11 May 2026 06:20:35 ET | 60 pages

Alibaba Group Holding (9988.HK, BABA)

Deep Dive

Decoding Alibaba Cloud AI Full Stack & Introduce MaaS Revs Forecast

Alicia Yap, CFA AC
+852-2501-2773
alicia.yap@citi.com

Nelson Cheung
+852-2501-2728
nelson.cheung@citi.com

CITI'S TAKE

This report analyzes Alibaba Cloud's vertically integrated AI stack, from its T-Head semiconductor offering to its IaaS, PaaS, and MaaS layers. Our MaaS analysis focuses on Qwen LLM, Model-Scope platform, and progress in application and agentic AI. We forecast AI-related revs to grow at 90% CAGR from FY26-31E to reach 70% of total cloud revs by FY2031 with MaaS revs growing at 235% CAGR to reach Rmb438.6bn, representing 53% of total cloud revs by FY2031. We reiterate our Buy rating on BABA with TP of US$205/HK$204 for ADR/H shares respectively and anticipate significant value-unlocking opportunities as Alibaba demonstrates the improving strength of its AI full-stack capabilities with cost synergies and improving cloud margins potential riding on continuous AI innovation. We believe Alibaba is well-positioned to capitalize on the rapid growth of the token economy and reaffirm its status as our top pick for China AI play.

Raising cloud revs forecast - We raise our FY2028-2029 cloud revs assumption and introduce FY30-31 forecast and now model consolidated cloud revs to grow at 39% CAGR from FY26-31E to reach Rmb833.4bn by 2031E with external cloud revs growing at 44% CAGR to Rmb666.8bn or US$95.2bn (vs company target of US$100bn) by 2031E. This is driven by 90% CAGR for AI-related revs to reach Rmb585.5bn or 70% of total while non-AI cloud revs to grow at 13% CAGR with revs contribution declining from 85% in FY26 to 30% in FY31.

MaaS revs to reach 53% of total by 2031 - We forecast MaaS related revs to grow at 235% CAGR to reach Rmb438.6bn (US$62.6bn) by 2031 and accounted for 53% of total cloud revs. Thanks to BABA's multi-chip architecture strategy and contribution from T-Head AI accelerator, we forecast AI-IaaS revs to grow at 46% CAGR to Rmb128.7bn and account for 15% of total cloud revs by 2031 while we model AI-PaaS revs to grow at 40% CAGR to Rmb18.2bn but only represent 2% of total cloud revs.

Cumulative capex spend – We forecast capex to be maintained at Rmb130-140bn per year with cumulative capex estimated at Rmb822.9bn for six years (FY26-31) and our estimated capex spend suggest avg of 32% of cloud revs for six years period.

															Current Fiscal Year		Next Fiscal Year	
						Rating		Short-Term	Target Price						EPS		EPS	
Company	Ticker	Ccy	Price	Mkt Cap (M)	Date & Time	Old	New	View	Old	New	ESPR (%)	Div Yld (%)	ETR (%)	Last Rpt Yr	Old	New	Old	New
Alibaba Group Hldg	9988.HK	HK$	139.00	2,667,744	08 May 16:10	1	nc	-	204.00	nc	46.8	1.4	48.2	Mar-25	3.837	nc	6.121	nc
Alibaba Group Hldg	BABA	US$	140.06	336,011	08 May 16:00	1	nc	-	205.00	nc	46.4	1.4	47.8	Mar-25	30.693	nc	48.236	nc

1 = Buy, 2 = Neutral, 3 = Sell, H = High Risk
ESPR = Expected Share Price Return, ETR = Expected Total Return, nc = no change
Source: Citi Research
^Catalyst Watch

See Appendix A-1 for Analyst Certification, Important Disclosures and Research Analyst Affiliations.

Not for distribution in the People's Republic of China, excluding the Hong Kong Special Administrative Region and Qualified Foreign Institutional Investors.

www.futoucaixin.cn

Earnings Estimates

				Last Reported Year					Current Fiscal Year					Next Fiscal Year				
Company Name	Ticker	Last Rpt Year	Currency	1Q	2Q	3Q	4Q	FY0	1Q	2Q	3Q	4Q	FY1	1Q	2Q	3Q	4Q	FY2
Alibaba Group Hldg	9988.HK	Mar-25	Rmb	2.089	1.907	2.790	1.580	8.176	1.844	0.545	0.886	0.564	3.837	1.262	1.369	2.038	1.454	6.121
Alibaba Group Hldg	BABA	Mar-25	Rmb	16.474	15.255	22.317	12.638	65.406	14.749	4.361	7.089	4.514	30.693	9.945	10.787	16.059	11.460	48.236

Source: Citi Research

www.futoucaixin.cn

Contents

BABA's AI Full-Stack Capability 4
The Rmb380bn capex investment 5
History of Alibaba Cloud 6
Stages of Ali Cloud revenues and growth trend 6
US$100bn revenue target and MaaS as key driver 8
Raising Cloud revenues forecast 9
AI-related and MaaS to reach 70%/53% of total by 2031 9
Cumulative Capex spend 10
Decoding AliCloud's AI Full-Stack 11
Silicon (Chips) Layer - T-Head (Pingtouge) 11
Background of T-Head Unit 11
Chip architectures by T-Head 12
Alibaba Cloud Infrastructure 15
Domestic and international data centers 15
Apsara full-stack cloud operating system 17
Platform-As-A-Services (PaaS) 22
Alibaba Token Hub (ATH) 23
Tongyi Lab (Tongyi Large Model Business Unit) 24
Model-as-a-Service (MaaS) 24
Qwen Business Unit 25
Wukong: AI-Native Agentic Platform 25
AI Innovation Business Unit 25
Foundation model – Qwen series 26
Qwen model series & benchmark ranking comparison 28
AI Application integration with Ali's ecosystem 31
Agentic AI development for AliCloud 31
Token Economies and usage scale 34
Global foundation model market size 35
China AI market TAM 36
Cloud industry market size and market share 40
Token usage forecast by IDC 42
Our forecast of China public cloud MaaS size 42
Bull/Bear: Alibaba (9988.HK) 47
Bull/Bear: Alibaba Group Holding (BABA.N) 48
Alibaba 49
Company description 49
Investment strategy 49
Valuation 49
Risks 49
Alibaba Group Holding 49
Company description 49
Investment strategy 49
Valuation 50
Risks 50
Appendix A-1 51

www.futoucaixin.cn

BABA's AI Full-Stack Capability

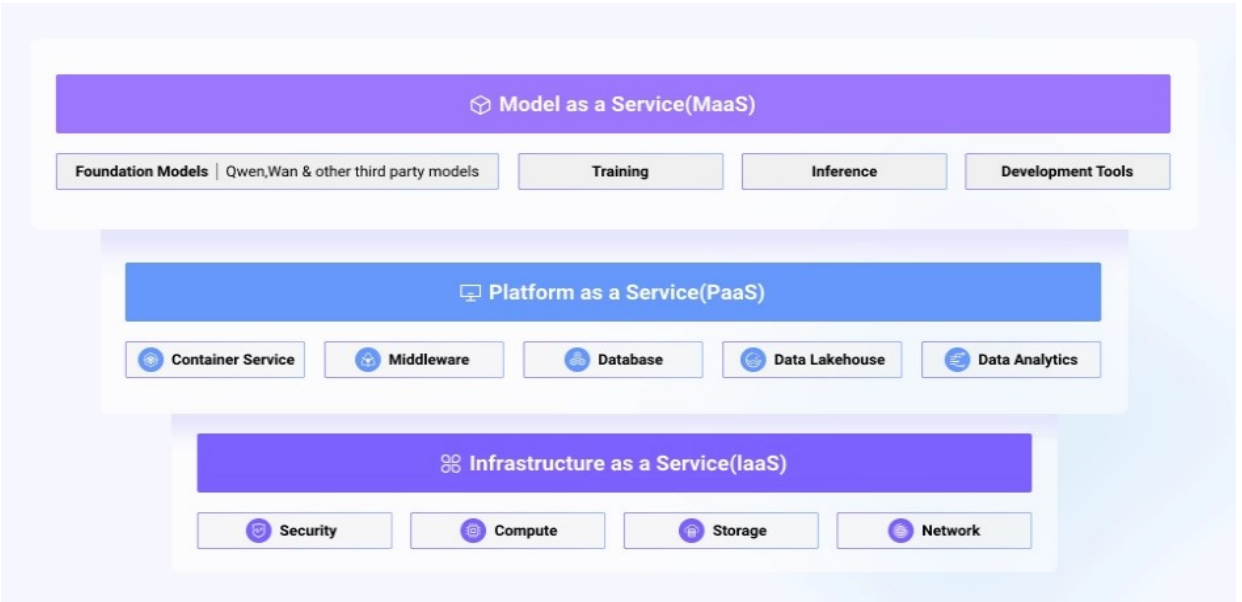
As highlighted by CEO Mr. Eddie Wu during latest FY3Q26 earnings call that Alibaba has developed a comprehensive, **full-stack AI capability** to meet the exponential growth in AI demand. Its AI strategic roadmap provides end-to-end coverage, from the AI infrastructure layer (chips and cloud computing) to the application layer. The latter is anchored by the Alibaba Token Hub and includes the Qwen foundation models, Model as a Service (MaaS), and various enterprise and consumer applications, creating a fully integrated AI ecosystem.

The success of this full-stack strategy is evident in the performance of Alibaba's **infrastructure layer**. Driven by sustained, strong AI demand, the Cloud Intelligence Group's revenue from external customers accelerated to 35% in FY3Q26, with AI-related product revenue delivering triple-digit year-over-year growth for the tenth consecutive quarter. This momentum has expanded its market share to 36% and pushed its cumulative external revenue through February for FY2026 surpassed Rmb100bn, per management.

A primary catalyst for this expansion was attributed to MaaS platform, where token consumption has grown six-fold in the last three months, positioning it to become the Cloud Intelligence Group's largest revenue product. Underpinning this growth is the T-Head division's proprietary AI chips, which have achieved cumulative shipment of 470,000 units. With >60% of these chips serving >400 external enterprise customers, T-Head is providing the critical compute capacity that strengthens the overall competitiveness of Alibaba's cloud and MaaS offerings.

The **application layer** comes with core mission of creating, delivering, and applying tokens. With the establishment of new Alibaba Token Hub (ATH) business group, which comprised of the Tongyi Lab, the MaaS Business Line, the Qwen Business Unit, the Wukong Business Unit, and the AI Innovation Business Unit, ATH aims to provide the organizational foundation for executing Alibaba's AI strategy and serves as the hub for efficient coordination across its various AI businesses.

