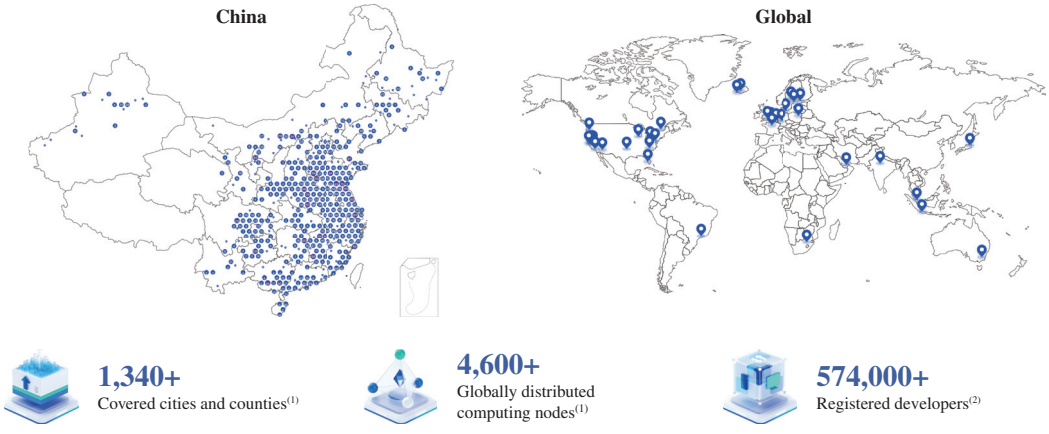
OVERVIEW

We are a leading independent distributed cloud computing service provider in China. According to CIC, we operate the largest computing power network in China in terms of the number of compute nodes as of December 31, 2025, and we are the largest independent edge cloud computing service provider in China in terms of revenue in 2025. We also rank first among independent AI cloud computing service providers in China in terms of average daily token consumption volume in 2025.

Our business began in the area of edge cloud computing services, as we identified significant mismatches between massive demand for and scattered, often underutilized supply of computing power. We established a computing power scheduling platform to aggregate and manage idle computing power from heterogeneous sources and deploy it in an efficient fashion to customers to accelerate content delivery and enhance fast-response, low-latency computing experience. In 2023, we strategically commenced AI cloud computing services, as the rapid growth of large models spurred a surge in demand for GPU computing power as well as AI inference capabilities. Leveraging our global compute scheduling network and comprehensive AI-native capabilities, we transform heterogeneous, cross-regional computing resources into low-cost, low-latency and highly scalable computing services, serving AI-native applications, developers and enterprises worldwide. Rather than simply providing access to computing resources, our core value lies in delivering reliable, efficient and optimized model services that customers can seamlessly integrate into their products and workflows without building or managing the underlying infrastructure. We are also extending our platform into "agentic infra," creating the foundational layer that will power AI agents' autonomous and durable task execution. As agentic AI proliferates, we expect the primary callers of cloud computing infrastructure to shift from human users toward AI agents, giving rise to entirely new demand for computing resources.

As of December 31, 2025, our computing power network covered more than 1,340 cities and counties globally with over 4,600 compute nodes. As of April 30, 2026, our AI cloud computing services had over 574,000 registered developers throughout the world. In April 2026, our average daily token consumption reached 1,028 billion, ranking us the largest independent AI cloud computing service provider in China, according to CIC.

Our Computing Power Network



Notes:

(1) As of December 31, 2025.

(2) As of April 30, 2026.

We have gained wide industry recognition for our technological strength and service capabilities. We provide edge cloud computing services and AI cloud computing services to many leading Chinese and international technology companies (including a majority of the top 10 Chinese major internet companies, according to CIC), cloud computing companies and "unicorn" startups.

During the Track Record Period, our revenue grew rapidly from RMB358.4 million in 2023 to RMB558.0 million in 2024, and further to RMB770.3 million in 2025, representing a CAGR of 46.6% from 2023 to 2025. In particular, our AI cloud computing services have experienced exponential growth since 2025, with revenue increasing by more than tenfold from RMB10.4 million in 2024 to RMB119.2 million in 2025.

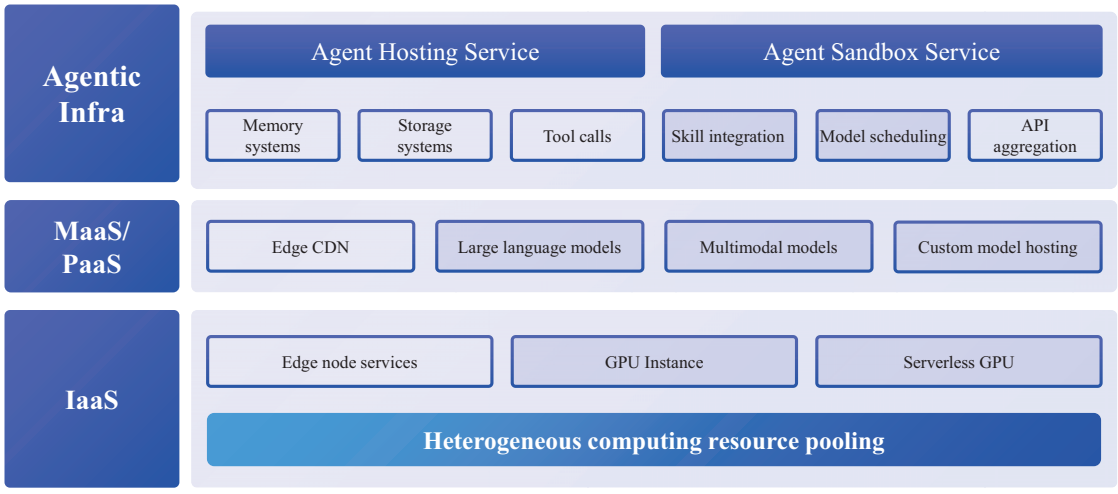
Our Market Opportunity

The rapid advancement of AI has created enormous demand for computing power. Despite increased supply, computing power globally is still used inefficiently with an average utilization rate between 50% and 70%, due to the heterogeneous nature, widespread geographic distribution and temporal misalignment of computing resources. While securing additional computing resources is important, an even greater challenge lies in transforming such dispersed capacity into stable, commercial-grade services.

Distributed cloud computing platform, with its flexible resource allocation and asset-light operating model, can effectively address these constraints. In particular, for AI computing demands, a distributed cloud platform provides a standardized accesses to dispersed, multi-generational computing resources across regions, integrating them into a unified, manageable capacity pool. Leveraging geographic and temporal differences in demand (for example, midnight in the western hemisphere is typically off-peak, while simultaneously, daytime in the eastern hemisphere is peak), such platform dynamically schedule underutilized computing resources to high-demand inference tasks, maximizing the overall utilization of resources across the network on a continuous basis. Distributed cloud computing has experienced explosive growth in recent years. According to CIC, in terms of revenue, the global edge cloud computing services market amounted to RMB209.8 billion in 2025; the Chinese edge cloud computing services market reached RMB15.3 billion in 2025, and is expected to approach RMB41.9 billion in 2030; the global AI cloud computing services market reached RMB188.2 billion in 2025, and is expected to approach RMB4,666.8 billion in 2030, representing a CAGR of 90.1%.

The development of AI and the introduction of large models are revolutionizing everyday life and work scenarios. While vast amounts of computing power have been utilized by AI developers to train their models to ever higher degrees of capability and sophistication, increasing resources are now devoted to AI "inference," the process of using trained AI models to generate insights in application scenarios. Currently, AI inference has exceeded AI training in terms of computing power consumption. AI inference applications permeate consumer and enterprise workflows at massive scale, generating persistent, high-frequency demand for efficient, elastic and cost-effective inference services. Looking ahead, the agentic AI era is emerging as AI transitions from passively generating answers to around-the-clock, proactive execution of complex, multi-step tasks. As AI agents proliferate, the primary callers of cloud computing infrastructure will shift from human users toward AI agents autonomously initiating tasks, giving rise to entirely new demand for computing resources to support multi-step, tool-intensive agentic workloads. We believe that, in the long run, the primary driver for the growth of demand for computing power will be agentic AI tasks, because agentic AI will increasingly seep into people's life and work scenarios and result in computing power needs on a massive scale. For the same reason, we believe there will be tremendous demand for highly elastic and cost-effective computing power services among enterprises and individual users.

Our Products and Services



We have built a cloud computing platform structured across three seamlessly integrated layers, as illustrated in the diagram above. At its foundation is our global computing power network, or the Infrastructure-as-a-Service (IaaS) layer, which unifies fragmented and heterogeneous computing resources worldwide into a highly efficient, scalable and standardized compute base. Above this, our MaaS (Model-as-a-Service) layer then provides access to ready-to-use models, enhanced by proprietary inference optimization technologies that significantly improve their performance beyond out-of-the-box capabilities. As AI progresses to the agentic era, we are further expanding into an agent infrastructure layer, laying the groundwork for AI agents capable of autonomously executing high-frequency tasks. Together, these layers form a cohesive platform designed to power the next generation of AI innovation at scale.

Specifically, we provide the following offerings under our edge cloud computing services and AI cloud computing services segments:

- *Edge cloud computing services*. Our edge cloud computing services integrate dispersed and heterogeneous computing resources from third-party providers to meet our customers' needs for edge computing and transmission. Instead of transmitting data to a centralized cloud data center, our edge cloud computing services allow the computation to be processed at the "edge" of the internet, near where the data are generated and used. This capability is essential for customers who process huge volumes of data at high speed in real time (for example, audiovisual content providers), where low latency is critical for good user experience. Specifically, we provide the following services:

  - *Edge node services*. Leveraging our widely distributed network of edge compute nodes, edge node services relocate computing power from central nodes to edge nodes closer to users, aimed at supporting scenarios such as edge storage, audio and video processing, video transcoding and distribution. Specifically, our offerings include edge containers and edge bare metals. Edge containers are virtualized units deployed at the edge that allow applications to run efficiently at or near where data are generated, reducing latency and improving performance for real-time applications. Edge bare metals refer to physical, standalone servers deployed at the edge that provide dedicated hardware resources for computing.

- *Edge content delivery network (Edge CDN)*. Our edge CDN is built on an expansive network of edge nodes and draws on our expertise in intelligent scheduling, distributed caching and network transmission technologies. The primary application scenarios of our edge CDN include file download, video-on-demand and live streaming. Unlike traditional CDNs, our edge CDN is more deeply integrated into local networks, placing servers closer to end users. This proximity improves transmission quality, providing users with smoother video playback and faster content delivery experiences.

• *AI cloud computing services*. Leveraging the distributed GPU computing resources integrated within our computing power network, we launched AI cloud computing services in 2023. As AI models continue to grow in scale and complexity, our comprehensive product portfolio addresses a wide range of demands in the field of AI inference. This includes providing elastic GPU computing resources required by large enterprises, as well as diverse model services tailored for small- and medium-sized enterprises (SMEs) and developers. Our services are designed to overcome the cost and technical barriers faced by clients worldwide in adopting new technologies. Specifically, we offer the following products with basic to advanced functions:

  - *GPU cloud services*. Our GPU cloud services provide customers with access to powerful computing resources without the need to purchase physical GPUs or manage GPU servers. The architecture is highly elastic, automatically scaling in or out based on demand, and operates on a pay-as-you-go model, where customers are charged only for actual usage. Our services go beyond merely consolidating GPU resources. We deliver a robust operating environment specifically tailored for the unique needs of generative AI and machine learning workloads. Key features of our GPU cloud services include an optimized inference engine, pre-built templates for rapid deployment and automated scaling driven by demand. These features establish our GPU cloud services as a comprehensive platform for AI inference, offering efficiency, flexibility and reliability for developing and scaling AI applications.

  - *Model API*. Our model APIs are gateways for our customers to access pre-built, ready-to-use models (which are mainly large language models (LLMs) and multimodal models) that cater to customers who lack the capability to build, train or manage highly complicated AI models themselves. We provide two types of services: first, our model APIs provide direct access to leading open-source models, allowing our customers to use such open-source models directly and conveniently; second, for customers who prefer to use their self-developed models, we host such models on our infrastructure. This allows our customers to retain control over their data and model behaviors, while benefiting from the scalability, efficiency and reliability of our infrastructure.

    A defining feature of our model API services is that we do not simply provide a pass-through gateway to underlying models. Our platform incorporates a suite of proprietary inference optimization technologies that materially enhance the performance of hosted models beyond their out-of-the-box capabilities. This enables customers to benefit from model services that are significantly faster, more efficient and more cost-effective than those available through standard model deployments. For example, based on our evaluation in 2026, the performance of a well-known large model delivered through our model API services substantially exceeded that of the same model deployed using SGLang (a benchmark model inference engine), reducing first-token latency by 95% and increasing total throughput efficiency by 300%.

- *Agentic Infra*. AI agents are AI systems that autonomously complete tasks by reasoning, planning and taking actions, without constant human oversight. Agentic infra comprises a suite of infrastructure designed to build a cloud environment in which AI agents can operate autonomously, reflecting a future in which AI agents, rather than humans, will be the primary users of cloud services. This comprehensive technical system spans AI coding, sandboxed execution environments, a broad range of tools and skills, and both short-term and long-term memory systems.

COMPETITIVE STRENGTHS

We believe that the following competitive strengths have played a key role in our success and differentiate us from our competitors.

China's leading independent distributed cloud computing service provider

We are a leader, pioneer and innovator in the distributed cloud computing services industry:

- *Leader*. We are a leading independent distributed cloud computing service provider in China. According to CIC, we operate the largest computing power network in China in terms of the number of compute nodes as of December 31, 2025, and we are the largest independent edge cloud computing service provider in China in terms of revenue in 2025. We also rank first among independent AI cloud computing service providers in China in terms of average daily token consumption volume in 2025.

- *Pioneer*. Our founding team comprises a group of pioneers with over 20 years of experience in distributed systems. Recognizing significant mismatches between massive demand and scattered, often underutilized supply, we were one of the earliest in the industry to establish a computing power scheduling platform to aggregate and manage idle computing power from heterogeneous sources and deploy it in an efficient fashion to customers to accelerate content delivery and enhance low-latency computing experience.

- *Innovator*. We have innovated our platform and services in response to the growth of large-scale AI models in the generative AI era. This evolution has driven the demand for GPU computing power and AI inference capabilities. We launched our AI cloud computing services by integrating AI capabilities to our computing power network. Our AI cloud computing services equip our customers with an AI computing power platform and a suite of AI inference tools essential for the generative AI era, such as the high-performance inference engine, distributed cloud platform and model APIs. We are among the earliest players in the Chinese distributed cloud computing market to offer AI cloud computing services.

The dawn of the generative AI era has driven significant demand for computing power, and the need for low-cost, highly flexible AI inference computing power is steadily increasing among enterprise users and individual developers. Based on our experience in the distributed cloud computing industry and our strategic positioning in the AI cloud computing field, we believe we have a first-mover advantage to capture the vast market opportunities brought by AI.

Efficient and highly scalable business model that facilitates organic growth

Our computing power network is central to our operations and connects various stakeholders within our business ecosystem, including computing power providers, users and AI developers. We benefit from this efficient and highly scalable platform-based business model which facilitates our organic growth:

- *Organic growth*. Our platform aggregates global computing resources to empower users and AI developers worldwide, effectively addressing mismatches in the computing power market. This drove the rapid growth for our platform during the Track Record Period. The number of compute nodes in our computing power network increased from over 2,800 as of December 31, 2023 to over 4,000 as of December 31, 2024 and further

to over 4,600 as of December 31, 2025. In line with our business growth, our revenue grew rapidly from RMB358.4 million in 2023 to RMB558.0 million in 2024 and further to RMB770.3 million in 2025, representing a CAGR of 46.6% from 2023 to 2025.

- *High efficiency*. We employ an asset-light business model with minimal capital expenditure, which affords us the flexibility to channel more resources into research and development activities and enhance our capacity to innovate and adapt swiftly to market changes while consistently improving profitability.

- *High scalability.* Our distributed cloud computing services platform accommodates rapid and seamless scalability, enabling efficient replication across regions. By leveraging our interconnected computing power network, we rapidly expanded our business from certain regions to almost all regions in China. This scalability optimizes resource utilization while catering to the increasing global demand for computing power. Our capability to replicate and expand swiftly contributes to a resilient ecosystem that accommodates the growing needs of AI developers and end-users.

Strong technological foundation and robust R&D capability

In the ever-evolving landscape of cloud computing services, we have carefully developed our core technologies to enhance flexibility, scalability and accessibility. With a commitment to compatibility and independence, our distributed cloud computing platform offers advantages in providing cutting-edge solutions tailored to the adaptive and diverse demands in the generative AI era. We possess the following core technologies:

- *Proprietary distributed computing architecture*. At the heart of our operations is a robust distributed computing power scheduling platform that seamlessly manages our expansive compute nodes network. Our system integrates diverse computing resources into a unified resource pool. Through intelligent workload scheduling, our architecture reduces the heterogeneity of underlying hardware, enabling seamless integration of diverse compute resources while serving as a standardized access point for upper-layer applications. In addition, based on a lightweight cloud-native architecture, our system efficiently organizes fragmented computing resources along two complementary dimensions: an AI computing layer, which handles AI inference and high-performance computing workloads; and an edge computing layer, which drives latency-sensitive services through a distributed edge node network. This design allows workloads to be intelligently routed to the optimal resource tier based on performance requirements, latency constraints and cost objectives, while significantly reducing system overhead compared to traditional solutions. Further, using machine learning techniques, our automatic fault-diagnosis system predicts potential node failures and dynamically adjusts task redundancy in real time.

- *Computing power network and scheduling capabilities*. We have established a distributed computing network, where the high-density compute node deployment enables millisecond-level latency (as short as 10 milliseconds) for local distribution, providing infrastructural support for scenarios such as real-time audiovisual content delivery and AI inference. Our global scheduling network enables us to achieve high utilization of computing resources on a continuous, around-the-clock basis. By leveraging cross-regional time-zone differences and the staggered demand profiles of our diverse customer base, we dynamically allocate globally idle computing capacity to wherever demand is high at any given moment, ensuring that our GPU resources are productively deployed rather than sitting idle during off-peak periods. As a result, our average GPU utilization rate is consistently maintained at above 75% in 2025, significantly outperforming the industry average of between 40% and 50%, according to

CIC. This intelligent "peak-shaving and valley-filling" capability materially improves overall resource utilization across our network, directly reducing the effective cost per token delivered to our customers.

