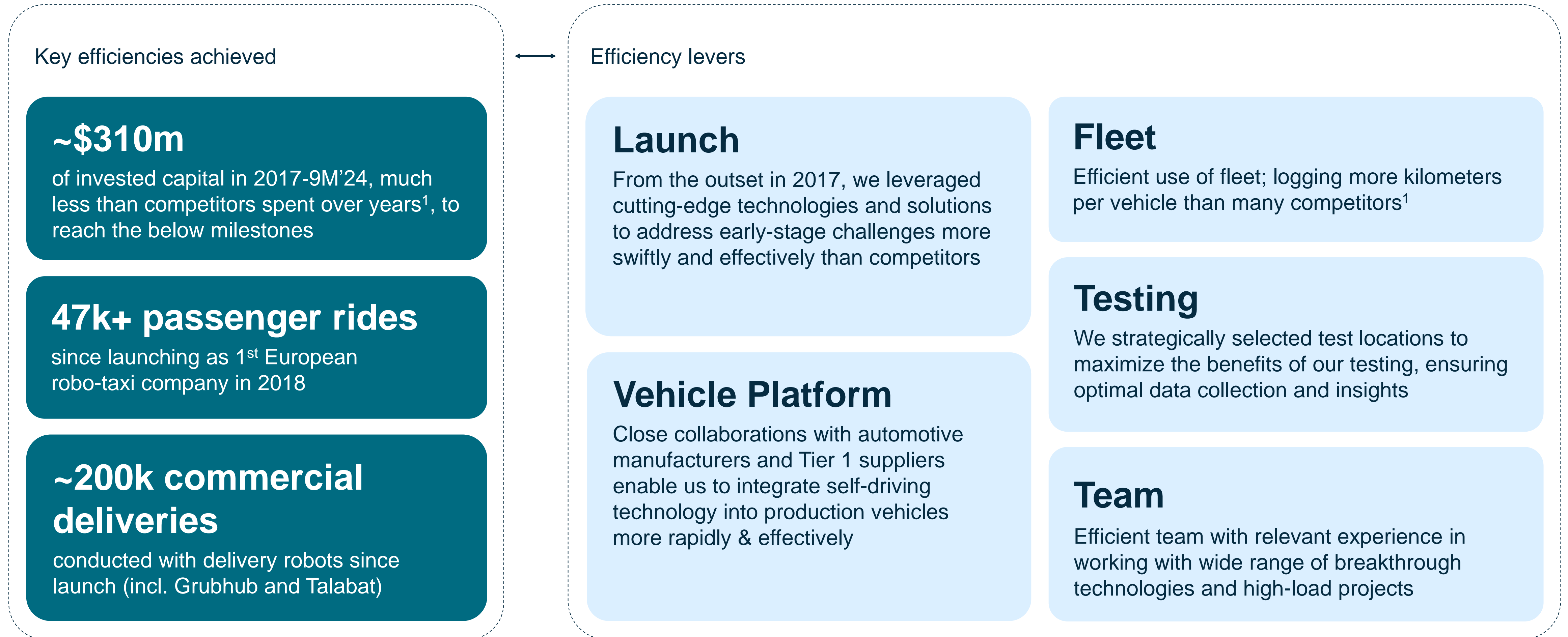
**Goal:** Meet the delivery demands of one of the largest university campuses in the U.S.

**Solution:** Deploy advanced robot dispatch algorithms to maximize efficiency and ensure smooth operations, avoiding traffic jams or crowds

**Result:**

- Up to 60 robots operating simultaneously
- Handling up to 1,000 deliveries per day

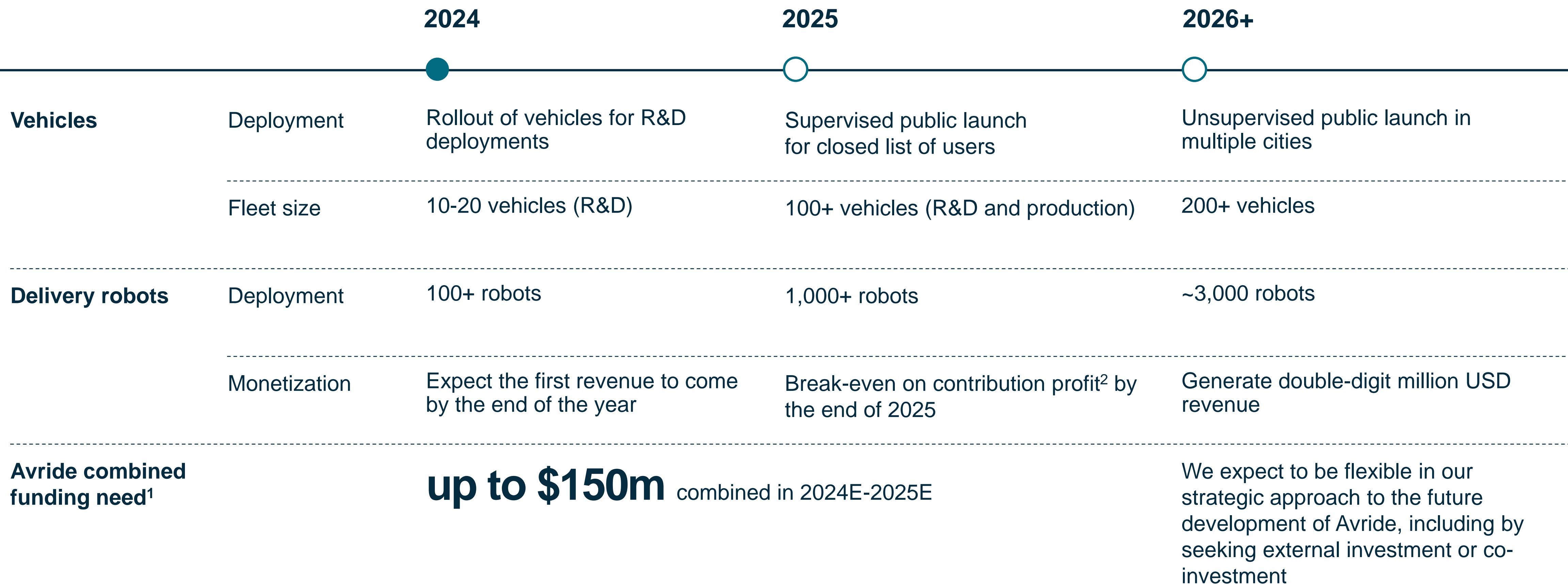
## Avride's autonomous vehicle offering has been highly efficient since launch