Figure 1. AliCloud all-in-one full stack services



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports, Citi Research

www.futoucaixin.cn

Driven by strong market demand and confidence in its full-stack capabilities, Alibaba's management is targeting significant growth opportunities in its cloud and AI businesses. The company has set a goal to **achieve US$100 billion** in combined cloud and AI external revenue within the next five years. Management expects MaaS to be the primary engine for this growth, projecting it will contribute the majority of the growth and become the largest revenue component by that time. We noted in FY3Q26 review note that by targeting AI and cloud business revs to >US$100bn in five years, driven by capitalizing on the enterprise adoption of large AI models for complex B2B workflows with key growth drivers including MaaS as the primary growth engine, bespoke AI and cloud solutions for enterprise-specific training and inference and an expansion of traditional CPU-centric cloud computing to support the operational needs of AI agents. By transitioning from selling resources to **selling "intelligence**" and leveraging proprietary T-Head chips for efficiency, management expects **non-linear growth in both revs and profitability.**

In fact, Ali management has commented on its full-stack capability several times in the past, this includes the keynote speech during 2025 Apsara Conference (on Sept 29, 2025), where Chairman and CEO Eddie Wu commented that "*Alibaba Cloud is strategically positioned as a full-stack AI service provider, dedicated to delivering robust computing with maximized efficiency for training and deploying large AI models on the cloud.*" During the Apsara conference, BABA also disclosed that as of Sept 2025, over 600mn models downloads of Qwen and Wan series model, with 170,000 derivative models and more than 800,000 AI agents build on Model Studio and over one million corporate and individuals using Qwen through Model Studio.

The Rmb380bn capex investment

Alibaba first provided its three-year Rmb380bn capex plan in Feb 2025, framing it as a "once-in-a-generation" opportunity presented by Artificial Intelligence. While it was initially set as a three-year target, recent management comment suggest that plan will be accelerated.

During the 2025 Apsara keynote speech (Citi note), Mr. Eddie Wu also highlighted that the AI-driven intelligent revolution is moving the transformation from AGI to ASI (Artificial SuperIntelligence) where LLMs will be the operating system in the AI era and AI Cloud will become the next generation of computers. Mr. Wu also commented that with the demand for AI infrastructure accelerating, Ali is committed to increasing its global **datacenter capacity by tenfold by 2032** to support the evolving ASI era. Mr. Wu also noted that only 5-6 global super cloud platforms are likely to exit to provide large-scale computational resources/AI services, we believe, with AliCloud being a world leading full-stack Ai service provide, Alibaba is well-positioned to capture the evolution upside translating to sustainable cloud revs growth with efficiency margins upside in coming years.

During FY2Q26 (Nov 2025) earnings call, Alibaba management noted that given supply chain constraint, mgmt expects AI demand will continue to outpace capacity in next 2-3 years, and its current 3-year Rmb380bn capex budget likely to deploy ahead of schedule.

www.futoucaixin.cn

History of Alibaba Cloud

Alibaba Cloud was established by Alibaba Group in Sept 2009, originally aiming to support the internal infrastructure of Alibaba's ecommerce ecosystem such as handling of vast amounts of data and transactions especially during Singles Day promotion. Since its inception, Alibaba Cloud has grown to become the largest cloud computing company in China and the 4th largest in the world, per BABA 2025 Annual report. As highlighted in FY2025 annual report (link), Alibaba Group is the world's fourth largest and Asia Pacific's largest Infrastructure-as-a-service provider by revenue in 2024 in U.S. dollars, according to Gartner April 2025 report.

Between 2010 and 2014, Alibaba Cloud had transitioned from an internal solution to a public cloud services provider, offering elastic computing, data storage, and analytics to external clients. A key development during this time was the 2013 debut of Apsara, its proprietary cloud computing engine, which became the foundation for its scalable services in web hosting, big data, and machine learning.

Alibaba Cloud began its global expansion in 2015 by opening a data center in Silicon Valley, signaling its intent to compete with established international players like AWS and Microsoft Azure at that time (but BABA has since scaled back its presence in the US since 2018 amid political tension as noted by CIO Dive article dated Sept 4, 2018). Gartner included Alibaba Cloud in its Magic Quadrant for Cloud Infrastructure as a Service (IaaS) in 2016, which suggests an important recognition. In 2017, Alibaba announced the 12-year partnership with international Olympic Committee to become the official "Cloud Services" and "Ecommerce Platform Services" Partner as well as a founding partner of the Olympic Channel. Since 2019, Alibaba Cloud has intensified its focus on artificial intelligence, launching the Machine Learning Platform for AI to provide developers with efficient tools for building and deploying AI models, positioning itself as a leader in the AI sector.

Stages of Ali Cloud revenues and growth trend

Since the inception in 2009, Alibaba Cloud has been through a few transition stages where we will discuss briefly the revenues scale during its early stage (FY2010-2014), fast growth stage (FY2015-2018), normalization growth (FY2019-2021), challenging stage (FY2022-2024), reaccelerating stage (FY2025-FY2026), AI high-growth stage (FY2027-2028).

- **Early stage (FY2010-2014) –** Based on the reported financial, cloud revs were Rmb144mn in 2010 and grew to Rmb425mn in 2011 before experiencing slow growth of 19% yoy to Rmb773mn in 2014 or 1.5% of total group revs. As noted in earlier part of this note, during this period, Alibaba Cloud had transition from an internal solution to a public cloud service provider offering elastic computing, data storage to external clients.

- **Fast growth stage (FY2015-2018) –** During this period, AliCloud also started to expand its services to global markets and cloud revs experienced fast acceleration with 64%/138%/121%/101% yoy growth from FY2015 to 2018, respectively, demonstrating fast ramp of cloud customers and revs generated per customer. Ali first reported the breakdown of segment EBITA in 2016, AliCloud incurred a loss of Rmb2.65bn or -86.3% of revs in 2016. The loss ratio improved to -25%/-23% in FY2017-2018.

- **Normalization stage (2019-2021) –** After the strong growth stage, revenues growth had become more normalized in subsequent few years on the back of

www.futoucaixin.cn

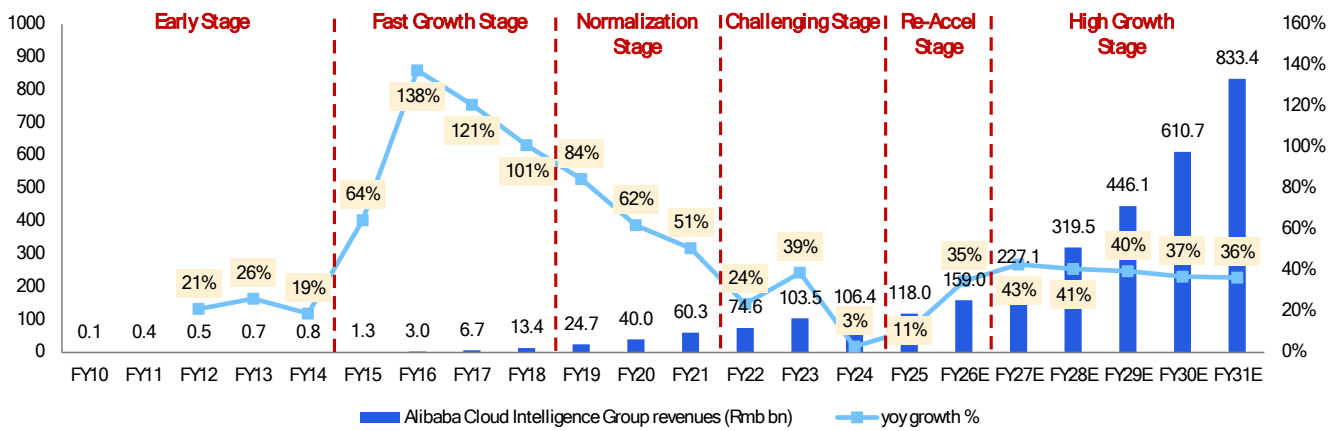
higher base, growing competition. During this stage, revs growth decelerated to 84%/62%/51% for FY2019-2021.

- **Challenging stage (FY2022-2024) –** In FY4Q21, cloud revs saw a big deceleration to 38% yoy, down from 50-60% from previous few quarters, mainly related to the transition of an Internet company's overseas cloud services from Alibaba Cloud to US cloud service provider due to the geo-political tension. Caixin reported on May 14, 2021, that the "internet" company is Bytedance's TikTok business. The transition had resulted in fast deceleration of revs growth to 24% in FY2022. With Bytedance launching its own enterprise-level cloud service Volcano Engine on June 10, 2021 (Reuters) we reconcile Bytedance had subsequently also moved out its domestic cloud business from Alibaba Cloud to its own platform, together with intense pricing competition in China and BABA's decision to also cutting back "low-margin" project-based cloud customers, AliCloud revs decelerated sharply to 4%/3% in FY2023/2024. AliCloud turned adjusted EBITA profitable in FY4Q22 with Rmb598mn or 3.2% margin. In FY2023-2024, adjusted EBITA reached Rmb4.1bn/Rmb6.1bn or 4%/5.8% margin.

- **Reaccelerating stage (FY2025-2026) –** After the challenging period, cloud revs started to show reacceleration with quarterly yoy growth continued to accelerate from 5.7%/7%/13%/18% in FY1Q25-4Q25 and further accelerate to 26%/34%/36% in FY1Q26-3Q26 and we currently model cloud revs to grow at 41% yoy in FY4Q26. The reacceleration growth mainly driven by 10 quarters of triple digit AI-related revs growth which accounted for ~20% of total cloud revs as of FY1Q26 as disclosed by the company. Adjusted EBITA margins further improved to 8.9% in FY25 and stabilized around 9% in FY2026.

- **High-growth stage (FY2027-2028)** - On FY3Q26 earnings call, management provided US$100bn revs target in next five years, which suggests revenues CAGR of >40% for the next five years. Nonetheless, consensus and we have raised our forecast for cloud revs growth to exceed 40% in next couple years, and we still model some deceleration in subsequent years based on our conservative view.