- *Distributed inference acceleration for large models.* Our platform incorporates a suite of proprietary inference optimization technologies that materially enhance the performance of hosted models beyond their out-of-the-box capabilities. This enables our customers to benefit from model services that are significantly faster, more efficient and more cost-effective. Our optimization capabilities span multiple dimensions, such as operator optimization, inference engine optimization, storage system optimization and model optimization tailored for cutting-edge hardware. Through these multi-layered optimizations, we are able to drive computing resource utilization to near its theoretical limits, substantially reducing costs. For example, in 2026, the input token pricing of a well-known large model delivered through our model API services is more than 40% lower than the average charged by major international cloud providers.

- *Integration of heterogeneous computing resources*. Unlike the capital-intensive model of traditional cloud computing providers, which rely on centrally owned, homogeneous computing resources, we employ proprietary technology that enables us to orchestrate heterogeneous computing resources (*e.g.*, GPUs of different brands, generations and locations) into a unified computing resource pool with standardized access for customers. This mix of computing resources provides materially superior cost efficiency compared to the high-end, homogeneous computing resources on which traditional providers depend, while still delivering sufficient service quality and reliability. By integrating these diverse, cost-effective resources at scale, we substantially reduce our baseline per-unit computing cost, which directly supports improved profitability.

Our research has led to significant technical advances in critical facets of distributed cloud computing, such as workload forecasting and fault awareness within heterogeneous edge environments, as well as intelligent resource scheduling and context-aware task dispatching. In collaboration with leading research institutions, we have published or presented 18 high-impact papers in top-ranked journals or academic conferences, such as the IEEE INFOCOM (Institute of Electrical and Electronics Engineers—International Conference on Computer Communications), a leading conference that focuses on key topics and issues related to computer communications.

Vibrant community that fosters growth

Based on our computing power scheduling platform and computing power network, we have created a community that effectively bridges idle computing resources with computing demand. Surrounding our platform is a vibrant community consisting of computing power users, AI developers, computing power suppliers and other business partners, such as research institutions and industry organizations.

For our edge cloud computing services, we effectively connect idle computing resources from our suppliers with computing demand from our customers. For suppliers of computing power, our platform addresses a persistent challenge in the industry: the high idle rate of computing resources. According to CIC, the average utilization rate of IDCs globally is only around 50% to 70%, and may be lower for other types of computing power suppliers, leaving significant capacity untapped. Our service is essential in unlocking this potential, enabling suppliers to monetize their idle resources and generate new revenue streams with minimal operational adjustments. For our customers, we provide cost-effective, low-latency and highly elastic cloud computing solutions tailored to meet a diverse range of needs. By using our platform, customers save on the substantial costs and complexities associated with building and maintaining their own data centers. In addition, elasticity of our services accommodates real-time resource adjustment in response to fluctuating demand, thereby enabling our customers to only pay for the actual usage of the computing power and optimizing resource utilization.

For our AI cloud computing services, we collaborate with global open-source communities (such as GitHub and OpenRouter) with extensive developer networks. We believe that open-source communities are the primary engine driving the most advanced progress in the AI industry, and that maintaining deep, active engagement with these communities is essential to staying at the forefront of AI technology. By integrating our products into partner ecosystems and pursuing joint go-to-market initiatives, we expect to accelerate new customer acquisition, broaden our industry coverage and increase revenue per customer. We ranked first on OpenRouter among third-party open-source large model providers in January 2026 and ranked first on Hugging Face by monthly request volume of inference services in December 2025. We also implement developer-friendly strategies to attract new customers at a relatively low acquisition cost. We actively participate in open-source community events, provide open-source integrations, and host online and in-person technical workshops and developer forums. In 2025, our AI cloud computing services had over 180,000 active users.

Furthermore, we have engaged in collaborative research or strategic alliance with various academic institutions and industry organizations, enhancing the diversity and vibrancy of our community. For example, we, along with several leading research organizations and enterprises, formed the Future Computing Power Network Collaborative Scheduling Alliance. This alliance is designed to be a collaborative development platform for Chinese industry players, aiming to build an extensive network with high computing power, transmission capacity and cost efficiency.

Commitment to environmental sustainability and social responsibility

AI computing is highly energy intensive. As AI workloads grow in complexity and scale, the associated energy consumption increases correspondingly, contributing to higher operational costs and environmental impact. Since our inception, we have embraced sustainability as a core principle within our business model and ethos. Our business is intrinsically anchored on energy conservation, carbon footprint minimization and fulfilment of social responsibilities.

Traditional IDCs often struggle with low energy utilization efficiency. In pursuit of economies of scale, many IDCs maintain a large number of servers under a single compute node, which results in substantial energy consumption for server cooling. Inefficiently operated IDCs can devote more than 30% of their total energy consumption for server cooling, leading to high PUE values. In contrast, our distributed platform can connect smaller compute nodes and those strategically located across diverse environments. This approach helps to reduce the reliance on higher PUE compute resources and ultimately increase overall platform energy efficiency.

In addition, our distributed cloud computing services contribute to environmental sustainability by optimized computing power scheduling. For example, we can allocate latency-insensitive computing tasks—those that do not require immediate proximity to end users—to compute nodes in high latitude regions (such as our compute nodes in Iceland and Norway), where the naturally lower temperatures reduce the energy required for server cooling. In addition, we can also direct tasks to computing resources powered by renewable energy, thereby further enhancing our environmental efficiency.

Meanwhile, our distributed cloud computing network leverages existing, underutilized compute nodes, and therefore reduces the need to build additional massive, centralized data centers. Our technology also effectively extends the useful life of computing equipment and reduces the need for constantly replacing or expanding centralized facilities. This reduces e-waste and the carbon footprint associated with production and disposal of computing equipment.

We also demonstrate our commitment to social responsibility through our support for local communities. The presence of data centers in low-tier cities contributes significantly to the local economy, serving as a catalyst for technological development and infrastructure improvement. By maximizing the computing power utilization of these local data centers, our network not only bolsters the productivity of the region but also fosters community growth.

Environmental sustainability and social responsibility are integral to our corporate values. We consistently aim for excellence in environmental stewardship and remain committed to our social responsibilities, striving to be a leader in the cloud computing industry devoted to creating a better future.

Experienced management team with extensive industry background

We have an experienced management team with solid industry background and entrepreneurial experience. Our core management team collectively bring over 20 years of professional expertise, coupled with strong technical abilities and commercialization skills.

Our CEO, Mr. Yao Xin, is a serial entrepreneur and investor with rich experience in building global-scale technology platforms and driving frontier tech investments. As the founder of PPTV, he pioneered the world's first P2P-streaming protocol in 2004, scaling it into China's premier internet TV platform. After exiting from PPTV, he became a venture partner at Lanchi Ventures, a leading technology-focused venture fund, where he focused on early-stage investments in the AI field globally. Mr. Yao's global vision drives our commercial and strategic expansion.

Our CTO, Mr. Wang Wenyu, is a highly accomplished tech serial entrepreneur. He co-founded and served as the CTO of PPTV, developing its distributed video streaming services and pioneering various P2P audiovisual streaming distribution technologies. During his second entrepreneurial venture, Mr. Wang co-founded and served as the CTO of Jidou Auto, where he designed the hardware and software-integrated intelligent cabin and China's first automotive smart operating system. Mr. Wang has published 15 papers on edge computing, artificial intelligence and large-scale model and holds 18 patents.

Our Chief Scientist, Professor Wang Xiaofei, is a leading expert in the fields of edge computing and AI systems. He has authored over 200 high-impact research papers, which have garnered more than 11,000 citations. Professor Wang has also received the Fred W. Ellersick Prize, a prestigious prize awarded by the IEEE Communications Society for outstanding papers. His seminal work, *Edge AI*, is among the earliest and most influential publications in the field.

Our Chair of Technology Committee, Professor Jin Hai, is a professor at Huazhong University of Science and Technology, and a pioneer in the field of distributed and parallel computing and cloud computing systems. Professor Jin serves as the Director of the Key Laboratory for Service Computing Technology and Systems of the Ministry of Education, and Director of China Computing NET, a nationwide digital infrastructure that enables the high-speed interconnection of large-scale computing power across China. Through their cutting-edge insights, Professor Wang and Professor Jin help steer us at the forefront of technological innovation, enhancing our ability to deliver superior solutions and sustain a competitive edge in this rapidly evolving field.

OUR STRATEGIES

We plan to implement the following strategies to further develop our business:

Continuously enhance our portfolio of services and solutions

As the rapid evolution of AI transforms industries and reshapes demands on cloud computing infrastructure, we remain committed to staying at the forefront of innovation. Our strategy is rooted in anticipating the future needs of AI technologies and we are committed to delivering low-latency, ready-to-use distributed cloud services that allow intelligent systems to operate seamlessly across a wide range of application scenarios. We will focus on the following perspectives to enhance our portfolio of services and solutions:

- *Scale up our AI cloud computing services leveraging the surge of agentic AI.* As agentic AI proliferates, the primary callers of cloud computing infrastructure will increasingly shift from human users to AI agents autonomously executing complex, multi-step tasks at massive scale. This paradigm shift will give rise to entirely new and sustained demand for cloud computing resources, characterized by high-frequency concurrency, persistent long-context memory and multi-agent workflow orchestration. Such demand from agentic AI for computing power is structurally different from, and incremental to,

existing AI inference workloads. We have established agentic infra as a new business sub-segment and will continue to expand its scale and regional coverage, improving speed, stability and reliability when many agents run at the same time. We will also introduce supporting features such as secure data storage, long-term memory, safe credential use and performance tracking, to provide one-stop agentic infra deployment and continuous optimization.

- *Steadily grow our edge cloud computing services*. We will continue to refine our edge transmission technologies to ensure more efficient and cost-effective service delivery. In addition, we will enhance our expertise in the autonomous operation of computing power network, which enables us to better serve application scenarios that require high-bandwidth and low-latency delivery, such as AI video generation and low-latency inference services.

- *Upgrade our multimodal API platform*. The advance of AI technologies requires seamless integration among textual, audio, visual and gestural inputs. Therefore, we aim to build a unified multimodal API platform that bridges models with real-world multimodal inputs and supports complex agentic workflows. Specifically, we plan to develop APIs for multimodal streaming, which supports continuous, low-latency input and output streaming covering audio, visual, gestural and 3D spatial data. These capabilities are crucial for constructing interactive and immersive AI systems.

Expand our global footprint

The global market for distributed cloud computing services has seen substantial growth in recent years. We are dedicated to seizing the opportunities arising from the ongoing expansion of overseas markets. Our strategy involves entering established and emerging markets to access untapped potential. Specifically, we intend to develop additional global compute nodes and collaborate with more computing power suppliers in these regions to bolster our resource-based competitive advantages.

In addition to our global expansion efforts, we are committed to significantly enhancing the capabilities of our local service teams across key regional markets. Our focus will be on providing customized services such as sales, deployment, maintenance and after-sales support, tailored to address the distinct and diverse needs of local clients. This initiative is designed to ensure our service offerings resonate deeply with each market's unique requirements and expectations.

By leveraging our comprehensive service portfolio and drawing upon our robust research and development capabilities, we will roll out state-of-the-art cloud computing solutions designed to closely align with the local customers' specific demands. Our approach not only seeks to improve user satisfaction but also strives to build long-lasting relationships rooted in understanding and responsiveness. Through this commitment, we will effectively position ourselves as the preferred partner in cloud computing services across a variety of international landscapes.

Continue to enhance our technology capabilities

We will continue to enhance our technology capabilities in the cloud computing services industry. We aim to address the growing complexity of AI applications that rely on ultra-responsive, efficient and scalable solutions. Our efforts will focus on the following key dimensions:

- *Enhance AI inference optimization and multimodal model capabilities.* We plan to further improve the performance, cost efficiency and scalability of our large model inference services and model APIs. We intend to optimize around cutting-edge hardware with higher compute power, larger high-speed memory and lower communication latency, in order to reduce per-token inference cost and enhance the overall return on our computing power investment while maintaining service quality and model performance.

- *Strengthen comprehensive agentic infra.* As agentic infra and multi-agent collaboration become more widely adopted, we plan to enhance our agentic infra so that customers' agent workloads can run on our platform in a stable, secure and scalable manner. We intend to build persistent data and state management capabilities across sandbox

instances so that agents can retain work content across different scenarios. We also plan to offer secure credential management, end-to-end encryption for data transmission and storage, automated backup and full data lifecycle management to deliver enterprise-grade availability, security and compliance for agentic data.

- *Enhance AI computing power network and scheduling.* We plan to further enhance our AI computing power network and scheduling capabilities to support ultra-low latency, large-scale and mixed-workload AI applications. We plan to introduce differentiated scheduling based on service level objective tiers by classifying requests according to latency and throughput requirements and dynamically routing them to different layers of AI computing resources, with a view to maximizing overall utilization without sacrificing per-request service quality. In parallel, we intend to further develop our ultra-low latency computing power network and cloud-edge collaboration so that edge devices can process local tasks while offloading complex workloads to the cloud.

Develop and strengthen our community

We place significant importance on our reputation and influence among global AI developers, as they represent a crucial component in the AI community. Through new product offerings, high-quality developer services and diverse growth strategies, we continually strengthen developer engagement, enhance commercial value and expand our business reach.

To accelerate our business growth, we have established dedicated teams for customer acquisition and developer relations, employing new approaches to drive expansion. We plan to continue boosting product visibility and brand influence through targeted digital marketing and social media strategies aimed at attracting potential customers. Additionally, we will maintain a strong presence within the community by hosting online and in-person events such as technical workshops and developer forums. We also aim to draw developers to our products by actively participating in open-source community events and offering open-source integrations.

At the same time, we will foster new partnerships with AI community platforms to integrate our products into their ecosystems and support joint sales initiatives. We also intend to collaborate with leading AI applications and prominent open-source platforms to diversify our customer base and increase brand recognition further.

On the supply side, we are committed to strengthening our computing power network by diversifying suppliers and expanding global collaborations. To meet the demands of diverse computing tasks, we will work with a broader range of computing resource providers and their ecosystems. In addition, we will partner with global telecom operators, data centers, intelligent computing enterprises and traditional cloud service providers. Using flexible pricing models such as on-demand, spot pricing and revenue-sharing, we will accommodate a variety of collaboration needs and business scales.

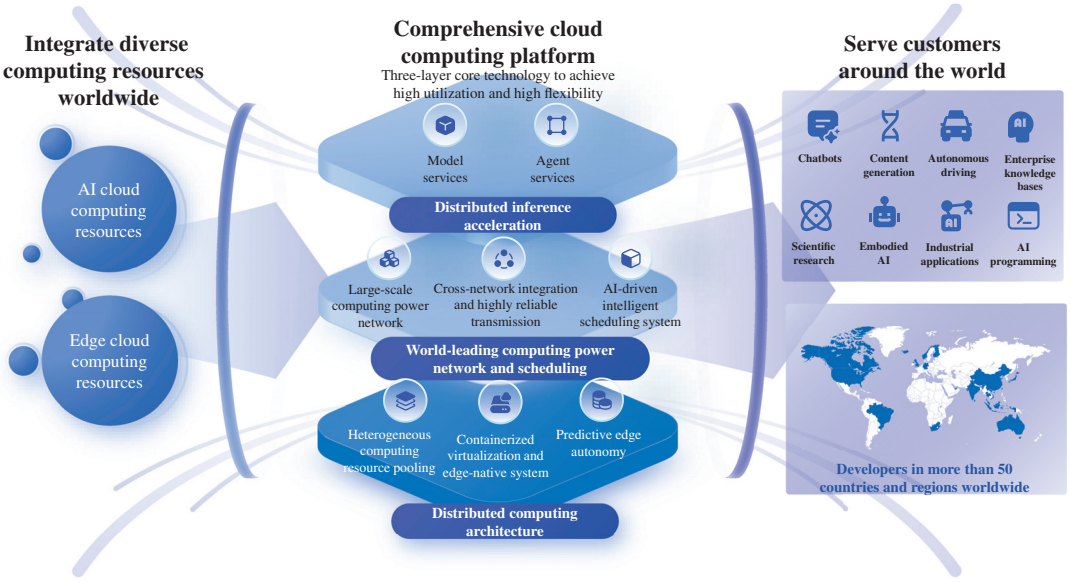
Our strong technical compatibility and flexible cooperation models position us to attract an increasing number of global computing power suppliers. This helps us enhance our infrastructure and allows us to deliver competitively priced computing resources to our clients. We believe these efforts will further solidify our position as a leader in the distributed cloud computing services market.

Explore strategic alliance, investment and acquisition opportunities

We plan to selectively pursue strategic alliances, investments and acquisitions to reinforce our technological capabilities and bolster our market position. With an eye on targeted expansion and innovation, our approach involves thoroughly exploring and evaluating opportunities that offer complementary or synergistic benefits aligned with our offerings. This pursuit is intended to bolster our competitive strengths and advance our leadership in the distributed cloud computing services industry. We believe these strategic initiatives may accelerate our capacity in delivering cutting-edge solutions and maintaining our industry leadership. As of the Latest Practicable Date, we had not identified any potential target company for acquisition or investment.