(1) Compared to Waymo, Cruise and Zoox.

# Financial Highlights

# Avride mid-term growth strategy and required investments



(1) In addition to ~\$49m invested in the 9M'24; (2) Calculated as Revenue minus robots direct per-unit costs, including costs like Remote assist and Mechanics (on the ground personnel).

Recap

## Avride key investment highlights

One of the most experienced teams in the autonomous vehicles industry

Fast track to commercial deployment at a fraction of costs

Partnerships with leading automakers and delivery platforms

Technology tested in diverse road conditions with outstanding safety record

# Nebius Group Financials and Outlook

# Nebius future growth driven by significant expansion in TAM and further penetration ...

## NEBIUS

### Management plans

Leverage GPUaaS and AI Cloud **tremendous market growth** & continue **expanding market share**

Continue **expanding and diversifying the client base**, securing **additional long-term contracts** and **entering new customer segments**

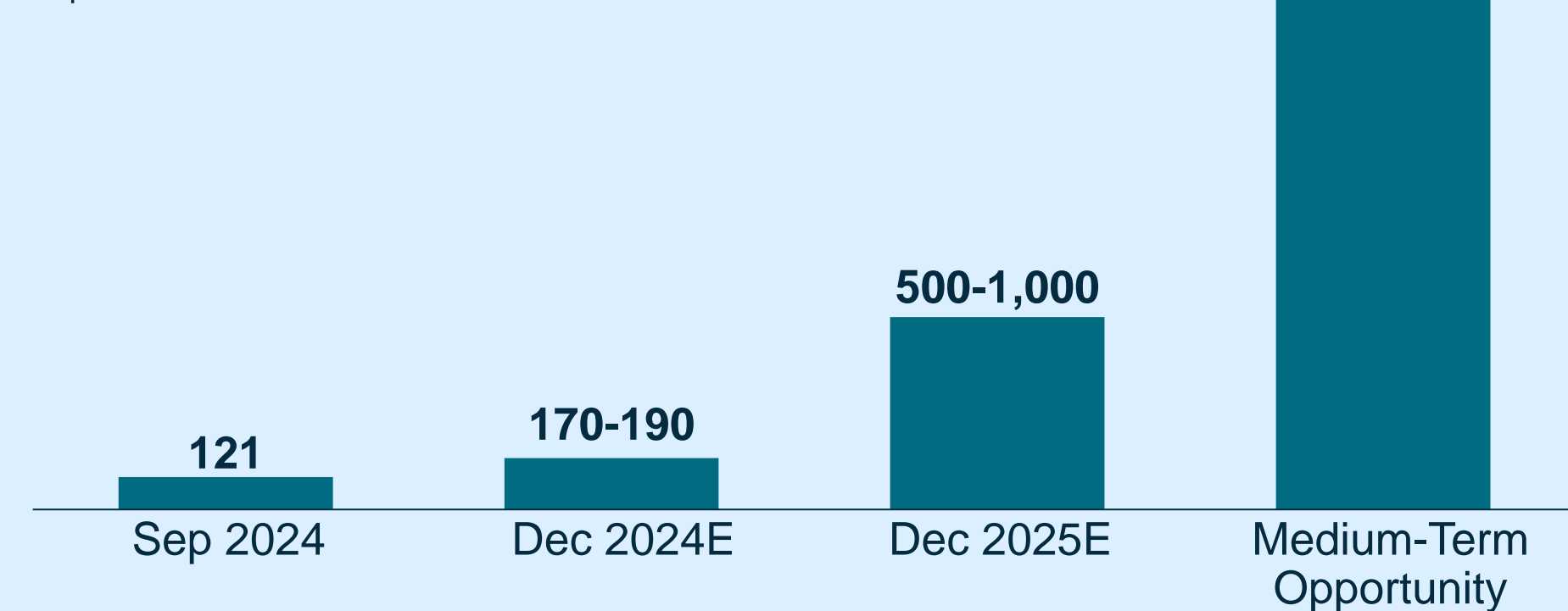
**Significantly expand the infrastructure** by adding **data centers** (owned, greenfield, and colocation) and **increasing GPU capacity**, to support **higher volumes** for existing and new customers

Continue **expanding product range**, offering more **value-add services**

# 4-6x

GPU capacity<sup>1</sup> expansion  
Sep'24 – Dec'25E

ARR<sup>3,4</sup>  
\$m



(1) GPUs available in H100 equivalent; (2) Based on the planned capacity expansion and assumed utilization and pricing as per the company's medium-term financial plan (as of September 2024); (3) Annualized run-rate revenue by the end of the period (revenue for last month of the period multiplied by twelve); Sep-24 ARR is based on preliminary revenue as per management accounts; (4) Depending on the amount of capital available to cover the company's investment program during the respective periods.