Figure 2. Alibaba Cloud Intelligence Group revenues

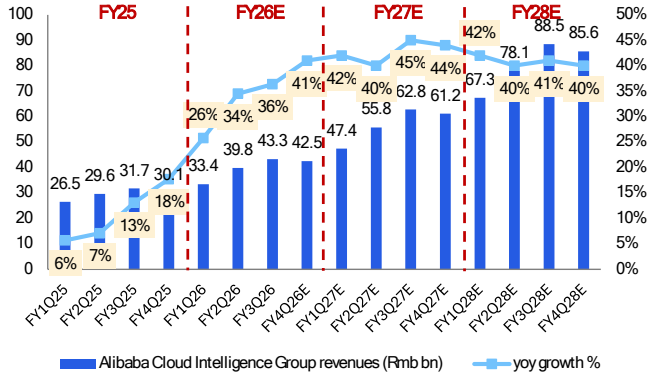


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports and Citi Research Estimates, Note: Alibaba changed its Cloud reporting revs from external to consolidated starting from June 2023 (FY1Q24)

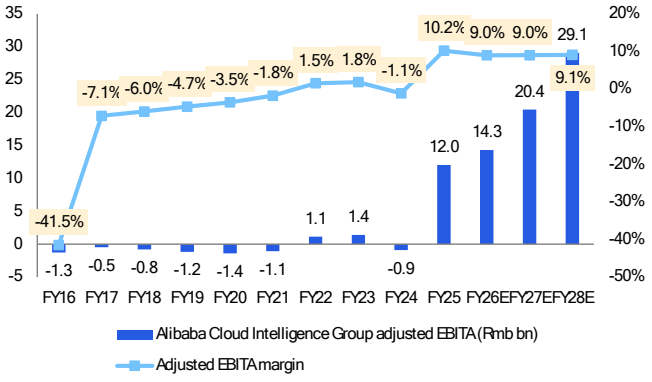
www.futoucaixin.cn

Figure 3. Alibaba Cloud Intelligence Group revenues by quarter



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports and Citi Research Estimates

Figure 4. Alibaba Cloud Intelligence Group adjusted EBITA



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports and Citi Research Estimates

External vs internal Cloud revs

Historically, BABA only reported external customer revs for its cloud revs, starting from FY1Q24 (or June quarter 2023), BABA changed its reporting to consolidated cloud revs, including both external and internal (ie: cloud demand from its ecommerce and other internal business units). We reconcile internal revs account for 25-30% of total cloud revs. Since the reacceleration of cloud revs starting FY1Q25, external revs have been growing slightly slower than total cloud revs growth, suggesting internal revs were growing slightly faster. For example, in FY4Q25, total cloud revs grew 18% while external customer grew 17% yoy and FY2Q26 total cloud revs grew by 34% yoy while external customer grew 29% yoy, mainly because internal business segment have been the early adopter of AI model and AI agentic tools, etc to strengthen business operation and optimize cost, hence internal demand has been very strong and make sense during nascent stage of AI revs take off, internal demand might be growing faster than external.

AI-related revs growing at triple digit yoy

As early as FY1Q26 (Aug 2025), BABA noted on its earnings call that AI related revs accounted for 20% of total cloud revs and with AI related revs growing at triple digit for 8th consecutive quarter. Later during FY2Q26 earnings in Nov 2025, BABA reiterated that AI related revs accounted for more than 20% of external customer revs with contribution continue to increase driven by accelerated adoption of AI product across broad range of enterprise customers with a growing focus on value added applications, including coding assistants. And in latest FY3Q26 earnings call in Mar 2026, BABA noted that AI-related revs continue to grow at triple digit yoy for 10th consecutive quarter. We traced back starting from FY4Q24 earnings call (May 2023), BABA commented that AI-related revs increased triple digit yoy, and specifically during that quarter, BABA noted that core public cloud offerings recorded double-digit yoy growth demonstrating that they are returning to growth.

US$100bn revenue target and MaaS as key driver

As noted above, during FY3Q26 earnings call, BABA management announced its goal to achieve US$100 billion in combined cloud and AI external revenue within the next five years with MaaS being the primary engine for this growth and projecting MaaS to contribute the majority of the growth and become the largest

www.futoucaixin.cn

revenue component by that time, implying over 40% revs CAGR for the next five years. We believe market remains skeptical whether the 40% revs CAGR is achievable, and we previously were modeling deceleration starting FY2029E after forecasting >40% for FY2027E and FY2028E.

In light of the continued acceleration of enterprise demand and continued improvement of AI model intelligence (including Qwen series) and declining token prices, we believe the adoption rate will further accelerate in the coming years. Combining top-down and bottom-up analysis, we raise our cloud revs forecast for FY2028 and FY2029 and introduce our forecast for FY2030 and FY2031.

Raising Cloud revenues forecast

AI-related and MaaS to reach 70%/53% of total by 2031

We now forecast consolidated cloud revs to grow at 39% CAGR from FY26-31E to reach Rmb833.4bn by 2031E with external cloud revs to grow at 44% CAGR to Rmb666.8bn or US$95.2bn (close to mgmt target of US$100bn) by 2031E.

Within the total consolidated revs, we forecast AI-related revs to grow at 90% CAGR to reach Rmb585.5bn or 70% of total cloud revs by 2031E (up from 15% in FY2026) while non-AI-related revs will grow at 13% CAGR to Rmb247.9bn with revs contribution declining from 85% of total in FY2026 to 30% in FY2031.

Within the AI-related revs, we forecast MaaS related revs to grow at 235% CAGR to reach Rmb438.6bn (US$62.6bn) by 2031 and accounted for 53% of total cloud revenues. We forecast AI-IaaS revs to grow at 46% CAGR to Rmb128.7bn and accounted for 15% of total cloud revs while AI-PaaS revs to grow at 40% CAGR to Rmb18.2bn and represent 2% of total cloud revs.

Figure 5. Growth drivers of Alibaba Cloud Intelligence Group revenues

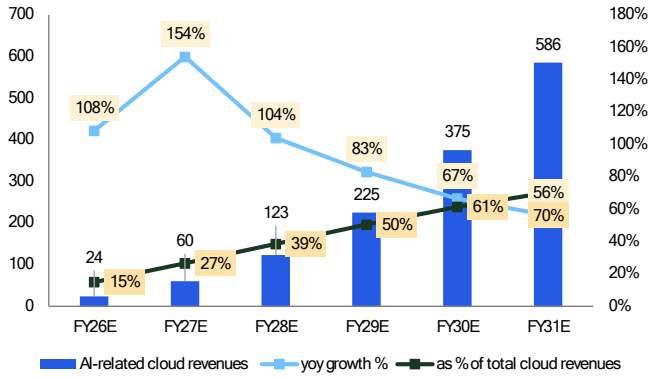
Rmb bn	FY26E	FY27E	FY28E	FY29E	FY30E	FY31E	FY26-31E CAGR
Cloud Intelligence Group revenues	**159**	**227**	**320**	**446**	**611**	**833**	***39%***
yoy growth %	*35%*	*43%*	*41%*	*40%*	*37%*	*36%*	
1. AI-related cloud revenues	**24**	**60**	**123**	**225**	**375**	**586**	***90%***
yoy growth %	*108%*	*154%*	*104%*	*83%*	*67%*	*56%*	
as % of total cloud revenues	*15%*	*27%*	*39%*	*50%*	*61%*	*70%*	
1.1 MaaS revenues	**1**	**10**	**54**	**133**	**260**	**439**	***235%***
yoy growth %		*859%*	*438%*	*146%*	*96%*	*69%*	
as % of total cloud revenues	*1%*	*4%*	*17%*	*30%*	*43%*	*53%*	
1.2 IaaS revenues	**19**	**43**	**59**	**80**	**100**	**129**	***46%***
yoy growth %		*124%*	*39%*	*35%*	*25%*	*28%*	
as % of total cloud revenues	*12%*	*19%*	*19%*	*18%*	*16%*	*15%*	
1.3 PaaS revenues	**3**	**8**	**10**	**12**	**15**	**18**	***40%***
yoy growth %		*122%*	*29%*	*28%*	*20%*	*23%*	
as % of total cloud revenues	*2%*	*3%*	*3%*	*3%*	*2%*	*2%*	
2. Non-AI-related cloud revenues	**135**	**167**	**196**	**221**	**236**	**248**	***13%***
yoy growth %	*27%*	*23%*	*18%*	*12%*	*7%*	*5%*	
as % of total cloud revenues	*85%*	*73%*	*61%*	*50%*	*39%*	*30%*	

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports and Citi Research Estimates

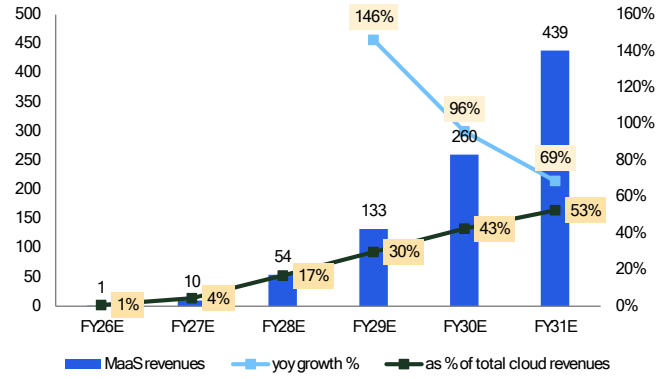
www.futoucaixin.cn

Figure 6. AliCloud AI-related cloud revenues (Rmb bn)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports and Citi Research Estimates

Figure 7. AliCloud MaaS revenues (Rmb bn)

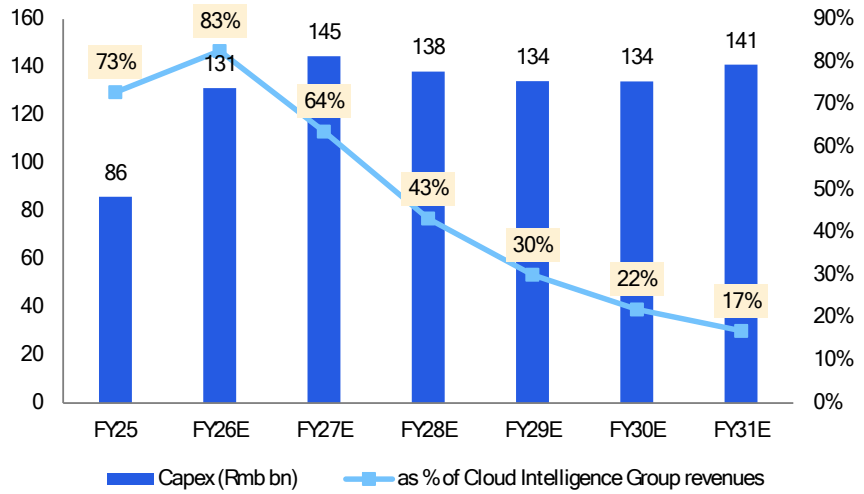


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports and Citi Research Estimates

Cumulative Capex spend

We had previously forecasted BABA capex spend to remain stable at around Rmb130-140bn per year and we now introduce our FY2030-31E capex forecast to sustain the similar momentum. Cumulatively, our estimated capex spend amounts to Rmb822.9bn for six years from FY2026 to FY2031. And we forecast capex as % of total cloud revs to decline from 83% in FY2026 to 17% in FY2031. Our forecast suggest that capex will account for average of 32% of cloud revs for 6 years period.