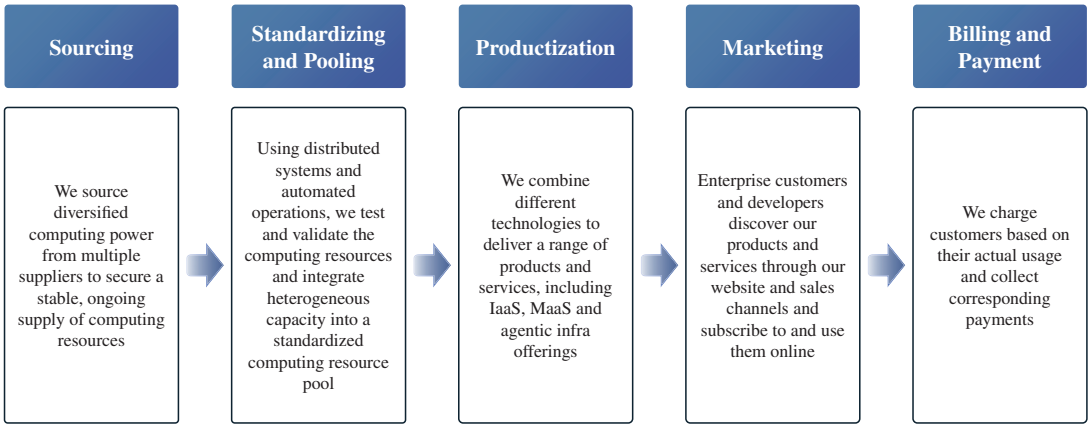
OUR SERVICES AND SOLUTIONS

Our Business Model



We have established a comprehensive cloud computing platform that bridges global computing supply with AI-driven demand through three integrated components. On the supply side, we aggregate two categories of computing resources from around the world: AI cloud computing resources, comprising distributed GPU capacity suited to inference and high-performance computing workloads; and edge cloud computing resources, drawn from a large-scale network of geographically dispersed edge nodes. These heterogeneous resources are channeled into our cloud computing platform, which operates across three technology layers: a proprietary distributed computing architecture that achieves standardized access to and unified scheduling of diverse hardware; an outstanding computing power network and scheduling layer that maximizes resource utilization through intelligent peak-shaving and valley-filling; and a distributed inference acceleration layer that optimizes model performance based on heterogeneous computing resources. On the demand side, our platform delivers services to developers across more than 50 countries and regions and supports a broad range of application scenarios for enterprise customers. Taken together, we create value by efficiently orchestrating globally dispersed, heterogeneous computing resources and converting them into reliable, low-cost and highly scalable cloud computing services for a growing global base of customers.

The following diagram sets forth our key operating work flows:



Specifically, our services include edge cloud computing services and AI cloud computing services:

- *Edge cloud computing services.* Our edge cloud computing services integrate dispersed and heterogeneous computing resources from third-party providers to meet our customers' needs for edge computing and transmission. Instead of transmitting data to a centralized cloud data center, our edge cloud computing services allow the computation to be processed at the "edge" of the internet, near where the data are generated and used. This capability is essential for customers who process huge volumes of data at high speed in real time (for example, audiovisual content providers), where low latency is critical for good user experience. Specifically, we provide the following services:

  - *Edge node services*. Leveraging our widely distributed network of edge compute nodes, edge node services relocate computing power from central nodes to edge nodes closer to users, aimed at supporting scenarios such as edge storage, audio and video processing, and video transcoding and distribution. Specifically, our offerings include edge containers and edge bare metals. Edge containers are virtualized units deployed at the edge that allow applications to run efficiently at or near where data are generated, reducing latency and improving performance for real-time applications. Edge bare metals refer to physical, standalone servers deployed at the edge that provide dedicated hardware resources for computing.

  - *Edge content delivery network (Edge CDN).* Our edge CDN is built on an expansive network of edge nodes and draws on our expertise in intelligent scheduling, distributed caching and network transmission technologies. The primary application scenarios of our edge CDN include file download, video-on-demand and live streaming. Unlike traditional CDNs, our edge CDN is more deeply integrated into local networks, placing servers closer to end users. This proximity improves transmission quality, providing users with smoother video playback and faster content delivery experiences.

- *AI cloud computing services*. Leveraging the distributed GPU computing resources integrated within our computing power network, we launched AI cloud computing services in 2023. As AI models continue to grow in scale and complexity, our comprehensive product portfolio addresses a wide range of demands in the field of AI inference. This includes providing elastic GPU computing resources required by large enterprises, as well as diverse model services tailored for small- and medium-sized enterprises (SMEs) and developers. Our services are designed to overcome the cost and technical barriers faced by clients worldwide in adopting new technologies. Specifically, we offer the following products with basic to advanced functions:

  - *GPU cloud services*. Our GPU cloud services provide customers with access to powerful computing resources without the need to purchase physical GPUs or manage GPU servers. The architecture is highly elastic, automatically scaling in or out based on demand, and operates on a pay-as-you-go model, where customers are charged only for actual usage. Our services go beyond merely consolidating GPU resources. We deliver a robust, operating environment specifically tailored for the unique needs of generative AI and machine learning workloads. Key features of our GPU cloud services include an optimized inference engine, pre-built templates for rapid deployment, automated scaling driven by demand and fine-tuning tools. These features establish our GPU cloud services as a comprehensive platform for AI inference, offering efficiency, flexibility and reliability for developing and scaling AI applications.

- *Model API.* Our model APIs are gateways for our customers to access pre-built, ready-to-use models (which are mainly LLMs and multimodal models) that cater to customers who lack the capability to build, train or manage highly complicated AI models themselves. We provide two types of services: first, our model APIs provide direct access to leading open-source models, allowing our customers to use such open-source models directly and conveniently; second, for customers who prefer to use their self-developed models, we host such models on our infrastructure. This allows our customers to retain control over their data and model behaviors, while benefiting from the scalability, efficiency and reliability of our infrastructure.

- *Agentic Infra*. AI agents are AI systems that autonomously complete tasks by reasoning, planning and taking actions, without constant human oversight. Agentic infra comprises a suite of infrastructure designed to build a cloud environment in which AI agents can operate autonomously, reflecting a future in which AI agents, rather than humans, will be the primary users of cloud services. This comprehensive technical system spans AI coding, sandboxed execution environments, a broad range of tools and skills, and both short-term and long-term memory systems.

The following table sets forth a breakdown of our revenue by product/solution category during the Track Record Period.

	Year ended December 31,					
	2023		2024		2025	
	Amount	%	Amount	%	Amount	%
	(RMB in thousands, except for percentages)					
Edge cloud computing services	358,120	99.9	547,650	98.1	651,128	84.5
– Edge node services	313,235	87.4	390,951	70.0	502,645	65.2
– Edge CDN	44,885	12.5	156,699	28.1	148,483	19.3
AI cloud computing services	265	0.1	10,387	1.9	119,217	15.5
Total	**358,385**	**100.0**	**558,037**	**100.0**	**770,345**	**100.0**

Edge Cloud Computing Services

Instead of transmitting data to a centralized cloud data center, our edge cloud computing services allow the computation to be processed at the "edge" of the internet, near where the data are generated and used. This capability is essential for customers that process huge volumes of data at high speed in real time (for example, audiovisual content providers), where low latency is critical for good user experience.

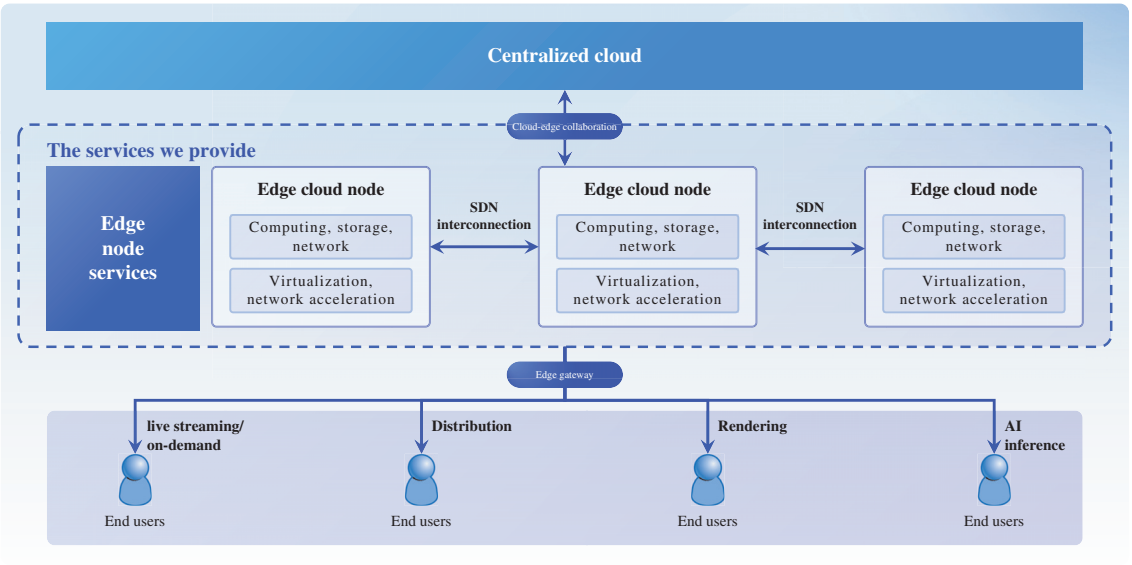
Edge Node Services

Leveraging our widely distributed network of edge compute nodes, edge node services shift computing power from central nodes to edge nodes closer to users, aimed at supporting scenarios such as edge storage, video-on-demand, live streaming, and video transcoding and distribution. Specifically, our products include edge containers and edge bare metals:

- *Edge container*. "Containers" are virtualized units that allow developers to package their applications into a single portable unit. Edge containers are functionally similar to traditional containers but are deployed on edge devices or nodes. They allow applications to run efficiently in the online edge compute nodes which are at or near where data are generated, reducing latency and improving performance for real-time applications.

- *Edge bare metal*. Edge bare metals refer to physical, standalone servers deployed at the edge that provide dedicated hardware resources for computing. Unlike virtualized environments where multiple users or processes share the same physical machine, edge bare metals deliver direct and unshared access to the infrastructure, offering high performance, low resource overhead and precise control.

The following diagram is a simplified illustration of our edge node services:



Amid the rapid development of audiovisual technologies and the expanding range of application scenarios, the demand for seamless, high-definition and low-latency media experiences continues to grow. Businesses now require computing resources that are more reliable and efficient. By leveraging cloud-edge integrated heterogeneous resource management technologies, our edge node services provide edge containers and edge bare metals with core characteristics such as secure resource isolation, high resource utilization and elastic scaling. Key features and advantages of our edge node services include:

- *Heterogeneous resource management*. We use advanced heterogeneous resource management technologies to provide direct access to physical edge bare metals. This eliminates the performance loss typically caused by virtualization, allowing optimal hardware utilization. Our services cater to compute-intensive workloads, ensuring maximum computational performance while maintaining user accessibility.

- *Edge container management*. Our edge container management platform supports process, network and file system level environment isolation, dynamic resource scheduling, seconds-level elasticity (*i.e.*, the system can automatically scale up or down within seconds to accommodate fluctuations in demand for computing resources) and rapid business deployment.

- *Comprehensive intelligent monitoring system*. A robust 24/7 monitoring framework underpins our edge node services, employing machine learning for fault prediction and anomaly alerts. This ensures proactive issue detection and resolution, safeguarding operational reliability.

Use Case: Edge Node Services for a Leading Personal Cloud Storage Platform

Our client is a leading personal cloud storage service provider in China with a vast number of active users. As user-generated content grew rapidly, the platform needed to process petabytes of unstructured data daily.

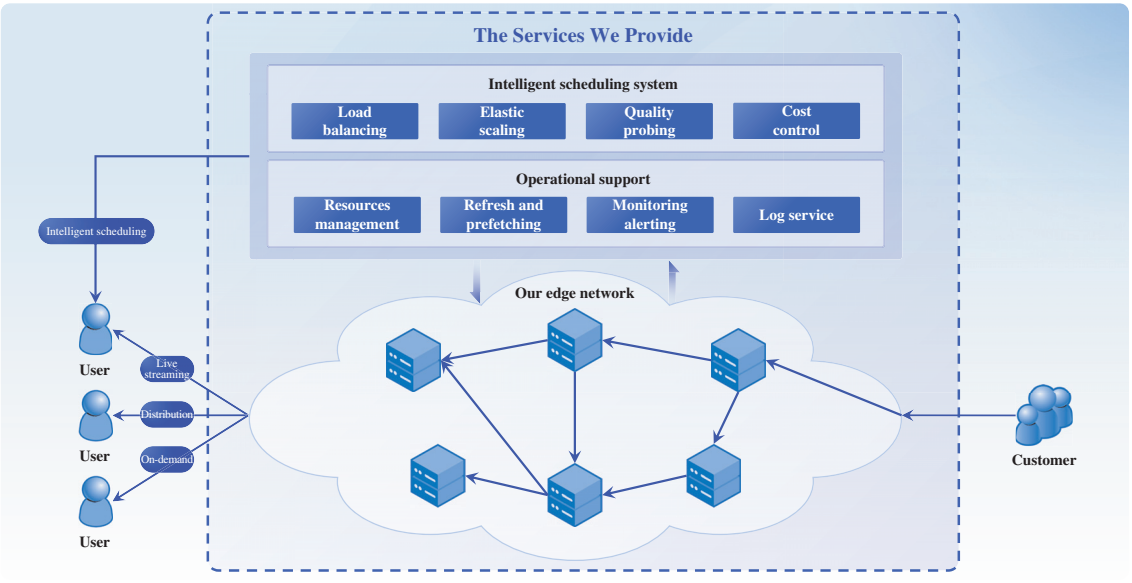
- *Our client's challenges.* Our client faced significant performance and cost pressures. During peak hours, such as mornings and evenings, heavy access to popular files frequently caused download speeds to drop, especially for large file transfers, leading to a frustrating user experience. The centralized storage architecture also drove up bandwidth costs, which accounted for over 35% of operational expenses.

- *Our solution.* To address these issues, we delivered an intelligent storage solution powered by edge node services: (a) we built a distributed caching network across edge nodes nationwide. Caching is a technique to temporarily store frequently accessed data in a location that is faster to retrieve than the original source. We developed intelligent algorithms to store popular content closer to users and to increase download speed; (b) the distributed caching network offers greater storage capacity and enhances cache hit rates. Cache hit rate is a measure of how effectively a cache is working. It represents the proportion of times that requested data is found in the cache (a "cache hit") compared to the total number of data requests made. By enhancing cache hit rates, we are able to increase download speeds and lower computing power resource costs; and (c) the application of acceleration technology optimizes download speeds in weak network environments, further improving the user experience.

  Our solution delivered significant benefits to the client. It improved the experience for hundreds of millions of monthly active users by enabling faster and seamless downloads, even under weak network conditions. The optimized bandwidth usage reduced operational costs by nearly 30%. Furthermore, strengthened data security and high service availability help our client improve user trust.

Edge CDN

Our edge CDN comprises an expansive network of edge nodes and draws on our expertise in intelligent scheduling, distributed caching and network transmission technologies. The primary application scenarios of our edge CDN include file downloads, video-on-demand and live streaming. Unlike traditional CDNs, our edge CDN is more deeply integrated into local networks, placing servers closer to end users. This proximity improves transmission quality, providing users with smoother video playback and faster content delivery experiences. The following diagram is a simplified illustration of our edge CDN:



Our edge CDN has the following key features and benefits:

- *Intelligent scheduling*. Our intelligent scheduling mechanism allocates and dispatches compute nodes at millisecond-level speed. It manages the operating status and file cache distribution of tens of thousands of devices and can determine the node closest to the user in real time. If a node does not function normally, our system can swiftly transfer the workload to another node nearby, ensuring that there is no disturbance to our users' requests.

- *Elastic computing*. Leveraging the "containerization" technique, our node deployment is rapid and highly flexible, allowing adjustment of distribution in real time. This setup ensures high scalability of services to cater to scenarios where, for example, e-commerce platforms encounter high concurrent requests during nationwide promotional events.

- *Cost control algorithms.* Our cost control algorithms operate based on flexible node management and scheduling systems. They predict bandwidth redundancy by analyzing historical trends. The algorithms seek to maximize the utilization of network capacity and minimize the idleness of computing power.

Use Case: Edge CDN Services for a Well-Known Chinese Short Video Platform

Our client is a well-known Chinese short video platform that is experiencing rapid growth. As viewer traffic increases, the platform faces challenges associated with unexpected usage surges and ineffective content distribution.

- *Our client's challenges.* Our client's platform encountered regional traffic surges during celebrity live streams, where large-scale live events caused concentrated traffic in specific areas, resulting in network congestion. In addition, users in remote, low-tier cities and counties experienced long playback buffering time, which negatively affected user experience. Traditional CDN architecture restricted the platform's ability to balance user experience and cost efficiency, resulting in difficulties managing peak traffic while facing high operational expenses.

- *Our solution*. Our edge CDN services implement intelligent computing power scheduling to establish dynamic load balancing mechanisms. This approach automatically detects imbalances in workload among different compute nodes and reroutes traffic in real time to optimize network performance.

  In addition, our edge nodes are also deployed in low-tier cities and counties, which helps optimize content transmission in the "last mile" and enhances overall content delivery efficiency. Our platform also enables bandwidth resource reuse, which preloads popular content using off-peak network resources, allowing the creation of a flexible resource pool that improves content accessibility.

AI Cloud Computing Services

Our AI cloud computing services provide cost-effective computing power for AI computation. Through our distributed computing power network, we consolidate GPU resources around the world. This enables us to offer computing resources for AI fine-tuning and large-scale AI inference tasks using a flexible, pay-as-you-go pricing model. Additionally, by integrating a broad array of well-recognized open-source LLMs and multimodal models, and further optimizing the inference process, we deliver API-based services to enterprises and AI developers worldwide. This approach reduces the barriers and costs associated with adopting AI technologies.

GPU Cloud Services

Graphics processing units (GPUs) excel in handling parallel processing tasks that require extensive computing power. Using our GPU cloud services, our customers have access to powerful computing resources but do not need to purchase physical GPUs or manage GPU servers themselves. Our architecture offers high elastic scalability, which can automatically scale in or out based on demand, and our customers only pay for the actual usage of the computing power. Also, our services are not a simple consolidation of GPU resources. We optimize the operational environment for AI model inference so that the inference tasks can be processed faster and more efficiently. In addition, our customers can easily deploy models on our platform.

Our GPU cloud services are primarily used in two types of application scenarios: AI fine-tuning and large-scale AI inference. For AI fine-tuning, we offer customers access to a wide selection of GPU resources through a pay-as-you-go pricing model. For large-scale AI inference, we provide highly scalable and distributed GPU resources, as well as access to a diverse range of AI models.

Our GPU cloud services have the following key features and advantages:

- *Elastic scalability*. Our architecture can automatically scale in or out based on demand, and our customers only pay for the actual usage of the computing power. This feature is critical for customers who require rapid scaling during peak demand, such as e-commerce enterprises.

- *Optimization of operating environment*. Our GPU cloud services are not a simple consolidation of GPU resources. We optimize the operational environment for AI model inference so that the inference tasks can be processed faster and more efficiently.