## ... supported by growth across our other businesses

### Toloka

Expand revenue across all Toloka revenue categories and **continue establishing presence in the growing market**

Secure **new customers and expand wallet share** with existing clients for **Classic GenAI Solutions**

Capitalise **growing demand** across Evolved GenAI, Self-serve platform, and non-core non-Gen AI services offering

2025E revenue expected to be **\$50-70m**  
Capital expenditures for Toloka are immaterial

### tripleten

Grow **market share in existing markets**, expanding in the US, **gaining share** from competitors

Continue to **grow B2C offering** and **expand B2B product**

Expand offering and **increase penetration in LATAM**

2025E revenue expected to be **\$40-60m**  
Capital expenditures for TripleTen are immaterial

### AVRIDE

Continue developing and expanding commercial reach to **scale operations globally**

Expand **vehicle and robot operating fleet size**

Continue being a **leader in efficiency in commercial deployment**

**Up to \$150m** combined funding need expected in 2024E-2025E; initial revenue expected from partnership with Uber and other sources



## Nebius Group expects to deliver \$500-700m revenue in 2025 and push towards Adj EBITDA breakeven

### Current trading and 2024 guidance

	Q3'2024 <sup>3</sup>	2024E
Revenue	\$43m	\$120-130m
Adj. EBITDA <sup>2</sup>	(\$52m)	(\$250-300m)
Capex	\$167m	~\$1-1.1bn

### 2025 Outlook

Depending on the amount of capital available in 2025 to cover the company's investment program we intend to:

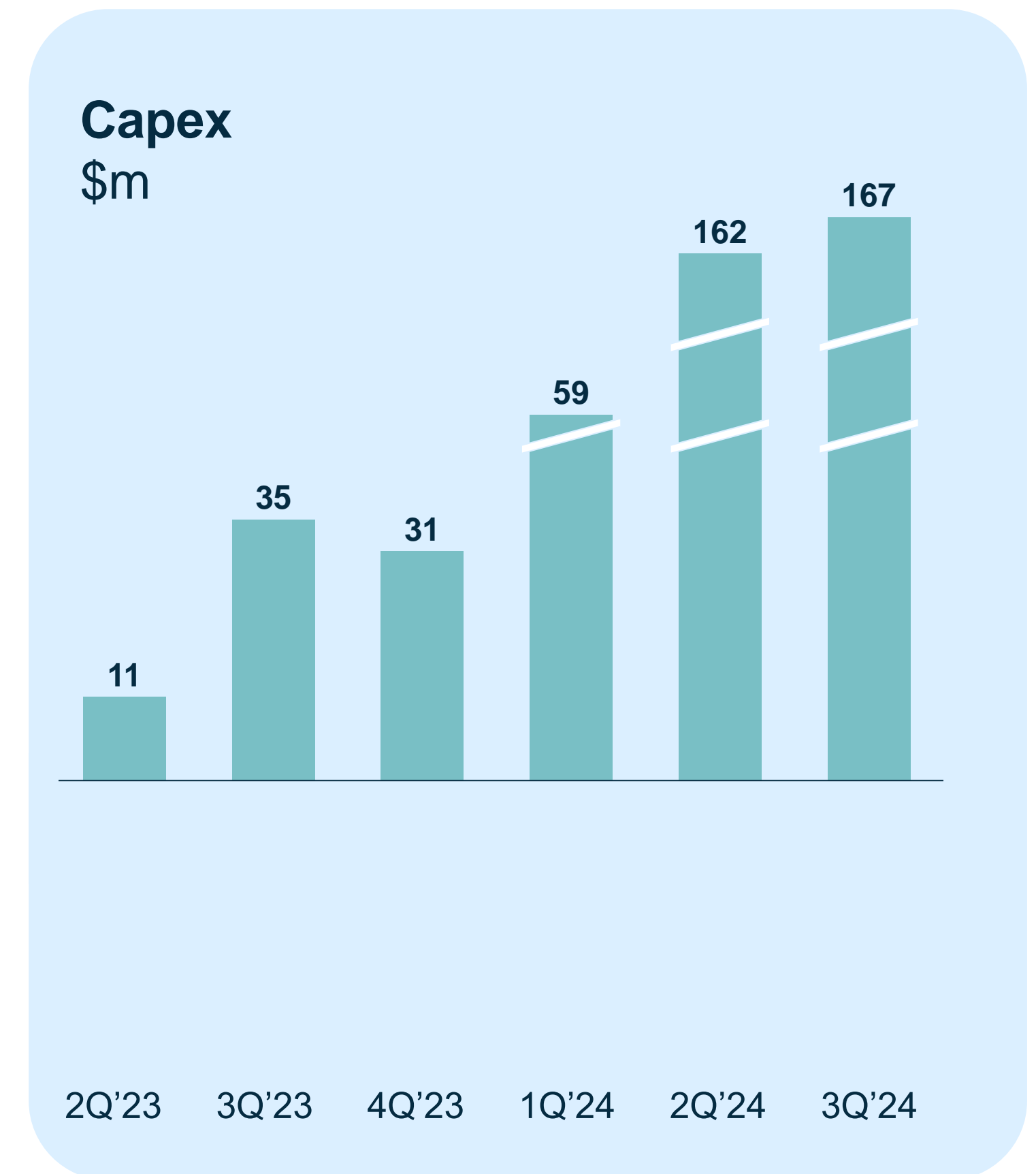
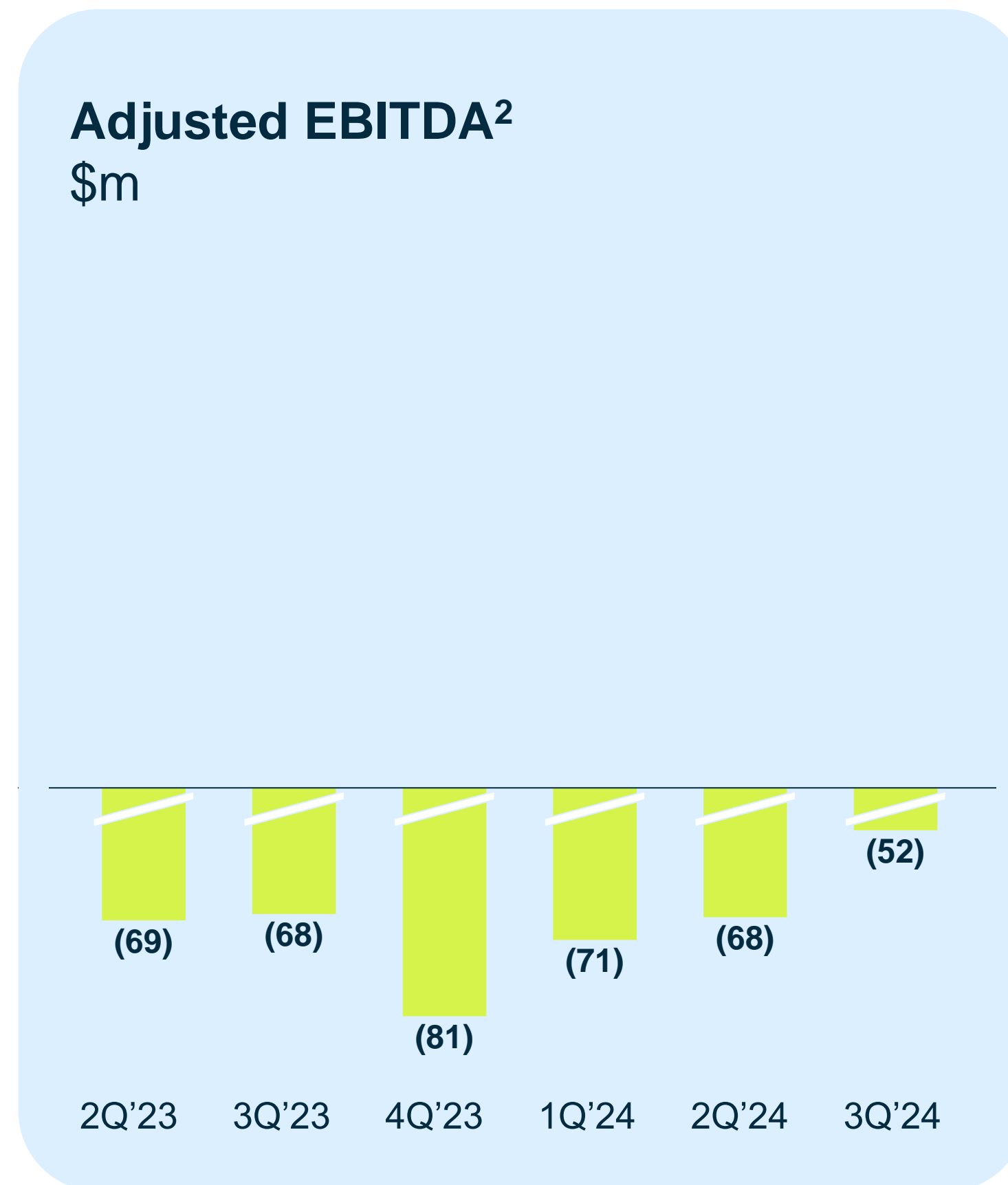
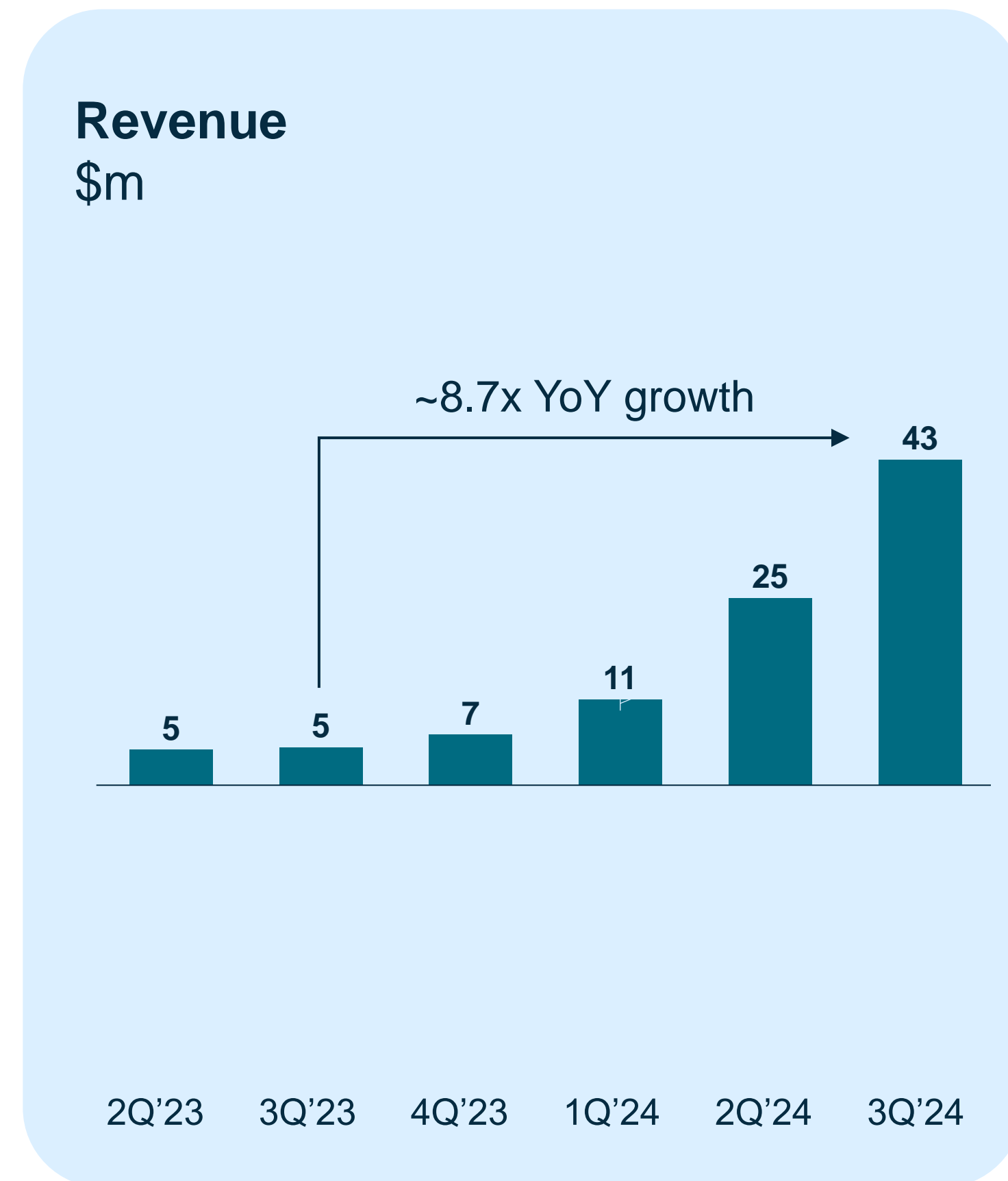
- spend between **\$600m and \$1.5bn on CAPEX**, primarily on the core Nebius business (GPUs); capex for Toloka, TripleTen and Avride is immaterial compared to Nebius
- grow **revenue** by **3-4x to \$500-700m** based on assumed capacity expansion, primarily driven by Nebius core business, as well as over 2x revenue growth in both Toloka and TripleTen
- move into a **positive territory on Adj EBITDA during 2025** (for the full year we will likely maintain below break-even Adj EBITDA)