Figure 8. Capital expenditure of Alibaba



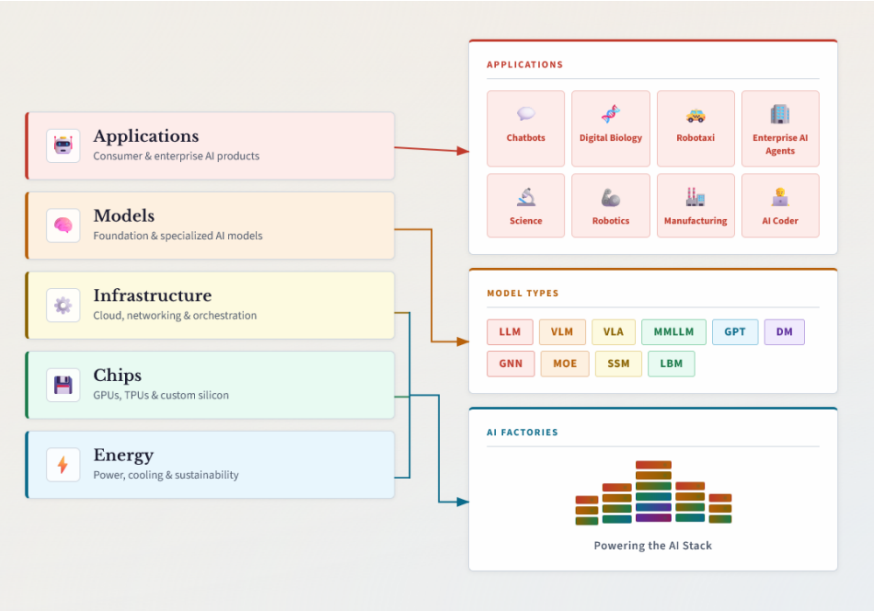
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports and Citi Research Estimates

www.futoucaixin.cn

Decoding AliCloud's AI Full-Stack

As we mentioned in our recent MiniMax's initiation report (link), Nvidia CEO Jensen Huang recently pointed out (10 March 2026; see company link) that AI, when viewed as infrastructure, resolves into a five-layer stack, like "five-layer cake." At the foundation of the AI stack is **Energy**. For every token produced, there is use of power, and energy is the first principle of AI infrastructure. The **Chips** layer sits above energy, which are the processors designed to transform energy into computation. **Infrastructure** is above chips, which consists of land, power delivery, cooling, construction, networking and the systems that orchestrate tens of thousands of processors into one machine. Industry experts now call these systems **AI factories**, which are designed to **manufacture intelligence**. The 4th layer is M**odel**. Different AI models understand different kinds of information differently. And the 5th layer is **Applications,** which sits on the top and is usually seen as where **economic value** is created. Other than energy, we believe **Alibaba Cloud has exposure in four out of the five layers of AI stack illustrated by Nvidia with evolving development in the application layer.**

Figure 9. AI's five-layer stack



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Nvidia, Citi Research

Silicon (Chips) Layer - T-Head (Pingtouge)

Background of T-Head Unit

Established on Sept 19, 2018, Pingtouge or T-Head is a wholly-owned semiconductor chip business entity of Alibaba Group. Before the establishment of Pingtouge, as announced by Xinhuanet on April 20, 2018, Alibaba had acquired an integrated circuit design house, Hangzhou C-Sky Microsystems Co., Ltd (founded in 2001). According to the article, C-Sky was the only embedded CPU volume provider in China with its own instruction set architecture and CTO of Alibaba Zhang Jianfeng commented that the acquisition was an important step for Alibaba's chip development. Following the acquisition, T-Head was spun-out of DAMO Academy and formed the T-Head unit. T-Head has a full-stack product

www.futoucaixin.cn

series for cloud and edge integration, covering data center AI chips and processor IP licensing.

Damo Academy and its relationship with T-Head

DAMO (Discovery, Adventure, Momentum and Outlook) Academy was established in 2017 as a global research institute by Alibaba Group with the vision "**Tech to the Future**" and positioned itself as a "research institution committed to unraveling the mysteries of science and technology, driven by the vision of advancing humanity". Per Alibaba's statement at the establishment of the academy, the research labs would focus on both foundational and disruptive technology research including data intelligence, internet of things, fintech, quantum computing and human-machine interaction. In addition, the labs will also focus on topics such as machine learning, network security, visual computing, Natural Language Processing (NLP), etc. According to DAMO website, it is noted that the academy has successfully incubated a number of well-known technical products including Tongyi, T-Head, Xiaomanly and others.

In 2023, DAMO Academy went through significant restructuring including 1) the announcement in May 2023 (Per SCMP May 15, 2023) of transferring the autonomous driving lab to Cainiao for application development (which later as reported by Reuters dated Jan 29, 2026, Cainiao was reportedly planning to merge its autonomous driving operation as part of its investment into Chinese startup Zelos Technology) and 2) shutdown of its quantum computing lab announced in Nov 2023 (per Reuters Nov 27, 2023) and donate the lab to Zhejiang University.

Post restructuring, as reported by SCMP dated Dec 20, 2023, DAMO Academy has established a new research unit that support >100 postdoctoral candidates to conduct researches in AI and semiconductor. The first batch of recruits was reportedly to do work on AI LLMs, multimodal data generation, advanced video-compression technologies.

Chip architectures by T-Head

T-Head focuses in three main architectures: **1) ARM – cloud-based; 2) RISC-V (IoT/Edge) and 3) Proprietary AI (for inference).** Since the establishment, T-Head has successfully launched a series of products which include: Hanguang 800 AI inference chip, Zhenwu 810E training and inference integrated Ai accelerator chip, Yitian 710 Arm server CPU and Zhenyue 510 SSD controller chip. For physical world, it also has Yuzhen ultra-high frequency (UHF) RFID electronic tag chip that has been applied in logistics, warehousing, asset management and IoT industry.

RISC-V series

- **XuanTie Series** – On July 25, 2019, as reported by Bloomberg, Alibaba announced the release of **XuanTie 910**, it was considered as the first fully formed product by T-Head, designed to serve more heavy-duty IoT applications including autonomous driving cars, networking and server computing.
  - On March 24, 2026, various news source including CNBC reported that Alibaba has released **XuanTie C950**, a CPU based on RISC-V architecture that compete with Arm's blueprints, targeting to serve agentic AI purposes. This is a 5nm, 3.2GHz processor features 1000-instruction Out of order window and is designed for Agentic AI. Ali noted that the new CPU is able to handle the processing of multi-step tasks carried by AI agents. XuanTie C950 will be installed in data centers as it is designed for running AI model inferencing. BABA noted that XuanTie CPUS can be customized for specific

www.futoucaifxin.cn

inference patterns to better support different usages of chips for different customers and able to achieves 30% improvement in performance compared to other mainstream products.

SSD IC

- **Zhenyue 510 –** On Nov 2, 2023, reported by SCMP, Ali released a new RISC-V-based controller chip for server storage, Zhenyue 510 during its annual Apsara cloud computing conference. This controller integrated circuit (IC) for enterprise solid-state drives (SSDs) will be deployed in Alibaba Cloud's data center to support application including AI training, online transaction and big data analysis. Based on open-standard RISC-V architecture, Zhenyue 510 provides chip developers the ability to configure and customize their designs. There are two algorithmic innovation for Zhenyue 510: 1) LDPC (Low-Density Parity-Check) and 2) optimal voltage search that achieve high reliability at lower cost through optimizing operating voltage of the NAND Flash.

Ultra-high frequency RFID chips

- **Yuzhen 611 & 612 –** On Nov 15, 2022, at the 2022 International Internet of Things Exhibition (IOTE), T-Head released the Yuzhen 611 and Yuzhen 612, which are ultra-high frequency (UHF) RFID electronic tag chips designed for Internet of Everything scenarios.
  - Yuzhen 611 is a single-port chip with an industry-leading sensitivity of -24dBm suitable for a variety of scenarios including supermarket retail, smart logistics, supply chain, airline parcel tracking, and asset management, greatly improving production and inventory efficiency.
  - Yuzhen 612 features a dual-port omnidirectional antenna design, providing greater versatility to meet the demands of complex scenarios such as stacked or obscured goods.

Server processor

- **Yitian series** – Announced on October 19 2021, Yitian is built on 5nm process technology, powered by 128 Arm cores with 3.2GHz to delivery "exceptional performance and excellent energy efficiency". Each chip has 60bn integrated transistors and is the first server processor that is compatible with Armv9 architecture and includes 8 DDR5 channels and 96-lane PCIe 5.0 that provide high memory and I/O bandwidth.
  - Reported by Yahoo Tech dated 30 April, 2024, according to a recent study published in the Transactions on Cloud Computing journal by IEEE, Yitian 710 was identified as the fastest ARM-based CPU for cloud servers. It is deployed at scale across Ali Cloud for video encoding, HPC and gaming workloads and also deployed in a leading Chinese telco's AI computing centers.

AI Chips

- **Hanguang 800 (AI Inference NPU):** On 25 Sept 2019, Alibaba unveiled its first AI inference chip, a neural processing unit (NPU) named Hanguang 800 that specializes in acceleration of machine learning tasks. It was initially used internally within Alibaba business operation in product search and automatic translation on ecommerce sites, personalized recommendations, advertising and intelligent customer service. The Hanguang 800's remarkable test performance is propelled by its integrated, self-developed hardware framework and highly optimized algorithm designs. These components are specifically

Citi GPS: Global Perspectives & Solutions

tailored for business applications, including the retail and logistics sectors of the Alibaba ecosystem. It powers Taobao's recommendation engine and image recognition at hyperscale.