- *Pay-as-you-go*. Our GPU cloud services adopt a pay-as-you-go pricing mechanism. This mechanism adeptly addresses our customers' needs for flexible resource allocation during critical demand periods by offering elastic pricing strategies aligned with actual usage. It provides a cost-effective solution for our customers, where expenditures correspond to resources actually utilized.

- *High-speed delivery for large-scale models*. By offering advanced network capabilities for massive delivery, our GPU cloud services substantially reduce waiting times, allowing users to access large models and datasets rapidly. This improves user's productivity and efficiency, making it easier to handle complex tasks in a limited timeframe.

- *Containerized solutions*. This method ensures the effective isolation of computing environments, thereby preventing competition for resources or conflicting requirements for dependencies. Our GPU cloud services also offer a user-friendly operational interface, thereby leading to increased productivity through streamlined workflows.

- *Shared storage*. This solution offers high-performance shared cloud storage facilities that facilitate the seamless sharing of data across multiple GPU cloud services, ensuring consistent access for all relevant users. Such functionality is pivotal in bolstering collaborative efforts and fostering efficient coordination and data management practices.

Use Case: GPU Cloud Computing Services for a Leading Automotive Software Company

Our client is a prominent provider of intelligent cockpit software solutions, serving global mid-to-high-end automotive manufacturers with comprehensive intelligent vehicle software and digital service solutions.

- *Our client's challenges*. The client's business required fine-tuning large models and they initially relied on closed-source models. However, this approach carried high costs as early-stage closed-source models were limited in version iteration. Although the client later switched to open-source models, the GPU rental costs during the fine-tuning process were still significant, with compute expenses accounting for a substantial portion of the budget. Securing long-term, stable supply for GPU resources was also a concern.

  Once deployed, the fine-tuned models faced challenges in handling inference services. Traffic fluctuated sharply between peak and off-peak hours, with sudden surges creating significant resource allocation challenges and making it harder to provide a seamless user experience.

- *Our solution*. (i) After carefully comparing open-source models with closed-source models in terms of cost-efficiency and performance, we assisted the client to transition to the open-source models, reducing their costs and stabilizing training data outputs. We implemented a unified fine-tuning and inference toolkit for LLMs, enabling the client to efficiently validate the effectiveness of fine-tuning using a single instance that supported both base models and fine-tuned models;

  (ii) we identified high-performance GPUs tailored to the client's needs and provided reliable GPU computing resources to ensure model stability and continuous optimization; and

  (iii) drawing on our expertise in distributed computing and computing power scheduling, we offered our client the pay-as-you-go pricing, which allowed precise billing based on GPU usage. Resources automatically scaled up during traffic peaks, while cost-efficient computational resources were utilized during off-peak periods to reduce expenses.

Our solution delivered substantial benefits to the client. Transitioning to open-source models reduced training costs and improved data consistency. Optimized GPU supply ensured a steady, long-term supply, supporting stable model iteration and fine-tuning. The elastic computing infrastructure provided scalable resources during demand surges while efficiently minimizing operating costs during off-peak periods. This comprehensive approach not only enhanced service performance and reinforced user satisfaction but also strengthened the client's competitive standing in the automotive software industry.

Use Case: GPU Rendering Services for a Global Leader in Architectural Software

Our client is an industry-leading architectural software company with advanced real-time rendering algorithms. To boost design efficiency and enhance the creative experience, they required highly optimized real-time rendering capabilities that demanded high computing power scheduling and inference performance.

- *Our client's challenges*. The client encountered issues such as low scheduling service efficiency and high latency in inference services, which disrupted the efficiency and fluidity of real-time rendering. Additionally, reliance on public cloud networks for data transmission caused security concerns and also led to high costs.

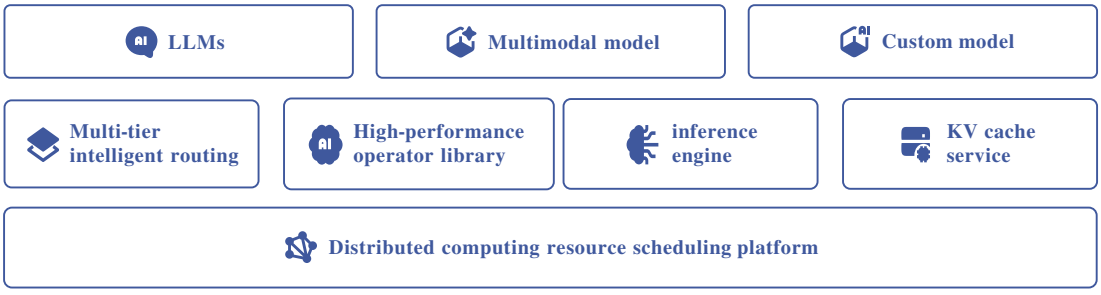
- *Our solution*. We delivered a comprehensive GPU/CPU containerized service that enabled seamless deployment and management of rendering workloads. By creating a secure virtual private cloud (VPC), we ensured fast and reliable internal data transmission between CPU and GPU instances, eliminating the risk of using public network directly, reducing latency and enhancing security. High-performance shared cloud storage allowed efficient multi-instance data access, streamlining workflows. To

ensure reliability, our 24/7 operations and maintenance support promptly addressed issues and kept the platform stable, helping the client meet their real-time rendering requirements with consistent fast-response performance.

Our solution significantly reduced network latency and optimized operational costs. By leveraging our support in AI infrastructure, the client was able to focus on enhancing their core rendering capabilities, thereby further strengthening their leadership in architectural software innovation.

Model API

Our model APIs are pre-built models that cater to customers who need AI technology but lack the capability to build, train and manage highly complicated AI models themselves. Our model APIs include two categories: large language models and multimodal models. Large language models handle various natural language processing tasks, whereas multimodal models process and generate images and videos. We provide two types of services: first, our model APIs provide direct access to leading open-source models, allowing our customers to utilize such open-source models directly and conveniently; second, for customers who prefer to use their self-developed models, we host such models on our infrastructure. This allows our customers to retain control over their data and model behaviors, while benefiting from the scalability, efficiency and reliability of our infrastructure. The following diagram is a simplified illustration of our model API:



Our model APIs have the following features and advantages:

- *Rapid integration*. Our model APIs offer access to an extensive range of the latest open-source models across various modalities. This streamlined approach allows businesses to deploy advanced AI tools directly, saving the time typically required for traditional model setup.

- *Low maintenance*. Our model APIs relieve our customers of the complexities of model deployment and operational maintenance. Our ready-to-use services are designed to handle the intricacies in maintaining the models, so that our customers can focus on achieving business goals without distractions.

- *Cost effectiveness*. Our model APIs adopt a pay-as-you-go pricing model, which aligns with the customer's actual usage. This pricing model is supported by a robust distributed elastic scheduling framework, which further leads to optimized performance and cost efficiency.

- *Diverse models*. Our platform hosts a dynamic variety of models, enabling effortless loading and switching within an expansive ecosystem. This flexibility removes the need for individual clients to manage models, providing an adaptable platform that evolves with customer needs.

Use Case: Integration of Model API with a Global Leading AI-Based Role-Playing Game Company

Our client is a global leading AI-based anime role-playing game developer. The client possesses a mid-sized R&D team, with solid capabilities in model optimization.

- *Our client's challenges*. Our client needed to maintain their own GPU servers, along with ensuring the stability of their network infrastructure. During times of increased traffic, user response times tended to become delayed.

- *Our solution*. Leveraging our ready-to-use model APIs, our client conveniently aligned them with their business needs and can further optimize the models from time to time in response to their evolving needs and preferences. By using our model API, we significantly reduced the client's infrastructure maintenance costs. Also, the pay-as-you-go pricing model helps the client lower overall expenses. Our model API also incorporates an automatic scaling mechanism, which targets to address the increased response time during peak traffic time. The unutilized GPU resources are repurposed for model refining and training, which further improves cost efficiency for the client.

Use Case: LLM API Services for a Global One-Stop AI Assistant Innovator

Our client is a leading artificial intelligence innovation company that provides a one-stop AI assistant offering features such as chat, search, collaboration and programming support. As AI technologies rapidly evolve, the client needed to stay ahead by integrating cutting-edge AI tools and functionalities into their product line.

- *Our client's challenges*. The client faced significant challenges in quickly validating the feasibility and effectiveness of new AI features to adapt their development strategies and avoid redundant resources. This required access to stable and reliable LLM APIs, but many available APIs lacked the requisite consistency, reliability and cost-effectiveness.

- *Our solution*. We provided a tailored solution leveraging our optimized distributed inference and inference acceleration capabilities designed for large-scale AI applications. Our high-reliability, scalable and cost-efficient LLM API services enabled the client to swiftly build and validate generative AI applications. Models backed by our services delivered exceptional performance in areas such as role-play, story expansion, code generation and empathetic interaction, combining high intelligence with emotional resonance. For programming support, our DeepSeek R1 API was particularly instrumental. Supporting function call capabilities, it empowered the client to efficiently test AI features and accelerate development cycles.

Agentic Infra

AI agents are AI systems that autonomously complete tasks by reasoning, planning and taking actions, without constant human oversight. Agentic infra comprises a suite of infrastructure designed to build a cloud environment in which AI agents can operate autonomously, reflecting a future in which AI agents, rather than humans, will be the primary users of cloud services. This comprehensive technical system spans AI coding, sandboxed execution environments, a broad range of tools and skills, and both short-term and long-term memory systems.

We provide two main types of capabilities. First, our agentic infra offers secure, isolated execution environments for AI-generated code, tool calls and browser operations. These sandbox environments prevent AI agents from accessing or tampering with resources outside the sandbox, so that agent behavior does not cause harm to customers' production systems. They support rapid start-up, pause and resume and cloning of running environments, which helps customers flexibly manage agent workloads that are long-running, intermittent or highly concurrent, and reduces resource costs during idle periods. Second, our agentic infra supports the hosting and orchestration of agents themselves. Through standard interfaces and developer tools, customers can deploy their own agents as services, manage their operation and scale their capacity based on demand, while being shielded from the complexity of the underlying infrastructure.

Our agentic infra has the following features and advantages:

- *Secure operating environment*. Hardware-level virtualization and sandboxing technologies create strong isolation boundaries for agent workloads, ensuring that AI-generated code and tool calls run within controlled environments, without affecting other services or systems.

- *Comprehensive management of operating status*. Native support for pause, resume, snapshot and cloning enables agents to preserve context over long horizons, sleep when inactive and resume on demand, improving efficiency for long-running or intermittent tasks.

- *Elastic scalability*. Our infrastructure can automatically scale agent execution environments up or down based on the volume and complexity of tasks, which is particularly important for scenarios with fluctuating traffic or large-scale concurrent agent interactions.

- *Developer-friendly experience*. Standard software development kits and deployment tools simplify the development and operation of agent-based applications, allowing customers to move quickly from algorithm experimentation to production deployment without rebuilding their environments.

Use Case: Agentic Infra for AI-Generated Coding

Our client is an AI agent developer whose AI-generated coding workloads run on our infrastructure.

- *Customer challenges*. In the AI-generated coding scenario, a sandbox environment often remains active for extended periods while awaiting model outputs or user input, even when it is performing little actual computation. When deploying its agents, the customer must also address the operational complexity of managing capacity and infrastructure expenditure at high levels of concurrency. The customer seeks to achieve elastic scaling and roll-out of agent services under a more lightweight operations, while leveraging our agentic infrastructure to improve efficiency and reduce cost.

- *Our solution*. Through our agent sandbox, we address the above challenges by providing the customer with an isolated runtime environment based on virtualization. The environment can be paused and resumed within seconds while it waits for model outputs or user confirmation, which materially reduces resource costs associated with long idle periods. The sandbox can also quickly create fresh environments from the customer's latest code, ensuring that each user works in an up-to-date, isolated workspace and mitigating the operational risks associated with shared environments. Together, these capabilities enable our agentic infrastructure to support the customer in running large-scale AI-generated coding workloads in a cost-efficient and operationally robust manner.

BUSINESS SUSTAINABILITY AND PATH TO PROFITABILITY

Our Historical Performance

During the Track Record Period, our revenue grew rapidly from RMB358.4 million in 2023 to RMB770.3 million in 2025, representing a CAGR of 46.6% from 2023 to 2025. In particular, our AI cloud computing services have experienced exponential growth since 2025, with revenue increasing by more than tenfold from RMB10.4 million in 2024 to RMB119.2 million in 2025. Our gross profit grew rapidly from RMB63.5 million in 2023 to RMB68.8 million in 2024 and further to RMB72.0 million in 2025, representing a CAGR of 6.5% from 2023 to 2025. Our gross profit margin decreased from 17.7% in 2023 to 12.3% in 2024 and further decreased to 9.4% in 2025.

While we achieved sustained business growth, we incurred a loss for the years of RMB189.4 million, RMB293.5 million and RMB222.9 million in 2023, 2024 and 2025, respectively, and an adjusted net loss (a non-HKFRS measure), which excludes fair value losses on convertible redeemable preferred shares and share-based payment expenses, of RMB37.1 million, RMB61.6 million and RMB104.6 million in the same years, respectively. As we launched our AI cloud computing services in 2023, this business segment was in the early stage of development during the Track Record Period and had not yet achieved profitability. For our edge cloud computing services business, we had maintained gross profit throughout the Track Record Period and we expect our investments in R&D and relevant infrastructures during the Track Record Period to drive sustainable growth of this business segment.

Our net losses during the Track Record Period were primarily attributable to the following factors:

Factors primarily related to our AI cloud computing services:

- *AI cloud computing services in ramp-up stage and has not yet reached economies of scale*. We launched our AI cloud services in 2023. This business segment was in the early stage of development during the Track Record Period and its revenue has not yet reached a scale that allows us to realize economies of scale. In 2023, 2024 and 2025, our revenue from AI cloud computing services was RMB0.3 million, RMB10.4 million, and RMB119.2 million, representing 0.1%, 1.9%, and 15.5% of our total revenue during the same period, respectively. Compared with our more mature edge cloud computing services, the number of compute nodes and paying customers for our AI cloud computing services was still ramping up. However, to ensure low latency, high availability and elastic capacity for our customers, we already needed to reserve a meaningful level of computing resources in advance, and to maintain a distributed infrastructure that can support large-scale inference workloads. These reserved resources and infrastructure costs were incurred regardless of short-term fluctuations in actual usage. As a result, our unit cost of revenue for AI cloud computing services remained relatively high, and our revenue was not yet sufficient to fully absorb these reserved capacity costs, which contributed to net losses for this business segment during the Track Record Period.

- *Significant investment for building AI cloud infrastructure.* We incurred substantial upfront investment to establish our AI cloud computing services. The number of compute nodes under our AI cloud computing services increased from 19 and 64 as of December 31, 2024 and 2025, respectively, to 77 as of April 30, 2026. Our upfront investment included procuring and integrating distributed computing resources, building and configuring servers and related hardware and laying the foundational architecture needed to support large-scale AI inference workloads. In line with the addition of our compute nodes, the computing resource costs for our AI cloud computing services grew rapidly. In 2023, 2024 and 2025, we incurred computing resource costs of RMB0.2 million, RMB9.4 million and RMB117.4 million, respectively, representing 29.8%, 46.5% and 89.0% of our cost of sales for AI cloud computing services during the same years, respectively.

  In addition, the establishment of our AI cloud computing services required us to invest in specialized hardware. Once these assets were placed into use, we began to recognize depreciation over their estimated useful lives, which resulted in relatively higher depreciation in the initial periods even though the related revenue from our AI cloud computing services was still at an early ramp-up stage. In 2023, 2024 and 2025, we incurred depreciation costs of nil, RMB7.0 million and RMB5.3 million, representing nil, 34.4% and 4.0% of our cost of sales for AI cloud computing services during the same years, respectively.

- *Substantial research and development expenses.* We recorded substantial research and development expenses to iterate and enhance our AI cloud computing services, with a clear focus on technologies that we expect will improve our profitability over the long term. For example, we advanced our research on distributed inference acceleration technology, which is designed to enable faster AI inference for our customers and optimize the utilization of our internal computing resources. This research involves extensive testing in real-world scenarios with large volumes of data input to verify the efficacy, stability and reliability of the technology. Such testing is highly compute-intensive and requires significant GPU computing resources and cloud-based services, which in turn contributed to our R&D expenses. From 2024 to 2025, our testing expenses increased substantially from RMB2.4 million to RMB24.3 million, representing 2.8% and 20.3% of our research and development expenses, respectively. Such rapid growth in testing expenses reflects our continued investment in pushing the upper boundaries of what our AI inference acceleration technologies can achieve and in rigorously validating the compatibility of our cloud computing services with newly launched AI models.

  In addition, developing AI cloud computing services requires highly skilled R&D personnel, who generally command competitive levels of compensation. During the Track Record Period, we continued to expand our R&D team, in particular personnel specializing in AI-related technologies such as large-scale model optimization and inference engine development, in order to enhance our R&D capabilities and deliver more up-to-date services and features to our customers. In 2023, 2024 and 2025, we incurred employee salaries and other benefits of RMB54.9 million, RMB71.6 million and RMB80.8 million, representing 80.9%, 82.9%, and 67.3% of our research and development expenses during the same period, respectively. Such expenses reflected both the growth in the size of our R&D team and rising average compensation levels for experienced AI engineers and researchers. We expect to continue to incur significant R&D staff costs as we further develop our AI cloud computing services and seek to maintain our technology leadership.

Factor related to both our edge cloud computing services and AI cloud computing services:

- *Ongoing expenses for operating and maintaining edge cloud and AI cloud computing infrastructure*. In addition to the upfront investment in building cloud computing infrastructure as discussed above, we also incurred recurring costs to operate and maintain such computing infrastructure, including computing resource costs (such as server costs, electricity and maintenance costs) and cloud service expenses. Operating AI workloads, particularly large-scale AI inferences, is power- and compute-intensive and requires us to maintain sufficient capacity to meet customers' latency and availability requirements across different regions. In addition, we incur ongoing expenses for data storage, security and system monitoring, and for regularly upgrading or replacing hardware to keep pace with advances in AI technology.