(1) Adjusted EBITDA/(loss) means net income/(loss) from continuing operations plus depreciation and amortization, certain SBC expense, interest expense, income tax expense, one-off restructuring and other expenses, less interest income, other income/(loss), net, and income/(loss) from equity method investments. Please refer to the Appendix for reconciliation of Q3 2024 Adjusted EBITDA figures; (2) Adjusted EBITDA for the group also includes corporate overheads; (3) Preliminary Q3 2024 financial results based on the management accounts.

# Appendices

Historical trading	<u>91</u>
Management overview	<u>92</u>
BoD overview	<u>93</u>
Employee snapshot	<u>94</u>
Shareholder structure	<u>95</u>
Sustainability overview	<u>96</u>
Reconciliation of Adjusted EBITDA to US GAAP Net income / (loss)	<u>97</u>

# Historical and current trading<sup>1</sup>



(1) Q3 2024 results are based on preliminary management accounts; (2) Adjusted EBITDA/(loss) means U.S. GAAP net income/(loss) from continuing operations plus depreciation and amortization, certain SBC expense, interest expense, income tax expense, one-off restructuring and other expenses, less interest income, other income/(loss), net, and income/(loss) from equity method investments. Please refer to the end of Appendix for reconciliation of Q3 2024 Adjusted EBITDA figure.