- **Zhenwu 810E** – As reported by various news article including SCMP on Jan 30, 2026, Alibaba unveiled its new AI chip, Zhenwu 810E, which is a ASIC chip supports 96GB of HBM2e memory for training and inference. It is based on self-developed inter-chip interconnect technology, Inter-Chip-Network (ICN), which offers high performance, high bandwidth and low latency which is suitable for large-scale model training and inference application.
  - Each Zhenwu 810E chip is equipped with seven ICN inter-chip interconnect ports. Paired with T-Head's proprietary interconnect acceleration library, it enables multi-card collaboration to efficiently support the demands of large-scale model training and inference. The chip is supported by self-developed AI product software stack features a complete software ecosystem and toolchain. It supports developers and businesses to expand at application level and support optimization of hardware performance.
  - SCMP noted that the performance is comparable to Nvidia's H20. The chip has been deployed in Ali Cloud's computing clusters and is used by external clients. Per product website, the use case scenarios are for autonomous driving, AI multimodal training and inferencing. Morning Star reported on Jan 29, 2026 that key enterprise customers adopting Zhenwu 810E including State Grid of China, XPeng Motors and Chinese Academy of Sciences. The article also reported that the new chip has more than 400 clients and also has been used internally for training for Qwen models.

Network interface card

- **Panmai 920 –** On April 28, 2026, at the 2026 Digital China Summit, Alibaba's T-Head released its first smart network interface card (NIC), the Panmai 920, which is the first 400G smart NIC in China with a built-in PCIe Switch and support a maximum throughput bandwidth of 400 Gbps. As reported by Sina finance article (28 April 2026), it can be applied in scenarios such as AI computing clusters with tens of thousands of cards, general-purpose computing clusters and high-performance storage. Accordingly, the Panmai 920 has already entered mass production and will be deployed in Alibaba Cloud's data centers.
  - The Panmai 920 is a smart NIC created for large-scale AI computing scenarios. It is equipped with a self-designed smart NIC chip and uses PCIe 5.0 and 112G PAM4 Ethernet technology.
  - It also supports multi-path RDMA, which can effectively reduce the completion time for training and inference tasks. At the same time, it integrates fine-grained network awareness tools and user-programmable congestion control algorithms to effectively handle and prevent congestion, ensuring that critical business operations are not affected. The Panmai 920 also introduces a chip-level network architecture with a built-in PCIe Switch, allowing the card to connect directly to GPUs and SSDs with extremely low latency.

A China Daily article dated Jan 30, 2026 reported that with the launch of Zhenwu 810E by T-Head, the "new chip" completes Alibaba's AI "golden triangle" combining cloud services at AliCloud, AI model development by Tongyi Lab and chips from T-Head.

Citi Global Markets

Figure 10. T-Head product summary

Product name	Date of release	Technical features	Use cases	Comparable products
AI chips				
Zhenwu 810E (真武810E)	29-Jan-26	- Proprietary Inter-Chip-Network (ICN) at 700GB/s - 96GB HBM2e memory - PCIe5.0 x 16 host bus - Compatibility to mainstream AI ecosystem and AliCloud products	- Autonomous driving - AI training - AI reasoning - Multimodal models	Nvidia H20, A800; Huawei Ascend 910B/910C, Kunlunxin P800, Biren BR100/BR104
Hanguang 800 (含光800)	25-Sep-19	- 820 TOPS peak performance - 78563 IPS inference performance - 500 IPS/W - Support mainstream deep learning frameworking including TensorFlow, Caffe, MXNet, ONNX	- Ecommerce marketing - Ecommerce search - AI reasoning	Nvidia T4 and L4; AWS Inferentia; Meta MTIA, Kunlunxin Gen1/2, Ascend 310/310P, Cambricon MLU series
RISC-V				
Xuantie c950 (玄铁c950)	24-Mar-26	- 22/GHz single-core performance - 3.2GHz clock speed - RVA23.1 mandatory extensions	- Data centers - autonomous systems - AI accelerators	- Intel Xeon6 - Arm V2 - AMD Zen5
Xuantie 910 (玄铁910)	25-Jul-19	- 16-core infrastructure - 7.1 coremark/MHz single-core performance - 2.5GHz clock speed - 8MB L2 cache	- AI - 5G - Autonomous driving	ARM Cortex-A72/A75; SiFive Performance P550; SiFive Intelligence X280; Ventana Veyron V1
Server Processor				
Yitian 710 (倚天710)	19-Oct-21	- 5nm process technology - 128 Arm v9 12-core CPU - 2.75 GHz clock speed - 8-channel DDR5 DRAM - 96-lane PCIe 5.0 I/O - 250W maximum power	- Video coding - Big data analysis - Ecommerce - AI reasoning	Intel Xeon 6, AMD EPYC 9005 series, Nvidia Grace, AmpereOne, AWS Graviton3/ Graviton 4, Azure Cobalt 100
Network interface card				
Panmai 920 (磐脉920)	28-Apr-26	- 400Gbps bandwidth - 400Mpps packet forwarding rate - 32-channel PCIe5.0 - 112G PAM4 network interface	- Smart compute cluster - High-performance storage - Generic compute cluster - Database and big data analysis	Nvidia BlueField-3 DPU; Nvidia ConnectX-7; Broadcom Thor 2 (BCM57608); Huawei HiSilicon
SSD IC				
Zhenyue 510 (镇岳510)	1-Nov-23	- 3400K IOPS - 4 microsecond latency - PCIe5.0 / DDR5.0 high-speed interface - 420K IOPS/W efficiency - Support 32TB+ storage	- AI training - AI reasoning - Distributed storage - Online transactions	Samsung PM1743 Controller; Marvell Bravera SC5; InnoGrit Tacoma (IG5666)
Ultra-high frequency RFID chips				
Yuzhen 611/612 (羽阵611/612)	15-Nov-22	- EPCglobal G2 V2 ISO/IEC18000-6C protocol - -24dBm read sensitivity - HBM±10KV electrostatic discharge protection - -40 ~ +85°C operating temperature range - 96-bit TID, 128-bit EPC storage	- Shoes, apparel retail - FMCG retail - Smart logistics - Suppy chain management	NXP UCODE 8; Impinj Monza R6

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

Alibaba Cloud Infrastructure

Domestic and international data centers

Alibaba Cloud operates data centers in 94 availability zones in 29 regions around the world, of which 13 regions are located in domestic China with the data center in Hangzhou opened in 2011 and the latest one in Wuhan opened in 2023. 15 regions are located in overseas countries including Singapore, Thailand, Malaysia, Indonesia, Philippines, Japan, South Korea, US, Mexico, Germany, UK, UAE, KSA and Hong Kong. As announced on Sept 25, 2025, Alibaba Cloud plans to launch more data centers in eight locations in the coming year, this includes first one in Brazil, France and the Netherlands while additional data centers to be added in Mexico, Japan, South Korea, Malaysia and Dubai. Alibaba CDN currently supports 2300+ nodes in China and 900+ international nodes covering 70+ countries.

www.tatouai.xyz

Figure 11. Regions and availability zones of AliCloud data center

Region Name	Country / Territory	Primary Cities	Year Est.	Availability Zones	Status	Notes
China Mainland						
China (Hangzhou) cn-hangzhou	China	Hangzhou	2011	3	Active	First Alibaba Cloud region; HQ city
China (Qingdao) cn-qingdao	China	Qingdao	2012	2	Active	Second region launched
China (Beijing) cn-beijing	China	Beijing	2013	3	Active	Capital region; major enterprise hub
China (Zhangjiakou) cn-zhangjiakou	China	Zhangjiakou	2014	2	Active	Green energy focus; cold climate cooling
China (Shenzhen) cn-shenzhen	China	Shenzhen	2014	3	Active	Tech hub; proximity to Hong Kong
China (Shanghai) cn-shanghai	China	Shanghai	2015	3	Active	Financial hub; largest enterprise region
China (Hohhot) cn-huhehaote	China	Hohhot	2017	2	Active	Inner Mongolia; renewable energy
China (Chengdu) cn-chengdu	China	Chengdu	2020	2	Active	Southwest China hub
China (Guangzhou) cn-guangzhou	China	Guangzhou	2020	3	Active	Pearl River Delta; manufacturing belt
China (Heyuan) cn-heyuan	China	Heyuan	2020	2	Active	Guangdong province backup region
China (Ulanqab) cn-wulanchabu	China	Ulanqab	2020	2	Active	Inner Mongolia; AI workload focus
Greater China & Asia Pacific						
China (Hong Kong) cn-hongkong	Hong Kong SAR	Hong Kong	2014	3	Active	International gateway; financial services
Singapore ap-southeast-1	Singapore	Singapore	2015	3	Active	International HQ; APAC flagship; AIGC hub
Japan (Tokyo) ap-northeast-1	Japan	Tokyo	2016	4	Active	4th DC launched Mar 2026; AI demand surge
Malaysia (Kuala Lumpur) ap-southeast-3	Malaysia	Kuala Lumpur	2017	3	Active	3rd DC launched Jul 2025; 4th planned 2026
Indonesia (Jakarta) ap-southeast-5	Indonesia	Jakarta	2018	3	Active	First intl. public cloud provider in Indonesia
South Korea (Seoul) ap-northeast-2	South Korea	Seoul	2022	2	Active	Additional DC planned; semiconductor ecosystem
Philippines (Manila) ap-southeast-6	Philippines	Manila	2021	2	Active	2nd DC opened Oct 2025
Thailand (Bangkok) ap-southeast-7	Thailand	Bangkok	2022	2	Active	2nd DC opened Feb 2025; GenAI focus
Europe						
Germany (Frankfurt) eu-central-1	Germany	Frankfurt	2016	3	Active	Operated with Vodafone; C5 certified; 100% green
UK (London) eu-west-1	United Kingdom	London	2018	2	Active	2 AZs launched Oct 2018
France (Paris) eu-west-2	France	Paris	2026*	–	Planned	Announced Sep 2025; launch expected 2026
Netherlands eu-west-3	Netherlands	Amsterdam	2026*	–	Planned	Announced Sep 2025; launch expected 2026
Americas						
US (Silicon Valley) us-west-1	United States	San Jose Silicon Valley	2015	2	Active	First US region; tech ecosystem proximity
US (Virginia) us-east-1	United States	Ashburn Virginia	2015	2	Active	East Coast US; enterprise & government
Mexico (Querétaro) mx-central-1	Mexico	Querétaro	2025	1	Active	First LatAm region; launched 2025
Brazil (São Paulo) sa-east-1	Brazil	São Paulo	2026*	–	Planned	Announced Sep 2025; launch expected 2026
Middle East & Africa						
UAE (Dubai) me-east-1	United Arab Emirates	Dubai	2016	2	Active	2nd DC launched Oct 2025; MENA hub
Saudi Arabia (Riyadh) me-central-1	Saudi Arabia	Riyadh	2023	2	Partner	Partner region (SAU); Vision 2030 alignment

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

www.futoucaixin.cn

As reported by various news media including CNBC on April 8th 2026, Alibaba and a leading Chinese telco launched a data center in China designed for AI training and inferencing. The data center will feature 10,000 of Alibaba's self-developed Zhenwu AI chips while the telco will own and operate the data center. Both Alibaba and the telco stated that the Shaoguan facility site could expand to 100,000 chips in the future.