Our Path to Profitability

Expand Economies of Scale Leveraging the Surge of AI

Currently, AI inference has replaced AI training as the primary demand for computing power. AI inference applications permeate consumer and enterprise workflows at massive scale, generating persistent, high-frequency demand for efficient, elastic and cost-effective inference services. Looking ahead, as agentic AI proliferates, the primary callers of cloud computing infrastructure will increasingly shift from human users to AI agents autonomously executing complex, multistep tasks at massive scale. This paradigm shift will give rise to entirely new and sustained demand for cloud computing resources, characterized by high-frequency concurrency, persistent long-context memory and multi-agent workflow orchestration. Such demand from agentic AI for computing power is structurally different from, and incremental to, existing AI inference workloads.

We have established agentic infra as a new business sub-segment and will continue to upgrade and expand our product offerings within this sub-segment. For example, we will further upgrade our agent sandbox, a product that provides a secure, isolated environment for executing AI-generated code. This sandbox environment prevents AI agents from accessing or tampering with resources outside the sandbox, ensuring that AI agents' behavior does not cause harm to the broader system.

As the agentic AI era unfolds, the usage of agentic infra will scale and we expect a further expansion of the economies of scale: a larger and more diverse base of agentic workloads will generate additional token consumption volume flowing through our global compute scheduling network, improving overall resource utilization, reducing per-unit compute costs and further strengthening the profitability of our AI cloud computing services.

Collaborate with Open-source Communities and Pursue Developer-friendly Growth Strategy

We collaborate with open-source communities (such as GitHub and OpenRouter) with extensive developer networks. By integrating our products into partner ecosystems and pursuing joint sales initiatives, we expect to accelerate new customer acquisition, broaden our industry coverage and increase revenue per customer. We ranked first on OpenRouter among third-party open-source large model providers in January 2026 and ranked first on Hugging Face by monthly request volume of inference services in December 2025.

We also implement developer-friendly strategies to attract new customers at a relatively low acquisition cost. We actively participate in open-source community events, provide open-source integrations, and host online and in-person technical workshops and developer forums. In 2025, our AI cloud computing services had over 180,000 active users. We plan to further tap into this user base by converting more developers into paying customers, upselling higher-value services and encouraging broader usage across their organizations.

Broaden Our Customer Reach and Expand Our Service Network

As we scale our service network and expand our customer base, we are committed to leveraging our leadership position to capture the significant growth opportunities in the distributed cloud service market. By broadening our global reach, cultivating relationships with key accounts across diverse industries, and building a thriving developer ecosystem, we aim to drive sustained growth. Paired with efficient operations and economies of scale, these efforts position us to achieve long-term profitability while delivering exceptional value to our customers:

- *Strengthening relationships with major customers of edge cloud services*. Our edge cloud computing services mainly serve mega internet companies in China. Our customers include majority of the top 10 Chinese internet companies, according to CIC. Our long-term cooperation with these customers has resulted in deep technical, service and product integration. Going forward, we will continue to provide high-quality customer support and full-lifecycle services. We have dedicated after-sales and project management teams that provide rapid responses, ongoing progress updates, technical consultation and hands-on implementation support, especially for customers with complex integration needs. This strengthens customer satisfaction and retention and creates opportunities for cross-selling and upselling additional services.

- *Growing our computing resources network globally.* We have continued to grow our computing resources network globally. In particular, for our AI cloud computing services, the number of our computing nodes increased from 19 as of December 31, 2024 to 64 as of December 31, 2025 and further to 77 as of April 30, 2026. We focus on the following key regions, with commercial reasons set forth below: (i) North America, where the supply of GPU resources is relatively abundant and customer demand for such computing resources is high; accordingly, we are proactively expanding our computing resources in North America in line with the growth of our customer base in this region;

(ii) Southeast Asia. In addition to Singapore, where we established our first compute node in this region, we have added compute nodes in countries such as Indonesia, Thailand and Malaysia, taking advantage of their lower electricity costs; and (iii) Japan, particularly its northern region, which offers favorable climate conditions (lower temperatures leading to lower energy consumption for heat dissipation) and where local governments provide energy subsidies that support the growth of computing services.

- *Expanding into new industry verticals with tailored solutions.* we will continue to broaden our customer base across diversified industry sectors. Currently, we mainly serve customers across pan-entertainment, telecommunications and information technology sectors. We intend to deepen our penetration in existing verticals and expand into additional sectors such as education, healthcare and financial services, where demand for low-latency, high-reliability and cost-effective computing power is expected to continue to grow. For example, we are exploring how our services can support content generation tools and marketing activities for e-commerce companies, enabling them to create content more efficiently and conveniently. Similarly, we are developing integrations between our AI cloud computing services and AI-assisted teaching tools for education companies, helping them leverage AI capabilities in a more cost-efficient way. By offering such industry-specific solutions, we aim to acquire new customers and increase the breadth and depth of usage among existing customers.

Upgrade Our Technology and Service Offerings

Leveraging the success of our current offerings, we plan to further optimize our services to improve their adaptability, accuracy and efficiency, thereby enhancing our competitive edge.

- First, we will continue to upgrade our agent sandbox, our new offering under the agentic infra business segment. Agentic infra is the infrastructure that is purposely built to support AI agents' autonomous and durable execution of tasks. Our agent sandbox is a secure, isolated environment for executing AI-generated code. This sandbox environment prevents AI agents from accessing or tampering with resources outside the sandbox, ensuring that AI agents' behavior does not cause harm to the broader system. We are upgrading our agent sandbox into a more powerful environment that not only isolates AI-generated code but also provides enhanced safety controls and supports more complex, long-running agent workflows.

- Second, we plan to develop ultra-low latency computing power network. Ultra-low latency computing power network is a critical infrastructure to support advanced AI systems to achieve millisecond-level responsiveness and make real-time decisions in scenarios such as autonomous driving, remote surgery and smart manufacturing.

- Third, we will augment our model-hosting capabilities by ensuring our platform supports an increasingly diverse array of models. This adaptability is vital for addressing the varied business needs and preferences of our customers.

- Fourth, we aim to continue to enhance the integration between cloud and edge systems. Edge devices will independently manage local tasks and send more complex workloads to the cloud for combined processing. This improved coordination can boost network throughput and reduce end-to-end latency.

- Lastly, we intend to facilitate the integration of next-generation intelligent hardware (such as NPUs) with our computing power network. This approach will enhance our ability to support sustained operations and manage situations involving high levels of simultaneous computing resource requests. These initiatives serve not only to enrich our service offerings but also to strengthen our position in the rapidly evolving AI landscape.

Optimize Operational Efficiency

Our ability to manage and optimize our expenses is critical to the success of our business and our ability to achieve sustainable profitability.

During the Track Record Period, we incurred substantial operating expenses, including research and development expenses, selling and marketing expenses and administrative expenses. The following table sets forth our research and development expenses, selling and marketing expenses and administrative expenses, as a percentage of revenue for the years indicated:

	Year ended December 31,		
	2023	2024	2025
		(%)	
As a percentage of revenue:			
Research and development expenses	18.9	15.5	15.6
Administrative expenses	7.1	5.6	7.4
Selling and marketing expenses	5.6	4.6	5.5
Total	31.6	25.7	28.5

- *Research and development expenses*. Our research and development expenses were RMB67.8 million, RMB86.3 million and RMB119.9 million, in 2023, 2024 and 2025, respectively, representing 18.9%, 15.5%, and 15.6% of our total revenue, respectively. Our research and development expenses as a percentage of revenue generally decreased, primarily attributable to the rapid increase in revenue and the achievement in economies of scale in line with our business growth. As we continue to drive service iteration and technological upgrades, we expect to further enhance our R&D efficiency by relying on our accumulated technological capabilities.

- *Administrative expenses*. Our administrative expenses were RMB25.3 million, RMB31.3 million and RMB57.3 million in 2023, 2024 and 2025, respectively, representing 7.1%, 5.6% and 7.4% of our total revenue, respectively. Excluding the impact of [**REDACTED**], our administrative expenses in 2025 would be RMB40.4 million, representing 5.2% of our total revenue. Our administrative expenses (excluding [**REDACTED**]) as a percentage of revenue had continued to decrease during the Track Record Period, primarily attributable to our effective cost control measures and our efforts in increasing our operational efficiency.

- *Selling and marketing expenses*. Our selling and marketing expenses were RMB20.2 million, RMB25.8 million and RMB42.4 million in 2023, 2024 and 2025, respectively, representing 5.6%, 4.6% and 5.5% of our total revenue, respectively. Our selling and marketing expenses as a percentage of revenue decreased primarily due to the rapid increase in revenue and our ability to increase the efficiency of our sales and marketing activities as a result of the economies of scale.

We recognize that disciplined management and control of operating expenses are vital to achieving and maintaining profitability. By combining technology-driven efficiency, effective cost management and a scalable operating model, we are well-positioned to enhance profitability while continuing to invest in innovation, growth and long-term success.

Benefiting from the revenue growth drivers and cost control measures described above, we expect our revenue to continue to grow. We are also committed to expanding our operations to capture economies of scale, thereby managing operating expenses more efficiently and supporting our path to profitability. Our Directors are therefore of the view, and the Joint Sponsors concur, that our business is sustainable.

Working Capital Sufficiency

Taking into account our available resources including cash and cash equivalents, time deposits, pledged deposits, operating cash flows, bank borrowings, the available banking facilities and the estimated [**REDACTED**] available to us from the [**REDACTED**], our Directors believe, and the Joint Sponsors concur, that we have sufficient working capital for our present requirements and for at least the next 12 months from the date of this document.

OUR CORE TECHNOLOGIES

Proprietary Distributed Computing Architecture

Heterogeneous computing resource pooling. Our system integrates diverse computing resources into a unified resource pool. Through intelligent workload orchestration, our architecture reduces the heterogeneity of underlying hardware, enabling seamless integration of diverse compute units while presenting a standardized access for upper-layer applications.

Containerized virtualization and edge-native system. Based on a lightweight cloud-native architecture, our system efficiently organizes fragmented edge resources along two complementary dimensions: an AI computing layer, which handles AI inference and high-performance computing workloads; and an edge computing layer, which drives latency-sensitive services through a distributed edge node network. This design allows workloads to be intelligently routed to the optimal resource tier based on performance requirements, latency constraints and cost objectives, while significantly reducing system overhead compared to traditional solutions. This ultra-compact design supports deployment on single-server edge nodes, maximizing resource utilization even in constrained environments. The system includes serverless GPU technology, which enables auto-scaling, fault recovery and hardware-agnostic deployment. Unified orchestration is achieved through the use of containers and overlay networks, ensuring seamless task migration and multi-tenant isolation across geographically dispersed nodes.

Computing Power Network and Scheduling Capabilities

Largest computing power network. We have established a computing power network covering more than 1,340 cities and counties globally with over 4,600 compute nodes as of December 31, 2025. This high-density compute node deployment enables millisecond-level latency (as short as 10 milliseconds) for local distribution, providing infrastructural support for scenarios such as real-time audiovisual content delivery and AI inference.

Global Scheduling Capabilities. Our global scheduling network enables us to achieve high utilization of computing resources on a continuous, around-the-clock basis. By leveraging cross-regional time-zone differences and the staggered demand profiles of our diverse customer base, we dynamically allocate globally idle computing capacity to wherever demand is high at any given moment, ensuring that our GPU resources are productively deployed rather than sitting idle during off-peak periods. As a result, our average GPU utilization rate is consistently maintained at above 75% in 2025, significantly outperforming the industry average of between 40% and 50%, according to CIC. This intelligent "peak-shaving and valley-filling" capability materially improves overall resource utilization across our network, directly reducing the effective cost per token delivered to our customers.

Distributed Inference Acceleration for LLMs

Model optimization. We enhanced widely adopted open-source models through a combination of hardware and software innovations. By intelligently combining compute tasks, simplifying calculations, and using techniques such as speculative sampling, we have boosted the model output efficiency by over seven times and reduced theoretical operating costs by 85.7%. This facilitates faster, more affordable AI performance without sacrificing quality.

Distributed inference optimization. Our optimization capabilities span multiple dimensions, such as operator optimization, inference engine optimization, storage system optimization and model optimization tailored for cutting-edge hardware. Through these multi-layered optimizations, we are able to drive computing resource utilization to near its theoretical limits, substantially reducing costs. For example, in 2026, the input token pricing of a well-known large model delivered through our model API services is more than 40% lower than the average charged by major international cloud providers.

Our optimization capabilities span multiple dimensions, such as operator optimization, inference engine optimization, storage system optimization and model optimization tailored for cutting-edge hardware. Through these multi-layered optimizations, we are able to drive computing resource utilization to near its theoretical limits, substantially reducing costs.

KEY OPERATING METRICS

We manage our business using the following key operating metrics. We use these metrics to assess the progress of our business and make decisions on how to allocate capital, time and technology investments.

	As of/In the year ended December 31,			As of/During the four months ended
	2023	2024	2025	April 30, 2026
Edge Cloud Computing Services				
Number of compute nodes[(1)]	2,849	4,012	4,632	4,956
Number of cities and counties covered by our edge cloud compute nodes[(2)]	999	1,296	1,345	1,441
Number of CPU cores[(3)]	299,166	387,011	446,525	479,148
Aggregate server storage capacity (TB)[(3)]	57,693	64,390	92,230	98,700
Number of paying customers	23	23	32	23
AI Cloud Computing Services				
Number of compute nodes[(1)]	5	19	64	77
Number of cities and counties covered by our AI cloud compute nodes[(2)]	5	15	42	53
Number of open-source models[(4)]	–	82	176	146
Number of registered developers[(5)]	12,112	125,545	524,436	574,485
Number of active users[(6)]	12,169	114,692	180,851	59,020
Number of paying customers[(7)]	7,454	47,860	91,538	13,832
Aggregate computing power (PFLOPS)[(8)]	238	1,482	17,430	25,793
Average purchase per user[(9)]	–	–	226,065	352,018

Notes:

(1) As we operate under an asset-light model, substantially all of our compute nodes are leased.

(2) As of December 31, 2025, all of our compute nodes for edge cloud computing services were located in China; as for our compute nodes for AI cloud computing services, 26 were located in China, 8 were located in the United States, 2 were located in the United Kingdom, and 1 was located each in Canada, Singapore, Iceland, Portugal, Indonesia and Malaysia.

(3) Referring to the volume during the last month of the relevant year/period.

(4) Referring to the number of open-source models made publicly available through our model API services during the relevant year/period. The number of open-source models decreased from 2025 to the four months ended April 30, 2026 as we removed low-usage, duplicate and experimental models to concentrate resources on core flagship models. This did not dampen platform usage: token consumption and aggregate computing power continued to grow over the same period, demonstrating that fewer models did not reduce user demand. Rather, the reduction signals a strategic shift from expanding model breadth toward optimizing model quality, inference efficiency and commercial conversion.

(5) Referring to developers who had registered on our platform as of December 31, 2023, 2024 and 2025 and April 30, 2026, respectively.

(6) Referring to users who used our model API services in 2023, 2024 and 2025 and the four months ended April 30, 2026. The number of active users for the four months ended April 30, 2026 was not annualized and therefore was lower.

(7) The number of our paying customers decreased from 2025 to the four months ended April 30, 2026 was primarily in line with the decrease in the number of open-source models. Please see Note (4) above for details. As a result, the number of low-paying customers declined, partially offset by an increase in the number of high-paying customers. This shift in customer mix reflected our continued commercial focus on enhancing revenue quality and monetization efficiency.

(8) PFLOPS is a standard unit to describe very large compute capacity. 1 PFLOPS equals 10^{15} floating point operations per second.

(9) Average purchase per user for AI cloud computing services is calculated as the total revenue from users whose spending exceeded RMB1,000 in a given year/period divided by the average number of such users during that year/period.

In addition, for our AI cloud computing services segment, the average daily token consumption, which represents the average number of tokens processed by our AI cloud computing platform per day during a given period, increased from 27.1 billion in December 2024 to 271 billion in December 2025, and further to 1,028 billion in April 2026.

RESEARCH AND DEVELOPMENT

Our deep passion for innovation coupled with our strong R&D capabilities have allowed us to keep up with the rapid advances in technology in the cloud computing industry. For the years ended December 31, 2023, 2024 and 2025, we incurred research and development expenses of RMB67.8 million, RMB86.3 million and RMB119.9 million, respectively, representing 18.9%, 15.5% and 15.6% of our revenue during the same years, respectively.

We place strong emphasis on the recruitment of technology specialists and senior engineers with extensive experience in the industry. Our technological capabilities are built by our talented and dedicated research and development team. As of December 31, 2025, we had a team of 170 R&D professionals, representing approximately 61.8% of our total employees. We have established various training programs to keep our R&D personnel abreast of the most advanced technologies in the relevant fields. Our R&D were primarily performed in-house during the Track Record Period.

Research and Development Process

The key stages in our research and development process consist of:

- *Concept stage*. Our R&D team identifies the needs of the market and technological gaps, drawing on input from existing customers and industry stakeholders. We establish clear objectives for what the product candidate should achieve.

- *Feasibility study*. We examine both technical and commercial aspects of the product candidate. Technical feasibility involves assessing available technologies, integrations and compatibility with existing systems, while economic feasibility evaluates costs and resource requirements.

- *Design and development*. We design the architecture of our product candidate, incorporating components such as software, hardware and algorithms. Further, we develop a prototype and conduct rigorous testing for functionality, security and scalability. The prototype also requires iterative revisions to refine features and resolve issues.

- *Deployment testing*. We release a trial version to select users for real-world testing. We collect feedback on usability and evaluate the performance of our product. We also conduct stress testing to evaluate reliability and scalability of our product under high demand.

- *Launch.* We launch our product with full functionalities to all users. We also formulate marketing strategies to promote the unique features and benefits of our products, and ensure sufficient and timely customer support is available for our customers.

Key R&D Projects

We set forth below examples of our key R&D projects:

- *AI-Powered Customer Service System.* This project focuses on developing a next-generation AI customer service platform that goes beyond scripted responses by deeply integrating large language models with business workflows. The goal is to build a system capable of making its own decisions and proactively serving customers, rather than simply reacting to queries.

- *CDN Traffic Scheduling System.* This project aims to develop a CDN product that is low-cost, high-performing and easy to operate and maintain. Research is focused on three areas: first, categorizing network resources by their characteristics and building a cost model to optimize how bandwidth is allocated; second, using historical traffic data to predict future demand patterns and guide bandwidth planning more efficiently; and third, building a visual, user-friendly operations management system that improves the safety, accuracy and speed of network changes.