## Nebius Group



**Arkady Volozh**  
CEO and Founder

Visionary founder; previously co-founded successful Tech companies including Yandex and CompTek International

### C-Suite



**Ophir Nave**  
COO

Proven leader and legal and corporate finance expert, having previously led Corporate M&A at Arnon, Tadmor-Levy



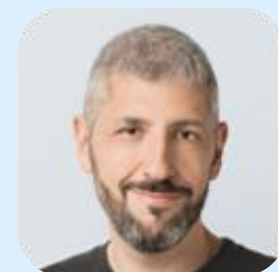
**Ron Jacobs**  
CFO

Finance expert with hands-on experience leading M&A and transformation programs in Europe's biggest telco companies



**Andrey Korolenko**  
Chief Product & Infrastructure Officer

20 years supervising IT infrastructure development and expertise in deploying supercomputers ranking among TOP-20 most powerful ones globally



**Roman Chernin**  
Chief Business Officer

12 years heading the development and deployment of digital B2B and B2C services (including Search and Maps) and an experienced startup mentor



**Tom Blackwell**  
Chief Comms. Officer

Strategic communications expert, chairman, founder and previously CEO of EM, an emerging markets focused PR/IR advisory firm

## Business units

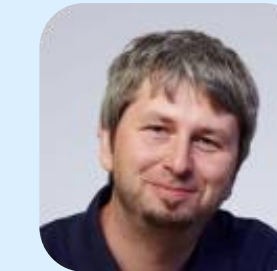
### Toloka



**Olga Megorskaya**  
CEO

10+ years of experience in AI & Data Labeling

### TripleTen



**Ilya Zalessky**  
CEO

6+ years driving edtech ventures expertise in startup consulting and go-to-market strategies

### Avride



**Dmitry Polishchuk**  
CEO

7+ years heading self-driving tech, 13 years of experience leading mobile services

# Nebius Group Board of Directors

 — based in

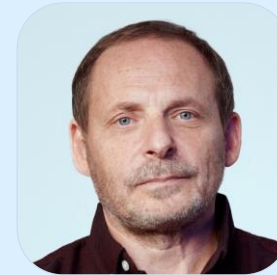


**John Boynton**  
Chairman

Founder, investor and board member of tech ventures with 30+ years in IT and tech-related fields

Non-executive

Independent



**Arkady Volozh**  
CEO

Visionary founder; previously co-founded successful Tech companies including Yandex and CompTek International

Executive



**Ophir Nave**  
COO

Proven leader and legal and corporate finance expert, having previously led Corporate M&A at Arnon, Tadmor-Levy

Executive



**Charles Ryan**  
Board Member

Founder and General Partner of an international VC firm Almaz Capital, previously CCO at Deutsche Bank and Chairman of UFG Asset Management

Non-executive

Independent



**Rogier Rijnja**  
Board Member

Independent management consultant and entrepreneur, previously senior HR executive at various global companies

Non-executive

Independent



**Esther Dyson**

Founder of Wellville and active investor in a variety of IT, health, logistics and space start-ups. Author of "Release 2.0: A design for living in the digital age"

Non-executive

Independent



**Kira Radinsky**

CEO and chairwoman of Diagnostic Robotics, previously serving on the board of directors of various companies

Non-executive

Independent



**Elena Bunina**  
Head of Science & Education

25+ years of experience teaching data analysis & maths in leading universities. Previously, Head of Academic & Educational Services as well as HR director at Yandex N.V.

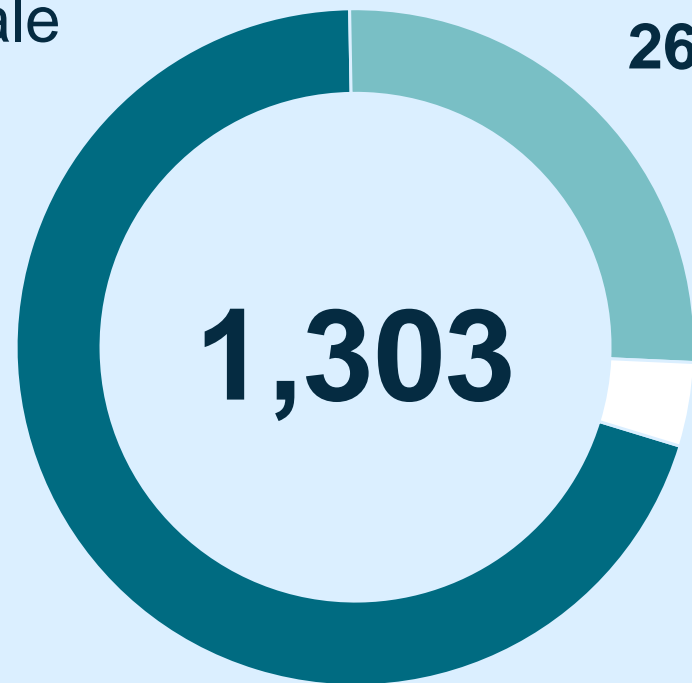
Non-Executive

# Nebius Group employee snapshot<sup>1</sup>

## Diversity

### Total headcount, people

70% male



26% female

4% non-binary / prefer not to state

Share of women is in line with industry trends for IT companies

# 43

nationalities



## Geographical presence

# 86%

of employees are based in the Netherlands, Serbia, Israel and USA



# ~2/3

of **Nebius** (core business) employees are based in **the Netherlands**

As well as **36%** of the total company headcount



(1) Data as of Q3 2024.