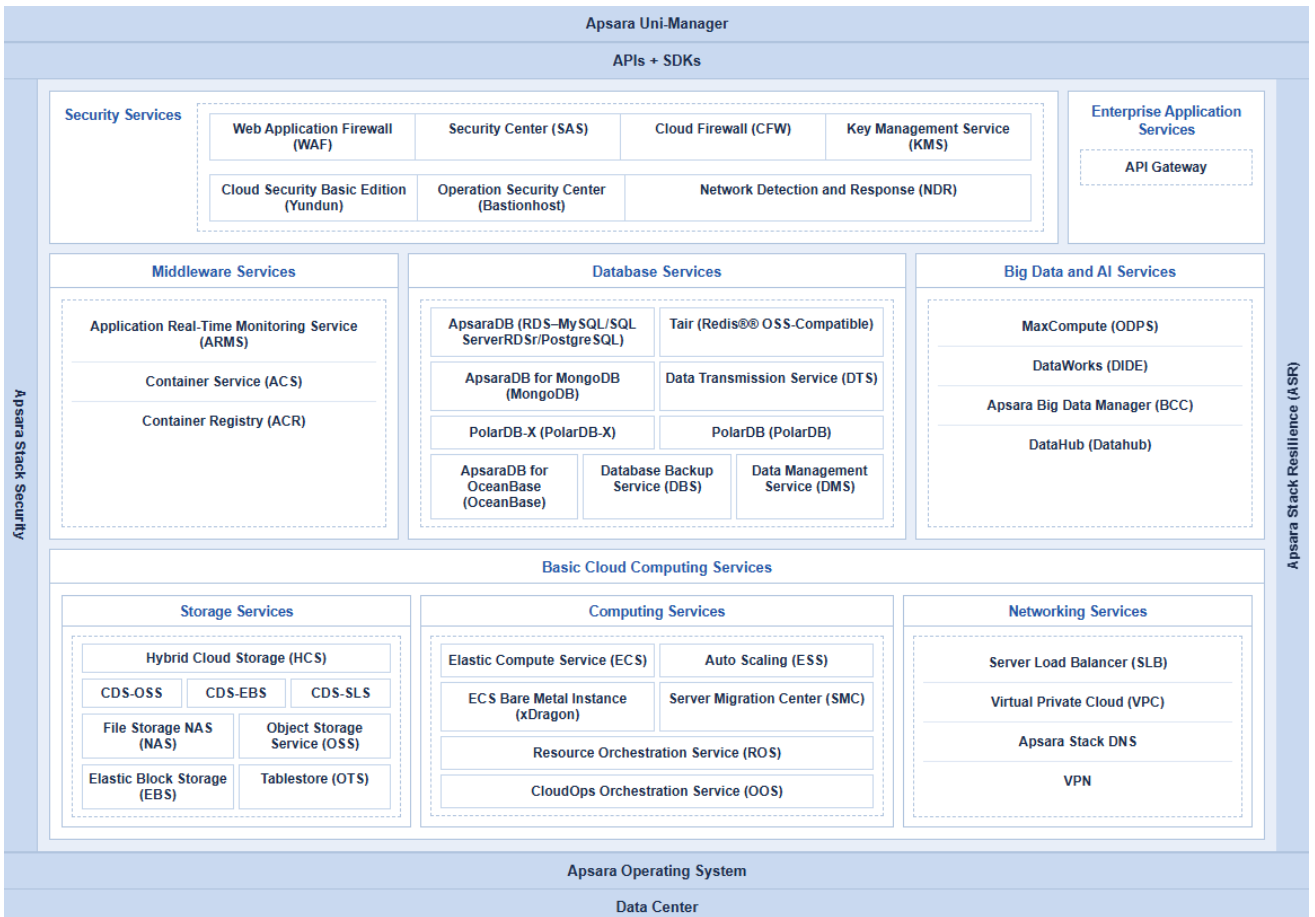
China context vs the world in data center

Data Center GAINs (Gen AI Names) Multi-Asset - Strong Absorption Supports Solid 2026 and LT Data Center Demand

According to a comment by Cargoson on 4 Jan 2026 (a transport management software company) quoted a report dated Nov 2025, China ranks # 4 globally with 449 data centers behind the US at 5,427 data centers, Germany 529 and UK 523. Within that, China has 190 hyperscale data centers or 16% of global total (rank # 2 behind the US at 642 or 54% of global total). Global data center capacity reached 122.2 gigawatts (GW) as of 1Q25, per Synergy Research Group, of which China accounted for 16% of global total or 19.6GW. Hyperscale operators accounted for 44% of global data center capacity and is expected to grow to 61% by 2030.

Apsara full-stack cloud operating system

Figure 12. Apsara full stack capabilities



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

www.futoucaixin.cn

Apsara Stack is an enterprise-grade cloud platform designed specifically for public service sectors and enterprise customers based on the Apsara distributed operating system to ensure the isolation of resources and cloud management. With its capacity for local deployment, Apsara Stack provides dedicated **computing, storage, and network resources** to meet the needs of public service sectors and enterprise customers for asset self-ownership, security compliance, and autonomous O&M. Thanks to its elastic and agile nature, Apsara Stack enables customers to maximize resource utilization and expedite service deployment with ease. Apsara offers more than 40 cloud products covering IaaS, PaaS, big data and security spanning across key solutions such as Disaster Recovery and Backup, Multi-region deployment, cloud-native applications, integrated O&M and Operational Industry cloud. On figure 8, we detail the product/services offered by Alibaba Cloud and the function feature.

www.futoucaixin.cn

Figure 13. Alibaba Cloud product/services by category and description

PRODUCT / SERVICE	DESCRIPTION
COMPUTE	
Elastic Compute Service (ECS)	Scalable virtual cloud servers (IaaS) for various workloads
Elastic Container Instance (ECI)	Serverless container runtime without managing servers
Auto Scaling	Automatically adjusts compute capacity based on demand
Batch Compute	Large-scale parallel and batch computing jobs
Function Compute	Serverless, event-driven function execution (FaaS)
Simple Application Server	Lightweight cloud server for small apps and websites
STORAGE	
Object Storage Service (OSS)	Scalable, secure cloud object/blob storage
Apsara File Storage NAS	Managed network-attached storage (NFS/SMB)
Cloud Disk (Block Storage)	High-performance persistent block storage for ECS
Hybrid Backup Recovery (HBR)	Cloud-based backup and disaster recovery
Tablestore	NoSQL wide-column data storage for massive datasets
NETWORKING	
Virtual Private Cloud (VPC)	Isolated private network environment in the cloud
Server Load Balancer (SLB)	Distributes traffic across multiple backend servers
Elastic IP Address (EIP)	Static public IP addresses for cloud resources
CDN (Content Delivery Network)	Accelerates content delivery globally
Express Connect	High-speed private connectivity between VPCs or on-premises
Global Accelerator	Improves global application access speed and reliability
NAT Gateway	Enables internet access for VPC instances
DATABASES	
ApsaraDB RDS	Managed relational DB (MySQL, PostgreSQL, SQL Server, MariaDB)
PolarDB	Cloud-native relational database with high performance
ApsaraDB for Redis	Managed in-memory key-value store
ApsaraDB for MongoDB	Managed NoSQL document database
AnalyticDB	Cloud data warehouse for real-time analytics
Data Lake Analytics (DLA)	Serverless SQL queries over data lakes
Lindorm	Multi-model database for IoT and time-series data
AI & MACHINE LEARNING	
PAI (Platform for AI)	End-to-end ML platform for model training and deployment
Machine Translation	Neural machine translation across multiple languages
Image Search	AI-powered visual product search
Intelligent Speech Interaction	ASR, TTS, and NLP voice services
Face Recognition	Facial detection, comparison, and attribute analysis
Alibaba Cloud Model Studio	Build and deploy large language model (LLM) applications
Qwen (通义千问)	Alibaba's proprietary large language model (LLM)

PRODUCT / SERVICE	DESCRIPTION
SECURITY	
Anti-DDoS	Protection against distributed denial-of-service attacks
Web Application Firewall (WAF)	Protects web apps from common exploits and bots
Security Center	Unified threat detection and security management
SSL Certificates Service	Managed SSL/TLS certificate issuance and deployment
Cloud Firewall	Network-level firewall for VPC traffic control
IDaaS (Identity as a Service)	Centralized identity and access management
BIG DATA & ANALYTICS	
MaxCompute	Large-scale data warehousing and batch analytics
DataWorks	Data integration, development, and governance platform
Realtime Compute (Flink)	Stream processing for real-time data pipelines
Quick BI	Business intelligence and data visualization tool
E-MapReduce	Managed Hadoop/Spark big data processing clusters
Data Integration	ETL service for moving data across systems
CONTAINERS & DEVOPS	
Container Service for Kubernetes (ACK)	Managed Kubernetes cluster service
Container Registry (ACR)	Managed Docker image repository
Alibaba Cloud DevOps (Yunxiao)	CI/CD, code hosting, and project management
Resource Orchestration Service (ROS)	Infrastructure-as-code (IaC) for cloud resources
Serverless Kubernetes (ASK)	Kubernetes without managing nodes
MIDDLEWARE & APP SERVICES	
Message Queue (MQ / RocketMQ)	Distributed message queuing for decoupled architectures
API Gateway	Manage, publish, and secure APIs
Microservices Engine (MSE)	Service mesh and microservices governance
EventBridge	Serverless event bus for application integration
Log Service (SLS)	Centralized log collection, query, and analysis
IOT & EDGE COMPUTING	
IoT Platform	Connect, manage, and communicate with IoT devices
Link IoT Edge	Edge computing for IoT data processing at the source
IoT Security Center	Security management for IoT device fleets
MEDIA & CDN	
ApsaraVideo VOD	Video-on-demand hosting, transcoding, and delivery
ApsaraVideo Live	Live streaming platform with low latency
Media Processing	Automated video/audio transcoding and processing
Alibaba Cloud CDN	Global content delivery acceleration
HYBRID & MULTI-CLOUD	
Apsara Stack	Private cloud solution for on-premises deployment
Cloud Enterprise Network (CEN)	Global private network connecting VPCs and on-premises
Smart Access Gateway (SAG)	SD-WAN solution for branch office connectivity