- *Virtual Machine Platform.* This project applies virtualization technology to our existing physical computing resources, enabling them to be divided into multiple virtual machines. This creates new and more flexible ways to deliver and operate services, and improves overall resource utilization.

INTELLECTUAL PROPERTY

Our patents, copyrights, trademarks, trade secrets and other intellectual property rights are critical to our business operations. We rely primarily on a combination of patents, copyrights, trademarks, trade secrets and anti-unfair competition laws and contractual rights, such as confidentiality agreements entered into with our employees and customers, to protect our intellectual property rights. We clearly state all rights and obligations regarding the ownership and protection of our intellectual properties in employment agreements and commercial agreements. In addition, we have implemented a set of comprehensive measures to protect our intellectual property.

As of the December 31, 2025, we had 34 patents, 37 trademarks, 32 domain names and 67 copyrights in the PRC.

During the Track Record Period and as of the Latest Practicable Date, we had not been subject to any material disputes or claims for infringement upon third parties' intellectual property rights.

DATA PRIVACY AND DATA SECURITY

Data security and protection are among our highest priorities. In this regard, we have obtained ISO 27001 Information Security Management System Certification. We have designed strict data protection and information security policies to ensure strict compliance with applicable laws, regulations, and prevalent industry practice.

Data Collection

In providing our edge cloud computing services, for purposes of integrating computing resources from third-party providers to meet our cooperate customers' needs for edge computing and transmission and completing transactions, that we typically collect data include: (i) the basic information, payment information and contact information of our customers; (ii) the basic information, payment information, contact information, device information, and certificate of qualification of the computing power suppliers.

In providing our AI cloud computing services, in addition to collecting the above data, we collect certain personal information through our website, such as mobile number, user name, email address, and identity authentication information (such as real name, ID number), and specify the

circumstances under which we collect personal information. We process such personal information only to the extent necessary for account registration and login, completing real-name identity authentication of our customers (including corporate customers and individual developer users), providing them with relevant services, and responding to service consultation.

During the Track Record Period, we do not purchase any personal information or other data from our customers or any other third parties. Our customers retain full ownership and control of their data. For the avoidance of doubt, as we operate solely as an intermediary technical service provider, we do not have access to our customers' or their end users' data when our distributed cloud computing services are being used.

Data Storage and Deletion

All information and data we received in the mainland China have been stored and preserved within mainland China. We differentiate data security storage based on different security levels. We will delete the data after the data retention period expires, in accordance with relevant laws and regulations.

Data Sharing and Transmission

To meet our customers' needs for edge computing and transmission, our edge cloud computing services do not transmit data to a centralized cloud data center, but to the edge servers through our edge transmission technologies, allowing computation to be processed at the "edge" of the internet, near where the data are generated and used. Our domestic and overseas systems are deployed on separate cloud servers, which are both physically and virtually isolated from each other, with no data transmitted from domestic to overseas. In addition, we have adopted internal rules and procedures designed to prevent illegal and/or unauthorized cross-border transmission of data. These measures ensure that our business operations do not involve any unauthorized and/or illegal cross-border transmission of data.

Pursuant to our internal control rules, the Privacy Policy and the contract between us and the customer, unless under legal requirement or we have obtained our customers' permission and authorization, we will not share, disclose or transfer our customers' data or personal information to third parties.

Internal Control Measures

We have implemented internal policies on protecting data privacy and security, with the purpose of ensuring data and information security, optimizing data governance, protecting the benefits of our customers, business partners, employees and other third parties, and ensuring compliance with all applicable laws and regulations. We implement a robust internal authentication and authorization system to ensure that our confidential and important business data and trade secrets can only be accessed for authorized use and by authorized personnel. We have established an information system in relation to data security requirements, national standards and industry best practices, and intend to continually invest heavily in data security and privacy protection. Our information system applies multiple layers of safeguards, including both internal and external firewalls, to identify and protect us against security attacks. Our major business systems have obtained record filing certificates for the graded protection of information system (信息系統安全等級保護備案證明) according to the Measures for the Administration of the Graded Protection of Information Security (《信息安全等級保護管理辦法》). All information and data we received in the mainland China have been stored and preserved within mainland China. Our domestic and overseas systems are deployed on separate cloud servers, which are both physically and virtually isolated from each other, with no data exchange between them. In addition, we have adopted internal rules and procedures designed to prevent illegal and/or unauthorized cross-border transmission of data. These measures ensure that our business operations do not involve any cross-border transmission of data.

Compliance with Data Privacy and Cybersecurity Laws in China and our Key Overseas Markets

During the Track Record Period and up to the Latest Practicable Date, (i) we had not received any claim from any third party against us on the ground of infringement of any third party's right to data and privacy protection as provided by any applicable laws and regulations in the PRC or other jurisdictions; (ii) we had not received any complaint relating to data privacy or security measures; (iii) we had implemented internal policies on protecting data privacy and security, with the purpose of ensuring data and information security and ensuring compliance with all applicable laws and regulations; (iv) during the Track Record Period, there was no material incident of data or personal information leakage; (v) during the Track Record Period, there was no investigation, legal proceeding or administrative penalty relating to the violation of relevant network security, data security and personal information protection laws or regulations, to our best knowledge, pending or threatened against us initiated by competent government authorities or third parties; and (vi) we will continue to pay close attention to the regulatory developments in data security and comply with the latest regulatory requirements. In view of the foregoing, our PRC Legal Advisor is of the view that we complied with currently effective and applicable laws and regulations on data privacy and security in all material respects during the Track Record Period and up to the Latest Practicable Date.

Based on the advice of our U.S. data laws legal adviser, our U.S. privacy law compliance obligations are primarily limited to the California Consumer Privacy Act ("CCPA"), as amended by the California Privacy Rights Act ("CPRA"). With regard to cybersecurity, while the U.S. does not have a standalone national cybersecurity law, regulatory guidance and enforcement, particularly by the Federal Trade Commission ("FTC"), require that companies implement and maintain reasonable security procedures and practices appropriate to the nature of the information handled. Based on the assessment of our U.S. data laws legal adviser, 1) our practice meet or exceed the baseline "reasonable security" expectations under U.S. law; 2) our adoption of contractual, technical safeguards with service providers and customers further supports compliance, and 3) our business operations are consistent with service provider obligations under the CCPA/CPRA and comparable U.S. privacy and cybersecurity laws. Further, our U.S. data laws legal adviser is of the view that the U.S. federal and state privacy and cybersecurity laws do not impose obligations that would materially adversely affect our business operations.

Based on the advice of our Singapore data laws legal adviser, the Personal Data Protection Act 2012 ("PDPA") is Singapore's principal legislation relating to data privacy, security and cross-border transfers. The PDPA sets out compliance obligations of general application that apply broadly to organisations that collect, use or process personal data in Singapore. There had been no enforcement action, investigation, regulatory direction or financial penalty imposed against us under the PDPA during the Track Record Period and up to the date of the opinion of our Singapore data laws legal adviser. With regard to cybersecurity, our Singapore data laws legal adviser has advised that, as neither we nor our products or services have been designated as critical information infrastructure under Singapore's Cybersecurity Act 2018 and we do not provide licensable cybersecurity services, the regulatory framework thereunder would not reasonably be expected to have any material adverse impact on our business operations.

ALGORITHM AND AI COMPLIANCE

In recent years, the PRC authorities have promulgated certain regulations in respect of algorithm and AI regulations in the PRC, including the Administrative Provisions on Deep Synthesis of Internet Information Service (《互聯網信息服務深度合成管理規定》) (the "Deep Synthesis Administrative Provisions") which took effect on January 10, 2023, the Management of Generative Artificial Intelligence Service (《生成式人工智能服務管理暫行辦法》) (the "AI Measures") which took effect on August 15, 2023, and the Measures for the Identification of AI-Generated and Synthesized Content (《人工智能生成合成內容標識辦法》) (the "Identification Measures") which took effect on September 1, 2025. For details, see "Regulatory Overview – Regulations Relating to Generative Artificial Intelligence Services" in this document.

Among the products and services we provide, only the Model API involves generative artificial intelligence services, which is available exclusively to:(i) corporate customers; and (ii) individual developer users with limited trials. We provide the Model API is subject to the Deep

Synthesis Administrative Provisions and the Identification Measures. Pursuant to the AI Measures, there are two elements for determining the applicability of the AI Measures: (i) whether the services are Generative Artificial Intelligence Services; and (ii) whether such services are offered to the public within China. In this regard, there has been no official interpretation or specific implementing rules; it essentially remains uncertain as to whether our Model API will be deemed to offer services to the public and thus subject to the AI Measures.

Despite the uncertainty, as of the Latest Practicable Date, we have taken a series of measures (considering the possibility that the AI Measures apply to us) to comply with the above regulations related to algorithms and AI, including but not limited to: (i) we have established a Dedicated Algorithm Security Department ("算法安全專職機構"); (ii) we have conducted regular security tests and evaluations on algorithms; (iii) we have set up an information security monitoring mechanism to review the content generated, provided channels for complaints and reports of illegal information, and we will timely dispose illegal content when we detect it; (iv) we have formulated and implemented the Algorithm Security Monitoring System, Algorithm Security Self-Assessment System, Algorithm Illegal and Irregular Handling System, Algorithm Security Incident Emergency Response System, and Technology Ethics Review System among other internal measures; (v) we have completed the filing of three deep synthesis service algorithms in the Internet Information Service Algorithm Filing System of the CAC, marked the record number in a prominent position on our website and provided links to public information; (vi) we have added explicit or implicit labels to AI generated synthetic content and implemented relevant technical preparations; (vii) we maintain communication with the relevant PRC authorities to fully understand and comply with the latest compliance requirements. Our Directors are of the view that, even if the AI Measures would apply to us and the Identification Measures is officially implemented, such measures would not have a material adverse impact on us. Based on the foregoing, our PRC Legal Advisers, our Joint Sponsors and our Directors are of the view that, during the Track Record Period and up to the Latest Practicable Date, we are in compliance with the above regulations related to algorithms and AI in all material aspects.

SALES AND MARKETING

As of December 31, 2025, we had a sales and marketing team of 43 personnel. We implement product-led growth and developer-friendly strategies to attract customers. We attract developers to use our products and services by participating in open-source community events and providing open-source integrations. This strategy helps us acquire new customers at a lower cost and expand our influence within the developer community. We also conduct sales through leveraging the network effect and referrals by stakeholders across the business cycle, and through on-site visits, industry conferences and other marketing events, to strategically expand our market presence and scale up our business in a cost-effective manner.

Pricing

We adopt different pricing models for specific products or services. The following table sets forth the pricing model for each of our major products or services:

Product/Service	Pricing Model
Edge Cloud Computing Services	
Edge Node Services	Charge by actual usage, ranging between RMB2 and RMB4 per MB.
Edge CDN	Charge by actual usage, ranging between RMB2 and RMB4 per MB.
AI Cloud Computing Services	
GPU Cloud Services	Charge by hours of usage (*i.e.* Total price = GPU unit price × number GPUs × usage time)

Product/Service	Pricing Model
Model API	Charge by actual consumption of tokens, number of images generated or the length of video generated. Specifically, (i) the fees for large language model services are calculated separately based on the volume of input tokens and output tokens consumed in each interaction. Total cost = (number of input tokens × input unit price) + (number of output tokens × output unit price); and (ii) for multimodal models, image generation is billed on a per-image basis; video generation is billed on a per-video basis, with pricing that scales upward according to the duration and resolution of the output; and speech synthesis services are billed primarily on a per-character basis.
Agentic infra	Sandbox services are billed on a per-second basis, with charges calculated independently for virtual CPU core usage and memory capacity respectively; billing ceases upon termination of the sandbox instance.

Customer Support and After-sales

We recognize that exceptional customer support is a cornerstone for nurturing enduring partnerships with our clients and strengthening our brand reputation over time. Our highly responsive and professional after-sales service not only boosts customer satisfaction and engagement but also plays a pivotal role in consistently achieving strong customer retention. We are committed to providing swift, reliable assistance. Our support teams deliver rapid responses to customer inquiries and provide ongoing progress updates. For clients with more complex after-sales needs, our specialized project managers offer comprehensive lifecycle support. They provide tailored technical consultations and hands-on guidance, enabling seamless integration of our services with the client's unique business requirements. Through continuous dialogue and attentive service, we gain valuable insights into our customers' evolving expectations and business challenges. This proactive engagement not only equips us to anticipate future needs but also positions us to uncover new commercial opportunities, reinforcing our role as a trusted, forward-looking business partner.

OUR CUSTOMERS

We have a broad base of customers across various industries, including, among others, pan-entertainment, social media, e-commerce, education, automotive, telecommunications, information technology and intelligent manufacturing. As our product lines expand and our service areas extend, we continue to grow our customer base. While maintaining high-quality service for major customers, we also actively cultivate relationships with emerging customers to drive ongoing business growth.

In 2023, 2024 and 2025, our revenue amounted to RMB358.4 million, RMB558.0 million and RMB770.3 million, respectively. Our revenue generated from our largest customer in 2023, 2024 and 2025 accounted for 44.1%, 35.2% and 30.0%, respectively, of our revenue during the same years. Our revenue generated from our five largest customers in 2023, 2024 and 2025 accounted for 92.5%, 89.5% and 79.0% respectively, of our revenue during the same years. See "Risk Factors—Risks relating to our operations—The loss of, or a significant reduction in usage by, one or more of our major customers would result in lower revenue and could harm our business."

Our Relationship with Our Largest Customers (Customer A and Customer B)

For the years ended December 31, 2023, 2024 and 2025, our revenue generated from Customer A accounted for 44.1%, 35.2% and 30.0% of our total revenue, respectively; and our revenue generated from Customer B accounted for 27.7%, 32.6% and 23.3% of our total revenue, respectively.

Our Directors consider that our customer concentration during the Track Record Period is primarily attributable to our ability to build and sustain strong, long-term relationships with our major customers. As of the Latest Practicable Date, our business relationships with Customer A and Customer B had been approximately five years and six years, respectively. We believe that we have maintained the loyalty of our major customers by providing a comprehensive suite of edge cloud computing and AI cloud computing services that consistently meet, and continue to meet, their evolving technical requirements, performance standards and cost expectations. Further, even though our major customers already possess their own computing infrastructure, they rely on us to supplement their capacity during peak demand periods. Our ability to deliver deep technical integration, responsive customer support and continuous service optimization has enabled us to secure recurring and growing business opportunities from these customers over the years. During the Track Record Period, we did not have any material dispute with any customer who was one of our five largest customers in any year of the Track Record Period.

Despite our revenue concentration from our largest customers during the Track Record Period, we consider that our business is sustainable on the following grounds:

(i) except for Customer A and Customer B, none of our five largest customers in any year of the Track Record Period accounted for more than 20% of our total revenue during such year of the Track Record Period;

(ii) we have maintained stable and longstanding business relationships with each of our five largest customers throughout the Track Record Period. As of the Latest Practicable Date, our business relationships with Customer A and Customer B had been approximately five years and six years, respectively, and the long-term cooperation with these customers has resulted in deep technical, service and product integration;

(iii) we believe that it is mutually beneficial for both our major customers and us to maintain these business relationships. According to CIC, major internet companies in China have significant demand for edge cloud computing services to provide low-latency content deliver to their customers. Major internet companies generally prefer to work with established, technically capable cloud service providers with proven infrastructure at scale, since such providers are better positioned to deliver the performance, reliability and cost efficiency required by large-platform workloads, and are able to accommodate rapid fluctuations in computing demand without service degradation. As a result, cloud computing service providers' customer concentration on major internet companies is an industry norm, according to CIC. We believe that our proprietary distributed computing architecture, leading scheduling capabilities, distributed inference acceleration technologies, extensive network of globally dispersed compute nodes and ability to deliver reliable and cost effective services at scale form the cornerstone of our strong relationships with our major customers, which our Directors believe will continue to remain stable; and

(iv) our customer concentration has improved materially over the Track Record Period. The revenue contribution of our five largest customers declined from 92.5% in 2023 to 89.5% in 2024 and further to 79.0% in 2025, reflecting the steady broadening of our customer base across diversified industry verticals, along with the rapid growth of our AI cloud computing services.

We will continue to diversify and expand our customer base through continuous technological innovation, joint development with customers, expansion into new industry verticals and geographic markets and the further development of our AI cloud computing services. In particular, our AI cloud computing services have grown rapidly during the Track Record Period, attracting a broadening base of customers across diverse sectors that are distinct from and complementary to our established edge cloud computing customer base. This diversification of our customer profile across two distinct and mutually reinforcing business segments meaningfully reduces our reliance on any single customer or industry vertical, and we expect this trend to continue as AI adoption deepens across industries and geographies. We remain committed to continuously strengthening our technology capabilities, expanding our global computing network and enhancing the depth of our service offerings, in order to attract and retain a growing base of high-quality customers and to mitigate our exposure to any material adverse change to, or termination of, our relationships with our major customers.