# Nebius Group ownership structure

## Total shares issued & outstanding

As of September 30, 2024

Class A

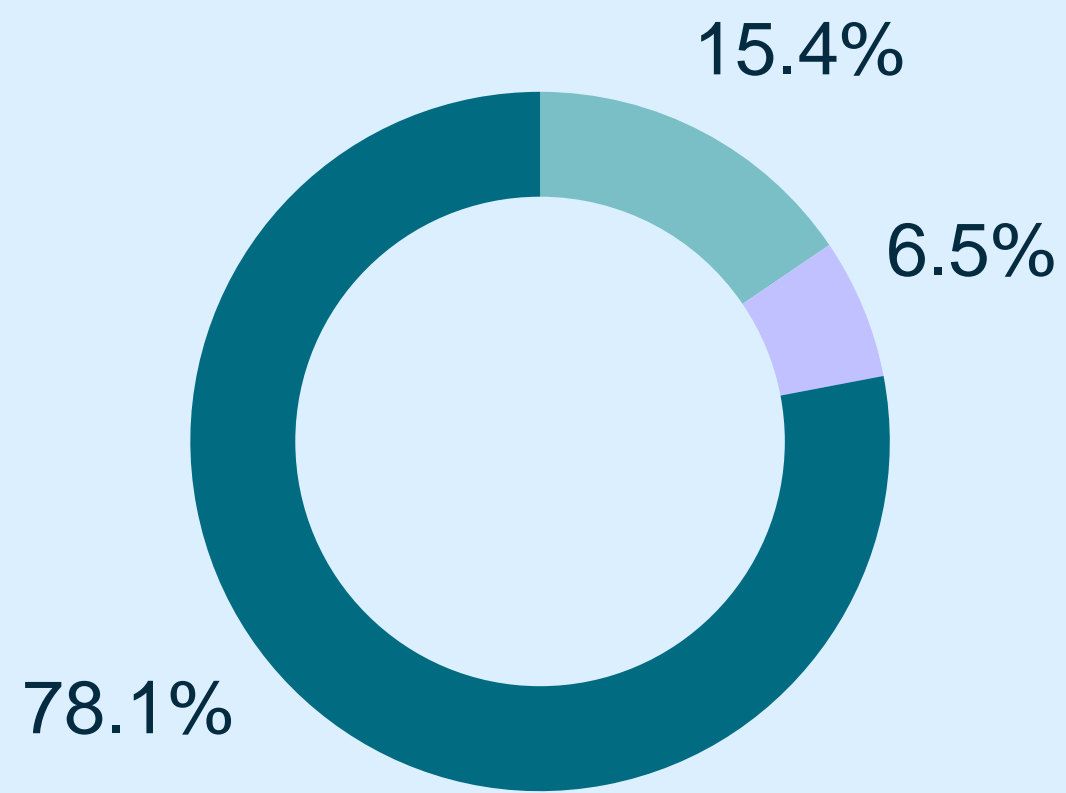
**163,641,444**

Class B

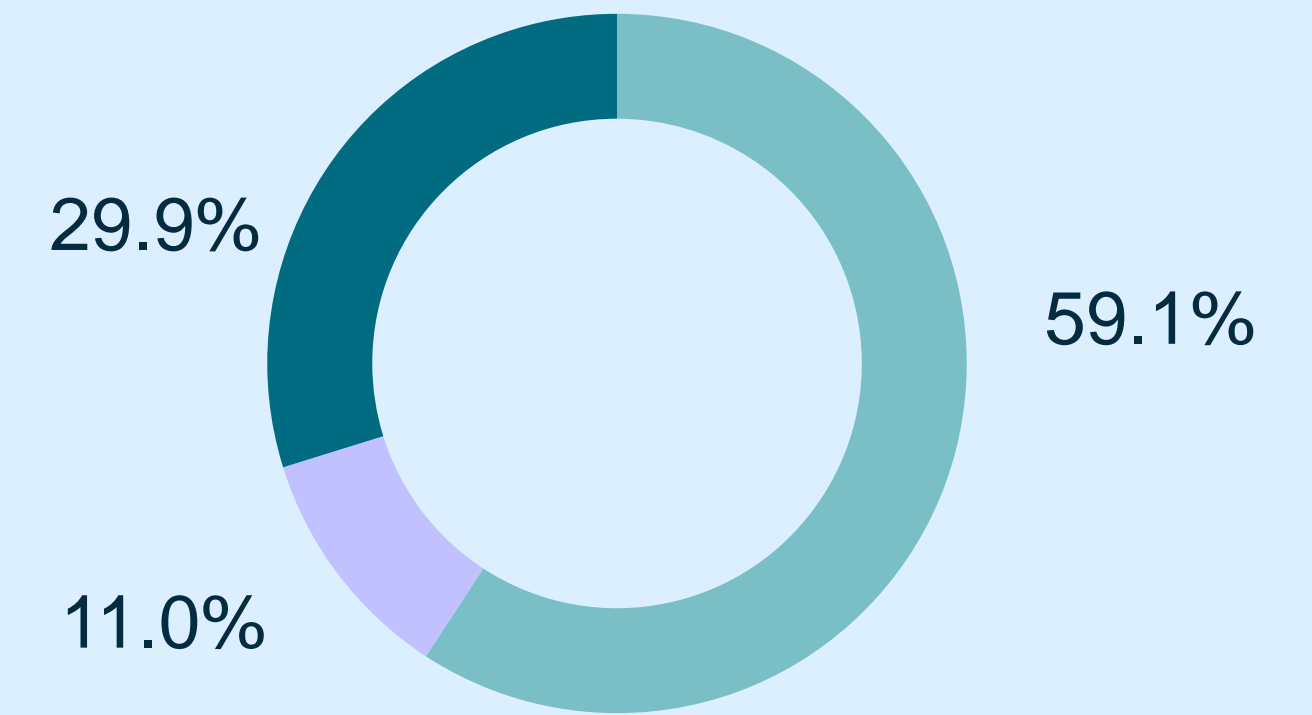
**35,698,674**

Excluding 162,700,826 Class A shares held in treasury pending use under our equity incentive plans and for further financing purposes

## Economic ownership



## Voting power



● — LASTAR Trust<sup>1</sup>   ● — Other directors, officers & employees, pre-IPO shareholders   ● — Free float

(1) Family trust established by Mr. Volozh in December 2019.

**Committed to doing business responsibly and sustainably at every stage of the journey, to becoming a leading player in AI, and to ensuring that products and solutions have a positive impact on society**

## Key areas of focus

### Information security & offline safety

Dedicated to safeguarding both the virtual and physical realms to ensure a seamless and secure experience for users

- Continuous monitoring of system performance
- Regular updates of cybersecurity systems and procedures
- Engagement of third-party audits
- ISO 27001, ISO 27701 among security certificates possessed



### Sustainable computing

Invest energy efficiency of the computing infrastructure to do more with fewer resources

**PUE 1.13 under max. loads** compared to world's average PUE of 1.58 in 2023<sup>1</sup>

Employ innovative tech such as free cooling, waste heat recovery, and supercomputing

**TOP 5% most efficient** supercomputer among 500 most powerful supercomputers<sup>2</sup>



### Social inclusion

Healthy and fair work environment for people to drive IT innovation

**Highly skilled international workforce:** 40+ nationalities, o/w 60% in STEM roles

Strive to shape a more diverse, dynamic, and accessible technology sector via digital solutions & edtech offering

**87%** TripleTen graduates' employment rate within 6 months of graduation in the US



(1) Uptime Institute Global Data Center Survey Results 2023; Nebius' figure represents PUE of Finnish data center under max. loads in 2020-2021; Due to corporate restructuring, the data center has not operated at full capacity since 2022, so the measurement pertains to the period before that year; (2) Top 500 & Green 500 list, June 2024.



Reconciliation of Nebius Group Adj. EBITDA

## Reconciliation of Adjusted EBITDA to US GAAP Net income / (loss)

	2024				2023				
	Q1	Q2	H1	Q3 <sup>1</sup>	Q1	Q2	Q3	Q4	FY
<b>Net income / (loss)</b>	<b>(316.5)</b>	<b>(97.5)</b>	<b>(414.0)</b>	<b>(91.1)</b>	58.5	176.3	84.3	(72.0)	247.1
Less: Net income from discontinued operations	236.9	58.2	295.1	-	(135.7)	(250.6)	(181.1)	(15.4)	(582.8)
<b>Net loss from continuing operations</b>	<b>(79.6)</b>	<b>(39.3)</b>	<b>(118.9)</b>	<b>(91.1)</b>	<b>(77.2)</b>	<b>(74.3)</b>	<b>(96.8)</b>	<b>(87.4)</b>	<b>(335.7)</b>
Add: depreciation and amortization	8.9	11.5	20.4	20.4	6.9	6.5	6.1	8.3	27.8
Add: certain SBC expense	2.2	(0.7)	1.5	4.7	1.8	3.3	3.9	3.9	12.9
Add: one-off restructuring and other expenses	-	43.5	43.5	19.6	-	-	-	-	-
Less: interest income	(0.4)	(12.7)	(13.1)	(28.6)	(1.2)	(1.0)	(0.5)	(0.6)	(3.3)
Add: interest expense	-	-	-	-	-	-	-	-	-
Less: Income / (loss) from equity method investments	-	(88.7)	(88.7)	29.3	0.8	1.4	9.8	(1.1)	10.9
Less: other income, net	1.0	14.7	15.7	(6.8)	4.1	(6.4)	7.7	(2.0)	3.4
Add: income tax expense / (benefit)	(3.0)	3.6	0.6	0.6	(0.5)	1.9	2.2	(1.2)	2.4
<b>Adjusted EBITDA</b>	<b>(70.9)</b>	<b>(68.1)</b>	<b>(139.0)</b>	<b>(51.9)</b>	<b>(65.3)</b>	<b>(68.6)</b>	<b>(67.6)</b>	<b>(80.1)</b>	<b>(281.6)</b>

(1) Q3 2024 reconciliation is based on preliminary management accounts.

Find more information on our website:  
<https://group.nebius.com/>

**Main Office**  
Gustav Mahlerlaan 300  
Amsterdam, Netherlands

**Investor Relations Contact**  
[askIR@nebius.com](mailto:askIR@nebius.com)