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

www.futoucaixin.cn

Figure 14. AliCloud's product availability by IDC zone

Region	Compute					Storage				Database					Network					AI / ML		
	ECS	GPU	ACK	FC	ECI	OSS	EBS	NAS	Backup	RDS	PolarDB	MongoDB	Redis	Analytic DB	VPC	SLB	CDN	CEN	GA	PAI	Model Studio	ENS
China Mainland																						
China (Hangzhou) cn-hangzhou	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
China (Qingdao) cn-qingdao	✓	⚡	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
China (Beijing) cn-beijing	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
China (Zhangjiakou) cn-zhangjiakou	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	✓	⚡	✓
China (Shenzhen) cn-shenzhen	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
China (Shanghai) cn-shanghai	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
China (Hohhot) cn-huhehaote	✓	⚡	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	✗	✓
China (Chengdu) cn-chengdu	✓	⚡	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
China (Guangzhou) cn-guangzhou	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
China (Heyuan) cn-heyuan	✓	⚡	✓	⚡	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	✗	✓
China (Ulanqab) cn-wulanchabu	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	✓	⚡	✓
Greater China & Asia Pacific																						
China (Hong Kong) cn-hongkong	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Singapore ap-southeast-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Japan (Tokyo) ap-northeast-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓
Malaysia (KL) ap-southeast-3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
Indonesia (Jakarta) ap-southeast-5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
South Korea (Seoul) ap-northeast-2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
Philippines (Manila) ap-southeast-6	✓	⚡	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	✗	✓
Thailand (Bangkok) ap-southeast-7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
Europe																						
Germany (Frankfurt) eu-central-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	⚡	✓
UK (London) eu-west-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
Americas																						
US (Silicon Valley) us-west-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	⚡	✓
US (Virginia) us-east-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	⚡	✓
Mexico (Querétaro) mx-central-1	✓	⚡	✓	⚡	✓	✓	✓	⚡	✓	✓	⚡	✓	✓	✗	✓	✓	✓	✓	✓	✗	✗	✓
Middle East																						
UAE (Dubai) me-east-1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	⚡	✓	✓	✓	✓	✓	⚡	⚡	✓
Saudi Arabia (Riyadh) me-central-1	✓	⚡	✓	⚡	✓	✓	✓	⚡	✓	✓	⚡	✓	✓	✗	✓	✓	✓	✓	✓	✗	✗	✓

COMPUTE PRODUCTS
- **ECS** — Elastic Compute Service
- **GPU** — Elastic GPU Service
- **ACK** — Container Service for Kubernetes
- **FC** — Function Compute (Serverless)
- **ECI** — Elastic Container Instance

STORAGE PRODUCTS
- **OSS** — Object Storage Service
- **EBS** — Elastic Block Storage
- **NAS** — File Storage NAS
- **Backup** — Cloud Backup (BaaS)

DATABASE PRODUCTS
- **RDS** — ApsaraDB RDS (MySQL/PG/SQL Server)
- **PolarDB** — Cloud-native relational DB
- **MongoDB** — ApsaraDB for MongoDB
- **Redis** — Tair (Redis-compatible)
- **AnalyticDB** — Real-time data warehouse

NETWORK PRODUCTS
- **VPC** — Virtual Private Cloud
- **SLB** — Server Load Balancer
- **CDN** — Content Delivery Network
- **CEN** — Cloud Enterprise Network
- **GA** — Global Accelerator

AI / ML PRODUCTS
- **PAI** — Platform for AI (ML training/inference)
- **Model Studio** — GenAI model deployment
- **ENS** — Edge Node Service (3,200+ nodes)

AVAILABILITY LEGEND
- ✓ Generally Available
- ⚡ Partial / Limited availability
- ✗ Not available / Not confirmed

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

Elastic computing services

SHENLONG architecture (virtualization technology)

SHENLONG architecture (or X-Dragon) is a proprietary bare metal server architecture developed by Alibaba Cloud that integrates software and hardware with dedicated chips used to reduce virtualization overheads. It offers direct access

www.futouqixin.cn

to CPU and RAM resources which built around a custom X-Dragon MOC card). It is the virtualization technology behind Alibaba Cloud Elastic Compute Service (ECS). The first two generations were called Xen and KVM and X-Dragon before renamed as SHENLONG in its 3rd generation. During 2021 Apsara conference, Alibaba released its 4th generation of SHENLONG architecture with significant improvement in storage IOPS, network PPS and network latency. It supports a variety of elastic computing product forms, such as elastic bare metal, virtual machines, and elastic container instances, serving multiple industries, such as healthcare, new government affairs, intelligent manufacturing, the Internet, and education.

Storage

Pangu (distributed storage)

According to Alibaba Cloud, Pangu is a high-reliability, high-availability and high-performance distributed file system developed by Alibaba Cloud. As a storage core of Alibaba Cloud, Pangu 1.0 supported the rapid development of multiple business lines of Ali Cloud, including ECS, NAS, OSS, Table Store, Max Compute and AnalyticDB. Following the success of Pangu 1.0, Alibaba Cloud introduced Pangu 2.0 in 2018, a redesigned storage engine built to leverage modern hardware advancements and meet evolving business demands. It serves as a unified storage core for Alibaba Cloud, Alibaba Group, and Ant Financial, supporting a wide range of services. One of the key upgrades for Pangu 2.0 is the separation of Storage and Computing so that storage cluster can delivery higher performance to serve multiple computing clusters efficiently.

Networking

Luoshen (networks on the cloud)

Apsara Luoshen represents Alibaba Cloud networks on the cloud that manages infrastructure network system. Apsara Luoshen has undergone three generations. The first generation is the classic network, which was mainly used to solve the problem of connectivity. The second generation launched the VPC network for security isolation. And the third generation Apsara Luoshen, which provides solutions to connect to the cloud, and offers the same capabilities as conventional enterprise networks. Alibaba Cloud defines the final stage as Networkless. In this stage, the networks are invisible to end users.

www.futoucaixin.cn

Figure 15. Selected Customer Case Studies

Customer	Region	Industry	Alibaba Cloud Service Used	Use Case	After / Key Improvements
Airbnb	Global (HQ: USA)	Travel & Hospitality	Qwen LLM	AI-powered customer service agent	Consolidated core customer service on Qwen; CEO cited superior **quality, speed, and cost-effectiveness** vs. alternatives
Dingdong	China	E-commerce & Retail	Qwen LLM via Model Studio	"DingXiaoDong" AI assistant for pre-sales, product recommendations, logistics & customer support	**70% reduction in model inference costs**; automated pre-sales, recommendations, and logistics support
RayNeo (TCL Electronics)	China / Global	Consumer Electronics & AR Wearables	Qwen LLM (Multimodal)	AI-powered voice assistant embedded in AR smart glasses	Context-aware multi-turn conversations, real-time AI translation, immersive navigation; **avg. AI query response time of 1.3 seconds**
Hongqi (FAW Group)	China	Automotive	Qwen LLM + T-Head AI Chip + Alibaba Cloud Infrastructure	"Lingxi" intelligent cockpit — voice-driven multi-stop trip planning, integrated Alibaba ecosystem services (Taobao, Amap, Damai)	Millisecond-level responses; complex multi-intent voice commands (e.g., plan multi-stop itineraries); car transformed into a **"service delivery terminal"**; debuted on Hongqi HS6 PHEV
ZEEKR (EV Brand)	China	Automotive & Manufacturing	Quick BI (AI-powered Business Intelligence)	Unified supply chain, manufacturing, and DTC sales data into "Jishu BI" platform with real-time OTD dashboards	Query latency reduced to **seconds**; **60% executive adoption rate**; industry-leading **76-day vehicle delivery time**; contributed to a record **37-month IPO**
PRO-NET	Malaysia (Southeast Asia)	Automotive & EV / Mobility Tech	Qwen LLM, Model Studio, Platform for AI (PAI)	Internal AI chatbot for operations; future expansion to customer-facing apps and vehicle infotainment	Cloud-first, data-driven foundation established; AI chatbot in development; advancing Malaysia's EV ecosystem and aligning with **Malaysia AI Nation 2030** goals
GoTo Group	Indonesia (Southeast Asia)	Digital Ecosystem / Super App	Alibaba Cloud Infrastructure (Compute, Database, Networking, Security, Analytics)	Full cloud migration of GoTo's services to Alibaba Cloud; digital talent development	Cloud migration began Oct 2024; data remains **onshore in Indonesia**; streamlined operations; Alibaba Cloud committed to training **800,000 individuals by 2033**; GenAI Hackathon co-hosted in May 2025
Olympic Broadcasting Services (OBS) — Paris 2024	Europe (France)	Media, Sports & Broadcasting	OBS Cloud 3.0, AI Multi-Camera Replay, Qwen LLM (for future AMD system)	Cloud-first Olympic broadcasting; AI-powered 3D multi-camera replay across 14 venues, 21 sports	Cloud became **primary distribution method** for first time in Olympic history; **11,000+ hours of content** produced (15% more than Tokyo 2020); **70% of global distribution over cloud**; 54 broadcasters used OBS Live Cloud; AI replays ready in **seconds**
EKS InTec GmbH	Europe (Germany)	Manufacturing & Industrial	Alibaba Cloud Workspace, Edge Computing, AI	Scalable digital twin technology for smart factory training and production optimization	Digital twins accessible from **any device via cloud streaming**; reduced hardware investment; launched in Germany & China by end of 2024; expanding to Asian markets by mid-2025
Salesforce (via Alibaba Cloud Partnership)	Europe / Global	Enterprise SaaS / CRM	Alibaba Cloud Infrastructure (hosting Salesforce Sales Cloud, Service Cloud, Platform)	Enabling multinational (incl. European) companies to run Salesforce in mainland China with data residency compliance	Compliant, localized Salesforce experience on Alibaba Cloud; **global CX consistency with local relevance**; Deloitte recognized as 2024 Salesforce on Alibaba Cloud Partner
Cainiao Network (Alibaba Logistics)	China / Global	Logistics & Supply Chain	AI, Robotics, Machine Learning, Route Optimization	Fully automated "Future Warehouse"; AI-driven parcel sorting, route optimization, demand forecasting	Order processing time **↓ 50%**; warehouse productivity **↑ 70%**; logistics costs **↓ 30%** via AI route optimization; **25% reduction** in stockouts/overstock; handled **1 billion+ orders** during peak sales
AliMe (Alibaba E-commerce Platform)	China	E-commerce & Retail	AliMe AI Chatbot (NLP + ML)	AI customer service agent handling pre-sales, order tracking, returns, and product queries	**95% of customer inquiries** handled by AI; response times **↑ 80%** faster; customer service costs **↓ 50%**; customer retention **↑ 30%**; handled **9 million queries** on Singles' Day 2017
30+ Automakers & Smart Driving Providers	China / Global	Automotive & Autonomous Driving	T-Head "Zhenwu" PPU AI Chips, Alibaba Cloud Infrastructure	Smart driving R&D — data processing, model training, simulation	**100,000+ T-Head Zhenwu PPU cards** deployed; used by **30+ automakers**; all major Chinese automakers reportedly adopted Alibaba Cloud for domestic and international operations by 2026 Singapore Motor Show
Saudi Cloud Computing Company (SCCC)	Middle East (Saudi Arabia)	Government & Cloud Infrastructure	Alibaba Cloud Full Stack (JV Partnership)	Delivering advanced cloud computing services tailored to Saudi businesses; supporting Vision 2030	Joint venture established; specialized cloud solutions and ecosystem support for Saudi businesses; aligned with **Saudi Vision 2030** digital transformation goals
Nokia + Cainiao (Logistics)	China / Global	Logistics & Telecom	Industrial-grade Private Wireless + AI Robotics	Digital transformation of Alibaba's Cainiao Smart Logistics Network warehouses	Enhanced warehouse efficiency via advanced robotic systems and industrial-grade private wireless; reduced operational costs