Below is the breakdown of our revenue derived from our five largest customers for each year of the Track Record Period, and their respective background information:

For the year ended December 31, 2023:

Rank	Customer	Types of services provided	Transaction amounts	Percentage contribution to total revenue	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
1	Customer A	Edge cloud computing services	158,195.9	44.1	A Chinese company headquartered in Beijing, focusing on developing and operating a range of content platforms	2021	60 days
2	Customer B	Edge cloud computing services	99,348.2	27.7	A Chinese company headquartered in Beijing, focusing on providing network technology services, software development, and data processing services	2020	30 days
3	Customer C	Edge cloud computing services	51,224.2	14.3	A Chinese company headquartered in Shenzhen, focusing on communication and social media services	2021	28 days
4	Customer D	Edge cloud computing services	13,100.2	3.7	A Chinese company headquartered in Beijing, focusing on providing internet-related services, including search engine services and online marketing	2021	30 days
5	Customer E	Edge cloud computing services	9,765.9	2.7	A Chinese company headquartered in Changsha, focusing on culture and entertainment business	2021	60 days
	Total		**331,634.4**	**92.5**			

For the year ended December 31, 2024:

Rank	Customer	Types of services provided	Transaction amounts	Percentage contribution to total revenue	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
1	Customer A	Edge cloud computing services	196,176.4	35.2	A Chinese company headquartered in Beijing, focusing on developing and operating a range of content platforms	2021	60 days
2	Customer B	Edge cloud computing services	182,169.8	32.6	A Chinese company headquartered in Beijing, focusing on providing network technology services, software development, and data processing services	2020	30 days
3	Customer C	Edge cloud computing services	58,526.1	10.5	A Chinese company headquartered in Shenzhen, focusing on communication and social media services	2021	30 days
4	Customer D	Edge cloud computing services	32,761.3	5.9	A Chinese company headquartered in Beijing, focusing on providing internet-related services, including search engine services and online marketing	2021	28 days
5	Customer F	Edge cloud computing services	29,440.8	5.3	A Chinese company headquartered in Shanghai, focusing on the development and operation of digital entertainment platforms	2020	90 days
	Total		**499,074.4**	**89.5**			

For the year ended December 31, 2025:

Rank	Customer	Types of services provided	Transaction amounts	Percentage contribution to total revenue	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
1	Customer A	Edge cloud computing services	230,861.3	30.0	A Chinese company headquartered in Beijing, focusing on developing and operating a range of content platforms	2021	60 days
2	Customer B	Edge cloud and AI cloud computing services	179,473.2	23.3	A Chinese company headquartered in Beijing, focusing on providing network technology services, software development, and data processing services	2020	30 days
3	Customer C	Edge cloud computing services	106,752.8	13.9	A Chinese company headquartered in Shenzhen, focusing on communication and social media services	2021	28 days
4	Customer F	Edge cloud and AI cloud computing services	61,387.0	8.0	A Chinese company headquartered in Shanghai, focusing on the development and operation of digital entertainment platforms	2020	90 days
5	Customer D	Edge cloud computing services	29,777.2	3.9	A Chinese company headquartered in Beijing, focusing on providing internet-related services, including search engine services and online marketing	2021	30 days
	Total		**608,251.5**	**79.0**			

All of our five largest customers for each year during the Track Record Period were Independent Third Parties. None of our Directors, their close associates or any of our shareholders (who, to the knowledge of the Directors, own more than 5% of our issued share capital) had any interest in any of our five largest customers for each year during the Track Record Period and as of the Latest Practicable Date.

The summary of the salient terms of our standard agreements with our customers during the Track Record Period are set out below:

- *Scope*. We generally set out the scope of our services within the terms of services.

- *Pricing*. We primarily charge our customers by actual usage based on resource use, data traffic and bandwidth. For details, see "Business—Sales and Marketing—Pricing."

- *Payment and credit terms*. We send an invoice to customers every month detailing their actual usage of our services in the previous month. Customers are typically required to settle the invoice on a monthly basis. We generally grant a credit term of 30 to 90 days to our customers.

- *Confidentiality*. Each party is obliged to treat all confidential information made known to it by the other party in strict confidence during and after the contract term.

- *Term and termination*. The agreement typically has a term of one year. The agreement can be terminated by either party with prior written notice of 30 days or payment in lieu of notice. The term of the agreement can be extended on the same terms until such time as our customers cease to use the service.

During the Track Record Period and up to the Latest Practicable Date, we were in compliance with the terms of agreements with our major customers in all material aspects, and we had not experienced any circumstances leading to any contractual disputes with or claims by our major customers that would have a material and adverse effect on our operations.

OUR SUPPLIERS

Our suppliers are primarily providers for computing power resources. Our transaction amounts with our largest supplier in 2023, 2024 and 2025 accounted for 4.0%, 6.0% and 5.5% respectively, of our cost of sales during the same years. Our transaction amounts with our five largest suppliers in 2023, 2024 and 2025 accounted for 15.2%, 20.3% and 19.5% respectively, of our cost of sales. During the Track Record Period, we generally settled our payments to our suppliers by bank transfer. We have established and maintained stable and good relationships with our five largest suppliers for each year during the Track Record Period, having a relationship of four years or above with a majority of them. Our major suppliers are located in China. Please see below for details.

Below is the breakdown of our five largest suppliers for each year during the Track Record Period, and their respective background information:

For the year ended December 31, 2023:

Rank	Supplier	Types of services provided	Transaction amounts	Percentage contribution to cost of sales	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
1 . .	Supplier A	Computing resources	11,869.2	4.0	A Chinese company headquartered in Beijing, focusing on technical services, data processing	2020	15 days
2 . .	Supplier B	Computing resources	11,380.4	3.9	A Chinese company headquartered in Shenzhen, focusing on information technology consulting services	2022	15 days

Rank	Supplier	Types of services provided	Transaction amounts	Percentage contribution to cost of sales	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
3	Supplier C	Computing resources	10,787.0	3.7	A Chinese company headquartered in Anyang, focusing on software development, big data services	2021	15 days
4	Supplier D	Computing resources	5,545.0	1.9	A Chinese company headquartered in Guangzhou, focusing on IDC services and information technology	2022	15 days
5	Supplier E	Computing resources	4,896.2	1.7	A Chinese company headquartered in Nanjing, focusing on computer network maintenance services, technical services	2021	15 days
	Total		**44,477.8**	**15.2**			

For the year ended December 31, 2024:

Rank	Supplier	Types of services provided	Transaction amounts	Percentage contribution to cost of sales	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
1	Supplier F	Computing resources	29,375.6	6.0	A Chinese company headquartered in Baoding, focusing on internet data services	2021	15 days
2	Supplier A	Computing resources	23,128.1	4.8	A Chinese company headquartered in Beijing, focusing on technical services, data processing	2020	15 days
3	Supplier G	Computing resources	18,636.3	3.8	A Chinese company headquartered in Hangzhou, focusing on software and information technology services	2022	15 days
4	Supplier H	Computing resources	14,660.0	3.0	A Chinese company headquartered in Beijing, focusing on edge cloud services, CDN acceleration, network security, and big data	2023	15 days

Rank	Supplier	Types of services provided	Transaction amounts	Percentage contribution to cost of sales	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
5	Supplier D	Computing resources	13,405.2	2.7	A Chinese company headquartered in Guangzhou, focusing on IDC services and information technology	2022	15 days
	Total		**99,205.2**	**20.3**			

For the year ended December 31, 2025:

Rank	Supplier	Types of services provided	Transaction amounts	Percentage contribution to cost of sales	Background and principal business activities	Year of commencement of relationship with the Group	Credit term
			(RMB'000)	*(%)*			
1	Supplier I	Computing resources	38,502.7	5.5	A Chinese company headquartered in Guangzhou, focusing on software development services	2025	30 days
2	Supplier J	Computing resources	32,505.3	4.7	A Chinese company headquartered in Guangzhou, focusing on information technology consulting services	2025	30 days
3	Supplier K	Computing resources	23,608.6	3.4	A Chinese company headquartered in Shandong, focusing on internet information services	2023	15 days
4	Supplier F	Computing resources	22,741.0	3.3	A Chinese company headquartered in Baoding, focusing on internet data services	2021	30 days
5	Supplier G	Computing resources	18,333.2	2.6	A Chinese company headquartered in Hangzhou, focusing on software and information technology services	2022	30 days
	Total		**135,690.8**	**19.5**			

All of our five largest suppliers for each year during the Track Record Period were Independent Third Parties. None of our Directors, their close associates or any of our shareholders (who, to the knowledge of the Directors, own more than 5% of our issued share capital) had any interest in any of our five largest suppliers for each year during the Track Record Period and as of the Latest Practicable Date.

The summary of the salient terms of our agreements with our major suppliers during the Track Record Period are set forth below:

- *Scope*. The scope of our procurement is set forth within the terms of services.
- *Pricing*. We are typically charged by actual usage.
- *Payment and credit terms*. We are generally required to settle payment to our suppliers on a monthly basis after the suppliers issue invoices for services accrued in the previous month.
- *Confidentiality*. Each party is obliged to treat all confidential information made known to it by the other party in the strictest confidence during and after the contract term.
- *Term and termination*. The agreement typically has a term of one year. The agreement can be terminated by either party upon mutual agreement. The term of the agreement can typically be extended for one year on the same terms unless otherwise agreed by both parties.

During the Track Record Period and up to the Latest Practicable Date, we complied with the terms of the agreements with our major suppliers in all material aspects, and we had not experienced any circumstances leading to any contractual disputes with or claims by our major suppliers that would have a material and adverse effect on our operations.

OVERLAPPING CUSTOMERS AND SUPPLIERS

During the Track Record Period, (i) Customer A was one of our five largest customers in 2023,2024 and 2025 and our suppier in 2025; (ii) Customer C was one of our five largest customers in 2023, 2024 and 2025 and our supplier in the same years; (iii) Customer D was one of our five largest customers in 2023, 2024 and 2025 and our supplier in 2025; (iv) Supplier E was one of our five largest suppliers in 2023 and our customer in 2023; and (v) Supplier H was one of our five largest suppliers in 2023 and our customer in 2023, 2024 and 2025 (collectively, the "Overlapping Customers-Suppliers"). We primarily procured computing power resources and provided edge cloud computing services to the Overlapping Customers-Suppliers. Such Overlapping Customer-Suppliers possess surplus computing power resources during off-peak periods, yet require additional computing power resources during peak periods. Consequently, we both procure computing power resources from them and supply such resources to them as needed. According to CIC, this approach is consistent with prevailing industry practice. The following table sets forth our sales to and purchases from the Overlapping Customers-Suppliers during the Track Record Period:

	For the year ended December 31,		
	2023	2024	2025
Sales to the Overlapping Customers-Suppliers			
Revenue (RMB'000)	59,697.2	64,350.3	305,421.4
As a percentage of total revenue (%)	16.7	11.5	39.6
Purchases from the Overlapping Customers-Suppliers			
Purchase amount (RMB'000)	15,010.0	25,453.8	43,468.4
As a percentage of cost of sales (%)	5.1	5.2	6.2

Our Directors confirm that our sales to and our purchases from the Overlapping Customers-Suppliers were (i) entered into after due consideration taking into account the prevailing purchase and selling prices at the relevant times and (ii) conducted on an arm's length basis. The price charged by/to the Overlapping Customers-Suppliers are comparable with those charged to/by other customers and suppliers during the Track Record Period. To the best knowledge of our Directors, we did not have any other overlap between our other major customers and major suppliers during the Track Record Period and up to the Latest Practicable Date.

COMPETITION

The distributed cloud computing services industry in China and globally is competitive. We compete mainly on product functionality and scope, performance, service scalability and reliability, technical strengths, marketing and sales capabilities, user experience, pricing, brand awareness and reputation. In addition, emerging and enhanced technologies are likely to further intensify competition of our industry. For details, please refer to "Industry Overview—Competitive Landscape of China's Edge Cloud Computing Services" and "Industry Overview—Competitive Landscape of Chinese Independent AI Cloud Computing Service Providers."

EMPLOYEES

As of December 31, 2025, we had 275 full-time employees. The following table sets forth the number of our employees by function as of December 31, 2025:

	As of December 31, 2025	
Function	**Number**	**%**
Research and development	170	61.8
Sales and marketing	44	16.0
Operations	17	6.2
Administration and others	44	16.0
Total	275	100.0

We believe that our success depends on our ability to attract, retain and motivate qualified talents. We primarily recruit our employees through recruitment agencies, campus job fairs, internal referral programs and online channels, including our company website and social networking platforms. As part of our recruiting and retention strategy, we have established comprehensive training programs that cover topics such as corporate culture, employees' rights and responsibilities, team building, compliance and job performance.

We offer competitive salaries, performance-based promotion, incentive stock options, bonuses and other incentives. We believe that we maintain a good working relationship with our employees and we had not experienced any material labor disputes or any difficulty in recruiting staff for our operations during the Track Record Period and up to the Latest Practicable Date.

We enter into standard employment, confidentiality and non-compete agreements with our employees. We participate in housing provident funds and various employee social security plans that are organized by applicable local municipal and provincial governments, including housing, pension, medical, work-related injury and unemployment benefit plans.

PROPERTIES

Our headquarters is located in Shanghai, China. We did not own any properties as of the Latest Practicable Date. As of the December 31, 2025, we leased six properties in the PRC with an aggregate gross floor area of approximately 2,912.44 square meters. Our leased properties are primarily used as offices.

As of the Latest Practicable Date, our leased properties in the PRC are required by the applicable PRC laws and regulations to be registered and filed with the relevant land and real estate administration bureaus in the PRC, all of which had not been so registered or filed.

As advised by our PRC Legal Advisor, failure to complete the registration and filing of lease agreements will not affect the validity of the lease agreements or result in us being required to vacate the leased properties. However, the relevant PRC authorities may impose a fine ranging from RMB1,000 to RMB10,000 for each unregistered lease.

Having considered the foregoing, our Directors believe that the non-registrations of leases described above will not, individually or in the aggregate, materially affect our business and results of operation, on the grounds that: (i) no penalty had been imposed on us for our failure to register and file the relevant lease agreements during the Track Record Period and up to the Latest Practicable Date, (ii) we were advised by our PRC Legal Advisor that, if the lease registration can be completed in accordance with relevant laws and regulations within the prescribed time limit ordered by the competent governmental authorities, the risk of governmental authorities imposing a material penalty on us with respect to these leased properties is remote.

INSURANCE

Our employee-related insurance consists of pension insurance, unemployment insurance, work-related injury insurance and medical insurance, as required by PRC laws and regulations. In line with general market practice, we do not maintain any business interruption insurance or product liability insurance, which are not mandatory under PRC laws. During the Track Record Period, we did not make any material insurance claim in relation to our business.

We believe our insurance policy complies with the relevant rules and regulations in the PRC. For details, please refer to the paragraph headed "Risk Factors—Risks Relating to Our Business and Industry—We may not have sufficient insurance coverage to cover our potential liability or losses, and our business, financial conditions, results of operations and prospects may be materially and adversely affected should any such liability or losses arise." in this document.

LICENSES, PERMITS AND APPROVALS

During the Track Record Period and up to the Latest Practicable Date, we had obtained all requisite licenses, approvals and permits from relevant regulatory authorities that are material to our operations in China. The following table sets out a list of our material licenses and permits as of the Latest Practicable Date:

License/Permit	Holder	Granting Authority	Grant Date	Expiry Date
Value-Added Telecommunications Business Operating License (IDC/CDN/ISP) (增值電信業務經營許可證 (IDC/CDN/ISP))	PPIO Shanghai	Ministry of Industry and Information Technology of PRC (中華人民共和國工業和信息化部)	24 July, 2025	30 May, 2030

License/Permit	Holder	Granting Authority	Grant Date	Expiry Date
Value-Added Telecommunications Business Operating License (ICP) (增值電信業務經營許可證(ICP)) . . .	PPIO Shanghai	Shanghai Communications Administration (上海通信管理局)	May 26, 2025	July 17, 2030
Value-Added Telecommunications Business Operating License (ICP) (增值電信業務經營許可證(ICP)) . . .	Generative Era	Shanghai Communications Administration (上海通信管理局)	30 June, 2025	30 June, 2030

LEGAL PROCEEDINGS AND NON-COMPLIANCE

During the Track Record Period and up to the Latest Practicable Date, we had not been involved in any actual or pending legal, arbitration or administrative proceedings (including any administrative penalties, bankruptcy or receivership proceedings), which we believe would have a material adverse effect on our business, results of operations or financial condition. As of the Latest Practicable Date, we were not aware of any pending or threatened legal, arbitral or administrative proceedings against us or any of our Directors, which we believe would have a material adverse effect on our business, results of operations or financial condition.

During the Track Record Period and up to the Latest Practicable Date, we had not been involved in any material non-compliance incidents that we believe would have a material adverse effect on our business, results of operations or financial condition.

During the Track Record Period and as of the Latest Practicable Date, we had not experienced any major error, defect, security vulnerability or service interruption in our services and solutions, nor have we been subject to any material claims brought against us by any of our customers.

Non-compliance with the No. 32 Notice

During the Track Record Period, we obtained under-utilized computing resources from heterogeneous sources, including from entities that did not possess telecommunication service licenses (the "Incidents"). Our distributed computing technology enables us to allocate dispersed and under-utilized resources efficiently and provide cost-effective solutions for our customers. Given the large number of suppliers involved, we did not verify in every case whether each supplier held the relevant license. In 2023 and 2024 and the nine months ended September 30, 2025, we engaged 2,347, 3,723 and 2,314 unlicensed suppliers, respectively. The number of unlicensed suppliers decreased significantly during 2025, because we began terminating relationships with such suppliers during this period, and the termination process was completed by September 30, 2025. In 2023, 2024 and 2025, our computing resources procurement costs attributable to unlicensed suppliers were RMB151.4 million, RMB189.1 million and RMB154.1 million, respectively, representing 54.3%, 42.2% and 28.5% of our total procurement costs of the respective year.

As advised by our PRC Legal Advisor, (i) according to the "Notice of the Ministry of Industry and Information Technology on Rectifying and Regulating the Internet Access Service Market [2017] No. 32" (the "No. 32 Notice"), among other things, IDC, ISP, and CDN enterprises should not use bandwidth resources provided by entities that do not possess telecommunication service licenses. IDC, ISP, and CDN enterprises are required to conduct regular self-checks and rectify their operations where necessary. Our use of bandwidth resources in such Incidents does not constitute material non-compliance under PRC laws and regulations. Under the No. 32 Notice, such Incidents are not characterized as "illegal business operations" but rather as "conducts that require rectification." Our business also does not constitute "sub-leasing layer by layer" pursuant to the No. 32 Notice; (ii) during the Track Record Period and up to the Latest Practicable Date, we had not been subject to any administrative penalties by any competent authorities in the PRC; (iii) based on the telephone interview with the PRC competent authority in December 2025, the competent authority confirmed that we would not be subject to any penalties for the historical Incidents, as

long as we had completed the rectification by terminating with all unlicensed suppliers according to relevant laws and regulations; (iv) we have obtained all necessary telecommunications service licenses as legally required to operate our edge cloud computing service business; and (v) the contracts executed between us and our major clients during the Track Record Period remain legally binding and enforceable.

We have implemented rectification measures accordingly. As of September 30, 2025, we had terminated procurement from all of the suppliers who did not possess such licenses. In addition, we have adopted more stringent standards for supplier qualification review. Shanghai PAI was the entity responsible for the management of individual suppliers. We had terminated the procurement from all of those unlicensed individual suppliers and transferred the management of those licensed suppliers to PPIO Shanghai. Shanghai PAI's sole shareholder resolved to deregister Shanghai PAI in June 2025, and the deregistration was completed according to relevant laws and regulations as of the Latest Practicable Date.

Given that we currently maintain a diverse group of computing power suppliers and there is an abundance of underutilized computing resources in China, the Directors are of the view that the Incidents, as well as the termination of such suppliers, had not and will not have a material and adverse effect on our business and results of operations. Our revenue from edge cloud computing services increased by 18.9% from RMB547.7 million in 2024 to RMB651.1 million in 2025, and continued to operate steadily subsequent to December 31, 2025.

We have implemented the following enhanced internal control measures to prevent the recurrence of similar incidents going forward:

- We have adopted a computing power procurement management policy that sets out clear and stringent requirements for supplier admission and qualification review. Under this policy, our procurement department must conduct rigorous due diligence on potential computing power suppliers at the admission stage, including collecting and reviewing their business licenses, computing power service qualifications, technical capability certifications and other relevant qualification documents.

- After passing the initial qualification screening, potential suppliers are further shortlisted through market research, tendering or inquiry processes. Our evaluation covers price competitiveness (with a weighting of 40%), computing power performance, network quality and supporting facilities (with a combined weighting of 30%), and service responsiveness and operation and maintenance capability (with a weighting of 30%). Our procurement personnel also conduct on-site inspections of the data centers of shortlisted suppliers as part of the evaluation.

- Prospective computing power suppliers are then subject to joint technical testing and assessment by our technology department and operation and maintenance department. Only suppliers that pass this joint assessment may proceed to the contract negotiation stage.

- At the contract negotiation stage, our procurement department negotiates and defines the scope of services, after-sales support arrangements and contractual remedies for breach. The draft framework contracts are reviewed by our legal department before execution.

- Following approval by both our procurement department and technology department, our procurement department will register the approved computing power suppliers in our supplier management system for future procurement.

ENVIRONMENTAL, SOCIAL AND CORPORATE GOVERNANCE

ESG governance

We regard ESG matters as an integral part of our operations. We are committed to being a socially responsible company, continuously enhancing our ESG management practices and promoting our long-term sustainable development.

To strengthen our capabilities in managing ESG matters, we have established an ESG management system that is led, overseen and implemented in a unified manner by our Board. Our Board ensures that our ESG strategies are aligned with our overall business objectives and is responsible for reviewing and approving our ESG management policies and our ESG reports.

In addition, we have set up an ESG working group directly under our Board as the primary body responsible for the day-to-day handling and coordination of ESG-related matters. The ESG working group is responsible for understanding the demands, opinions and suggestions of our stakeholders, coordinating relevant internal and external work, studying material ESG topics, guiding the daily implementation of ESG initiatives and overseeing the preparation of our ESG reports. Our various functional departments and subsidiaries act as the executing units for ESG work. Based on our overall plans, they are responsible for implementing ESG tasks and reporting regularly on their progress.

ESG risk identification and assessment

We integrate ESG risks into our top-level corporate strategy design and have established clear ESG management principles and strategies. Our Board performs an oversight role and regularly reviews ESG matters to ensure that they are advanced in alignment with our business objectives, effectively controlling risks and safeguarding implementation outcomes.

Taking into account our business characteristics and stakeholder feedback, we have established a comprehensive system for identifying and assessing material ESG topics. By adopting a multi-dimensional scoring approach that covers indicators such as the degree of impact, likelihood of occurrence and time sensitivity, we systematically assess related risks and opportunities. Our Board regularly leads this process to ensure that risk prioritization is robust and that our response strategies are well targeted, providing continuously optimized action guidance for our ESG risk management. Through this mechanism, we are able to identify priority risks in a timely manner and formulate corresponding response measures.

Environment

We have built a global computing power scheduling platform based on distributed computing technologies. By integrating idle computing resources and optimizing the allocation of supply and demand, we help to alleviate structural imbalances in global computing capacity. Through this platform, we standardize heterogeneous computing power into a comprehensive service portfolio covering cloud computing through to artificial intelligence inference, in order to meet the diversified needs of our customers.

At the same time, we recognize the profound impact of climate change on both ecosystems and business. We have therefore embedded climate action into our core strategy and are committed to reducing our carbon footprint through technological innovation, and to promoting our computing power network as a green and sustainable cornerstone infrastructure for the digital future.

Climate risk management

We recognize the close interconnection between our business and climate change and have identified and assessed climate-related risks that may affect our operations, supply chain and stakeholders. These risks include:

Type of risk		Potential impact	Management measures
Physical risks . .	Acute climate risks such as typhoons and floods, as well as chronic risks such as rising sea levels and long-term changes in climate patterns.	Disruption to employee commuting, damage to office infrastructure, interruption of computing facilities and services, and adverse impact on business continuity and operations.	We continue to refine our emergency response management policy, and to optimize emergency plans and recovery measures to address extreme weather events, safeguard employee health and safety, and ensure the stability of our products and services during customer use.
Transition risks . .	Policy risks arising from tighter carbon emission regulations, and reputational risks resulting from stakeholders' increased preference for green development.	Potential operational challenges and higher compliance costs, as well as pressure to align our products and services with low-carbon standards.	We continuously monitor changes in the regulatory environment, actively engage with stakeholders, and develop and implement more sustainable production and operating models so as to meet evolving regulatory and market expectations.

We provide direction for our environmental management work by setting environmental targets. We have established an ESG target review mechanism which requires management to review our targets on a regular basis. With reference to domestic and international ESG standards, industry peers and our own development plans, we have formulated the following environmental targets:

- Greenhouse gas emission reduction target: by 2030, reduce Scope 1 and Scope 2 greenhouse gas emissions per unit of revenue by 10 per cent compared with 2024.
- Energy efficiency target: by 2030, reduce energy consumption per unit of revenue by 10 per cent compared with 2024.
- Water efficiency target: by 2030, reduce water consumption per unit of revenue by 10 per cent compared with 2024.
- Waste reduction target: continue to implement green office practices and reduce the generation of waste.

Environmental and climate-related performance indicators

We include both our office premises and our computing power centers within the boundary for our climate and environmental performance indicators. Given the nature of our business, we do not engage in physical product manufacturing and do not operate our own vehicle fleet. As a result, we do not generate air pollutants or consume packaging materials.

Indicator	Unit	Year ended December 31, 2023	Year ended December 31, 2024	Year ended December 31, 2025
Total greenhouse gas emissions[1]	tons of CO_2 equivalent	120.78	132.94	183.25
- Scope 2 greenhouse gas emissions	tons of CO_2 equivalent	120.11	131.99	182.02
- Scope 3 greenhouse gas emissions[2]	tons of CO_2 equivalent	0.67	0.95	1.23
Greenhouse gas emissions per unit of revenue	tons of CO_2 equivalent/RMB10,000	0.34	0.24	0.24
Indirect energy consumption[3]	kWh	226,365.30	248,749.40	343,047.56
Total energy consumption	kWh	226,365.30	248,749.40	343,047.56
Total energy consumption per unit of revenue	kWh/RMB10,000	631.62	445.76	445.30
Total wastewater discharge	tons	740.0	1,055.0	1,365.00
Total wastewater discharge per unit of revenue	tons/RMB1,000,000	2.06	1.89	1.77
Total municipal water consumption	tons	925.00	1,318.75	1,706.25
Municipal water consumption per unit of revenue	tons/RMB1,000,000	2.58	2.36	2.21

Notes:

(1) We do not own any vehicles and therefore Scope 1 greenhouse gas emissions are not applicable to us. For the calculation of our electricity-related greenhouse gas emissions, we adopt the grid emission factor published by the Ministry of Ecology and Environment of the People's Republic of China in the Announcement on the Release of the 2023 Power Grid Carbon Dioxide Emission Factors. The carbon dioxide emission factor for electricity is 0.5306 kg CO_2/MWh.

(2) The calculation boundary for our Scope 3 greenhouse gas emissions covers greenhouse gas emissions generated from the treatment of wastewater discharged by us.

(3) We do not own any vehicles and we are not engaged in any physical product manufacturing activities. Our energy consumption consists solely of electricity used in our office premises, including lighting, air conditioning and office equipment.

Our business does not involve physical manufacturing and our impact on the environment is relatively limited. Nevertheless, we remain committed to green development, and strictly comply with environmental laws and regulations such as the Environmental Protection Law of the People's Republic of China. We monitor compliance on an ongoing basis and ensure that our operations meet applicable requirements. In addition, we have formulated dedicated environmental protection policies to manage environmental matters in our offices in a systematic manner. We strictly comply with the Law of the People's Republic of China on the Prevention and Control of Environmental Pollution by Solid Waste and other waste management laws and regulations, and we ensure the compliant treatment and disposal of waste. Given the nature of our business, the waste generated in our operations mainly consists of domestic waste and office waste. All non-hazardous waste is collected and treated in compliance with applicable requirements by the property management companies of our office buildings, and all hazardous waste is treated in compliance with applicable requirements by qualified third-party service providers engaged by us.

Social

Compliance in employment

We have formulated and continuously improved our employee handbook and other personnel management policies to regulate the entire process of recruitment and employment. In our recruitment activities, we follow the principles of fairness, impartiality, openness, suitability and merit-based selection. We have established evaluation criteria that are driven by job requirements and focus on work capabilities and job fit. We do not differentiate among candidates on the basis of gender, age, ethnicity, religion or other personal characteristics, and we endeavor to safeguard equal opportunities and diversity.

Occupational health and safety

We care about the long-term interests of our employees and are committed to building a comprehensive health and safety protection system. We enroll all employees in basic medical insurance and purchase supplemental commercial medical insurance, and we organize regular health check-ups for all employees to help identify potential health issues at an early stage. We also equip our offices with first-aid kits and basic emergency medicines, and we have established an emergency response mechanism for unexpected safety incidents to ensure that employees can receive timely on-site assistance when needed. As of the end of the reporting period, we had not experienced any work-related injuries or fatalities.

Remuneration and benefits

We have established a fair, reasonable and market-competitive remuneration management system to ensure that remuneration is aligned with job responsibilities, individual performance and contribution. We strictly comply with national requirements on minimum wage standards, working hour systems and social insurance, and make timely and full contributions to all mandatory social insurance schemes and housing provident funds for our employees, so that they can obtain labor remuneration and related social security entitlements in accordance with the law.

We continuously improve our employee care mechanisms and support employees in balancing work and life during their career development through diversified benefits and cultural activities. By organizing team-building activities, health protection programs and measures to facilitate daily life, we enhance employees' sense of belonging and team cohesion, and foster a positive, resilient and cohesive organizational culture.

Employee development and training

We regard talent development as an important foundation of our sustainable competitiveness. We place a strong emphasis on building our talent pipeline and supporting the long-term growth of our employees, and we are committed to fostering an organizational ecosystem in which employees and the company grow together. To this end, we have established a talent development system that is tiered and categorized, covers the full career cycle and is practice-oriented, and we have put in place a systematic capability development tracking mechanism. We continuously focus on enhancing employees' professional skills and overall competencies to ensure that talent development is aligned with our strategic development direction.

For the years ended December 31, 2023, 2024 and 2025, we had total full-time employee training attendances of 1,718, 840 and 1,125, respectively, and the proportion of employees receiving training was 62%, 80% and 69% for those years.

Supply chain management

We regard our suppliers as important strategic partners and is committed to building a long-term, stable, collaborative and mutually beneficial supply chain ecosystem. We have formulated internal policies such as supplier management implementation rules and procurement management regulations to regulate our procurement and supplier management activities in a systematic manner and to implement our compliance management requirements.

We have established a procurement governance structure led by the president's office and supported by cross-functional collaboration. This structure horizontally connects the procurement management committee, review panels and key departments such as asset management, business management, quality management and demand fulfilment, and vertically clarifies the division of responsibilities for decision-making, execution and supervision at different levels. This creates a management closed loop with clearly defined responsibilities and efficient collaboration. In supplier management, we implement a tiered and categorized control framework and carry out refined management based on suppliers' risk levels and strategic importance to ensure that control measures are precise and effective, and to enhance supply chain compliance and quality assurance capabilities from the outset.

Product responsibility

We strictly comply with the Product Quality Law of the People's Republic of China and other applicable national laws and regulations, and adheres to the principles of compliant operation and prudence throughout our research and development and marketing activities. We place business ethics and compliant operation at the core of our corporate development and strictly observe laws, regulations and industry norms in all business activities, thereby continuously strengthening the foundation of our product responsibility management. To systematically implement national requirements and industry standards, we have formulated and implemented R&D project management measures to standardize the entire process of product research and development. This ensures that each stage, from requirement analysis and design and development to testing and validation, is subject to institutionalized and standardized management, and provides a solid guarantee for product quality and market compliance.

Anti-corruption

We regard business ethics and compliant operation as the foundation of our corporate development. We regulate business conduct through systematic internal control processes and technical measures, and resolutely oppose bribery, fraud, money laundering and other behavior that violate business ethics or laws and regulations. We strictly comply with laws and regulations such as the Criminal Law of the People's Republic of China and the Anti-Money Laundering Law of the People's Republic of China, and have established and implemented internal policies such as the Anti-Fraud, Anti-Money Laundering and Anti-Bribery Management Policy. These policies apply to all employees, and we impose higher compliance requirements on directors, supervisors, senior and middle management and personnel in key positions, thereby supporting our continuous, stable and healthy development.

We have set up an internal audit system directly led by the Board of Directors to oversee the prevention, supervision and handling of business ethics and compliance risks. This system operates with independence and authority and systematically identifies potential risks through regular audits, special inspections and process controls. For major or complex matters, we engage external professional institutions to participate in investigations to enhance objectivity and credibility.

We conduct systematic risk assessments each year. By taking into account developments in our business, we identify potential risk points, dynamically review the effectiveness of existing control mechanisms and formulate improvement measures to address weaknesses in processes. At the same time, we have put in place a misconduct handling mechanism that covers investigation, accountability, rectification and remediation, enabling rapid response and effective control of risks. In determining responsibility, we differentiate between leadership responsibility and direct responsibility and hold both individuals involved in violations and relevant management personnel accountable in accordance with applicable rules, thereby encouraging departments to strengthen daily control.

RISK MANAGEMENT AND INTERNAL CONTROL

We are committed to developing and maintaining robust risk management and internal control systems tailored to our business operations, with a continuous focus on enhancing their effectiveness. We continually review the implementation of our risk management and internal control policies and procedures to enhance their effectiveness and sufficiency.

Financial Reporting Risk Management

Our financial reporting risk management includes a comprehensive set of accounting policies. We have established procedures to effectively implement these policies, and our financial department regularly reviews management accounts based on these procedures. Additionally, we provide ongoing training to our finance department employees to ensure they are well-versed in and can effectively apply our financial management and accounting policies in our daily operations.

Internal Control Risk Management

To ensure compliance with applicable regulations and internal standards, we have instituted stringent internal procedures. Our compliance team collaborates closely with the finance and business departments to: (a) perform risk assessments and advise risk management strategies; (b) improve business process efficiency and monitor internal control effectiveness; and (c) promote risk awareness throughout our Company. We maintain rigorous internal procedures to secure all necessary licenses, permits and approvals for our operations, with regular reviews by our internal control team to monitor the status and effectiveness of these authorizations. Our compliance team also coordinates with relevant departments to secure the necessary governmental approvals or consents for filings with appropriate authorities.

Human Resources Risk Management

We conduct regular and specialized training tailored to the diverse needs of our departments, ensuring that our staff's skills are current and aligned with our customer service objectives. We provide our employees with an employee handbook that outlines internal rules and guidelines on best commercial practices, work ethics, fraud prevention, negligence and corruption. Additionally, we have established a code of business conduct and ethics, and an anti-bribery and corruption policy. These guidelines outline best commercial practices and work ethics, providing clear anti-bribery guidance and measures. We maintain an open internal reporting channel for our staff to report any wrongdoing or misconduct, ensuring that all reported incidents and individuals are investigated, with appropriate actions taken based on the findings.

AWARDS AND RECOGNITIONS

During the Track Record Period, we received awards and recognitions in respect of our services, technology and innovation. The following table sets out the details of some of the notable awards and recognitions which we have received:

Award/Recognition	Award Year	Awarding Institution/Authority
InfoQ "2025 AI Infrastructure Excellence Award" (極客傳媒「2025年度AI基礎設施卓越獎」)	2025	InfoQ 極客傳媒
Quantum Bit "2025 AI Annual Outstanding Product" (量子位「2025人工智能年度傑出產品」)	2025	Quantum Bit 量子位
Quantum Bit "2025 AI Annual Potential Startup" (量子位「2025人工智能年度潛力創業公司」)	2025	Quantum Bit

Award/Recognition	Award Year	Awarding Institution/Authority
36Kr "WISE 2025 Business King – Annual Most Commercially Promising Enterprise" (三十六氪「WISE 2025商業之王年度最具商業潛力企業」)	2025	36Kr 三十六氪
Quantum Bit "AIGC Companies to Watch in 2025" (量子位「2025年值得關注的AIGC企業」)	2025	Quantum Bit
Quantum Bit "2024 AI Annual Potential Startup" (量子位「2024人工智能年度潛力創業公司」)	2024	Quantum Bit
Zhongguancun Intelligent AI Research Institute "2024 AI Application Benchmark" (中關村智用人工智能研究院「2024人工智能應用標桿」)	2024	Zhongguancun Intelligent AI Research Institute 中關村智用人工智能研究院
"2024 China Tech Industry Leading List" & Key Case in "China AI Computing Power Industry Development Report" (「2024中國科技產業引領榜」及《中國AI算力行業發展報告》重點廠商案例)	2024	Jazzyear 甲子光年
SegmentFault "2024 China Emerging Technology Pioneer Enterprise List" (SegmentFault「2024中國新銳技術先鋒企業榜單」)	2024	SegmentFault
CAICT 2024 Annual Distributed Computing Power "Star Shining" Case (中國信通院2024年度分布式算力「星耀」案例)	2024	CAICT 中國信通院
CAICT 2023 Annual Computing Network Infrastructure Excellent Case (中國信通院2023年度算網基礎設施優秀案例)	2024	CAICT
IDC China Edge Cloud Market Report (Top Ten) (IDC中國邊緣雲市場報告(前十))	2024, 2023	IDC
CAICT Edge Computing Pioneer Enterprise (中國信通院邊緣計算先鋒企業)	2023	CAICT