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

Platform-As-A-Services (PaaS)

Alibaba's PaaS, supported by Apsara Cloud operating system, is designed for application development, AI training, and data management without the infrastructure maintenance. Essentially, this is a cloud-based environment that allow developers to build, deploy and scale applications without managing underlying infrastructure like servers, OS or storage.

Key product offerings include Platform for AI (PAI), Model Studio, Apsara DB for RDS database, container services, Blockchain as a Service (BaaS).

Platform For AI (PAI)

PAI is a machine learning or deep learning engineering platform intended for enterprises and developers. It provides easy-to-use, cost-effective, high-performance, and easy-to-scale plug-ins that can be applied to various industry scenarios. With over 140 built-in optimization algorithms, Platform for AI provides whole-process AI engineering capabilities including data labeling (PAI-iTAG), model building (PAI-Designer and PAI-DSW), model training (PAI-DLC), compilation optimization, and inference deployment (PAI-EAS).

www.futoucaixin.cn

ApsaraDB RDS

ApsaraDB RDS is a stable, reliable, cost-effective, and scalable online database service that supports most mainstream database engines, including MySQL, SQL Server, PostgreSQL, and MariaDB. It automates infrastructure management, patching, backups, failover, and monitoring.

Container Services for Kubernetes (ACK)

Alibaba Cloud Container Service for Kubernetes (ACK) integrates virtualization, storage, networking, and security capabilities. ACK allows developers to deploy applications in high-performance and scalable containers and provides full lifecycle management of enterprise-class containerized applications.

AgentRun, a one-stop Agentic AI infrastructure platform built on the globally leading Function Compute (FC) technology. It integrates the extreme elasticity, zero maintenance, and pay-per-use features of Serverless with AI-native application scenarios, helping enterprises achieve optimal cost and efficiency improvements, with an average Total Cost of Ownership (TCO) reduction of 60%.

Alibaba Token Hub (ATH)

On March 16, 2026, Alibaba announced it has established Alibaba Token Hub (ATH) a major strategic business group to consolidate its AI operations and to focus on "token-based" economics to capture how AI has transformed from intelligence into action as the agentic era takes hold. Headed by Alibaba Group CEO Eddie Wu, ATH is structured around a three-layer framework which aligns with its core mission of creating, distributing and applying tokens. ATH comprises of five main divisions focusing on three-layer framework of 1) **creating (Tongyi), 2) distributing (MaaS) and 3) applying (Qwen, Wukong, Innovation) of AI tokens.**

As Alibaba pointed out that at the application layer, it deploys tokens in real-world settings to power **Wukong**, its newly launched enterprise-grade AI-native agent platform. Wukong empowers businesses to deploy "digital partners"—AI agents capable of executing daily workflows. This represents a strategic evolution of AI from passive "thinking partners," like chatbots, to active "doing partners" that can perform tasks autonomously. On the consumer front, Alibaba's flagship Qwen App is becoming China's first all-in-one personal AI assistant for life, work, and learning, with 300 million monthly active users as of February.

AI restructuring in April

On April 8, Group CEO Eddie Wu issued an internal letter announcing a new round of major organizational adjustments centered on the goal of "accelerating AI development." Core measures include establishing a new Group Technology Committee and upgrading the former Tongyi Lab to the Tongyi Large Model Business Unit, aiming to further integrate resources and strengthen investment and synergy in large models and AI infrastructure.

According to the internal letter as reported by Sina Finance dated Apr 8 (link), the newly established Group Technology Committee will be chaired by Mr. Eddie Wu himself. Core members include Zhou Jingren, Wu Zeming, and Li Feifei. Specifically, 1) Zhou Jingren serves as the committee's Chief AI Architect and will lead the newly upgraded Tongyi Large Model Business Unit.2) Li Feifei is responsible for Alibaba Cloud technology and AI cloud infrastructure construction and will also serve as Alibaba Cloud's Chief Technology Officer (CTO). 3) Wu Zeming

www.fuloucaifxm.cn

is responsible for the group's business technology platform and AI inference platform development, focusing on his role as Alibaba Group's CTO.

Figure 16. Alibaba Token Hub (ATH) Business Group Structure and Positioning

BUSINESS UNIT	STRATEGIC ROLE (TOKEN FLOW)	CORE RESPONSIBILITIES
Tongyi Lab	*Create Token (Production)*	Responsible for foundational model research and development, especially leading multi-modal models, and continuously pushing the capability limits of large models.
MaaS Business Line	*Deliver Token (Distribution)*	Builds an efficient and open model service platform and technical framework to support the entire industry's AI ecosystem and its partners.
Qianwen Business Dept.	*Apply Token (Consumer Entry)*	Dedicated to creating a globally leading personal AI assistant, focusing on the growth of C-side user interaction and traffic.
Wukong Business Dept.	*Apply Token (Enterprise Entry)*	Builds a B-side AI-native work platform, deeply integrating model capabilities into enterprise workflows to establish a consumption moat.
AI Innovation Business Dept.	*Apply Token (Innovation)*	Explores various cutting-edge AI application scenarios to rapidly validate new business models and identify incremental markets.

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports, Citi Research

Tongyi Lab (Tongyi Large Model Business Unit)

Tongyi Lab is the AI research lab under Alibaba Group focusing on R&D of large model technology including research, training, iteration of foundational model systems. The team is dedicated to promoting the open sourcing of model capabilities, ecosystem building and industry applications. As part of the AI restructuring announced on April 8, 2026, Tongyi Lab has been upgraded to Tongyi Large Model Business Unit under the Alibaba Token Hub (ATH) and led by Zhou Jingren. Tongyi Qianwen (Qwen) series is a family of large language and multimodal models developed by Tongyi Lab.

Model-as-a-Service (MaaS)

Alibaba Cloud Model Studio (Bailian Platform)

Launched in Oct 2023, the Alibaba Cloud Bailian Platform or Model Studio is a one-stop model service platform, integrating the full series of Qwen model and other maintain stream third-party LLMs and providing official Qwen APIs and OpenAI-compatible APIs and supports multimodal capabilities such as text, image and audio/video. It is often time called as the aggregator of models – as it allows customers to use non-Alibaba models while staying within the Alibaba billing ecosystem. It supports scenarios such as code generation, translation, data mining and intention recognition. It allows developers to call models on demand without managing infrastructure. It offers over 200 industry-leading models including proprietary Qwen series and Wan models, and DeepSeek, Kimi, GLM, MiniMax, and others. Over 1mn enterprise and individual users currently using Model Studio. As a unified entry point for enterprise AI implementation, Model Studio simplifies complex model calls into standardized services. It provides RAG (Retrieval

www.futongaixin.cn

Augmentation), Agent orchestration tools and Agent Store which offers enterprises more than hundreds of the ready-to-use industry-specific intelligent agents. More than 800,000 agents have been created on Model Studio supporting scenarios like content creation, intelligent marketing and production optimization. Bailian is the domestic version and Model Studio is now available in Singapore, Germany and the US.

During 2025 Apsara conference in Sept 2025, Alibaba Cloud's Bailian platform was upgraded to support the development and implementation of Agents on the basis of model into real world business operations. Bailian Platform's newly launched ModelStudio-ADK high a code framework that address the pain points of enterprises by enabling enterprises to develop Agents with autonomous planning and cyclic execution capabilities. Bailian has built seven enterprise-level capabilities, including the MCP Server for tool connection, the RAG Sever for multi-modal data fusion, the system-level sandbox, the intelligent memory server, the full-link observability system, dynamic reasoning scheduling and the Pay Server. MaaS Business unit will be led by Zhou Jinren.

Qwen Business Unit

According to Yahoo Finance article citing SCMP on Dec 10th , 2025 that Alibaba has set up a new division, Qwen Consumer Business Group, led by Wu Jia, to overseas Qwen chatbot app, Quark AI assistant, Cloud Drive, AI hardware products and UC Browser and online reading platform Shuqi. The move suggests that Alibaba tried to consolidate the management of consumer use of its AI offering. With the establishment of ATH, Qwen Business Group is now grouped under ATH with the focus and responsible on "applying" token on consumer end products.

Wukong: AI-Native Agentic Platform

With the establishment of Alibaba Token Hub, Alibaba also unveiled Wukong, an AI-native enterprise platform designed to bring advanced agentic capabilities directly into business workflows. Built on enterprise-grade security infrastructure, the platform can coordinate multiple agents to handle complex tasks within a single interface, making it a powerful productivity tool tailored to the demands of modern business environments. Wukong will be led by Chen Hang. Wukong is also integrated with Alibaba's broader ecosystem such as Taobao, Tmall, 1688, Alipay, and Alibaba Cloud with modular agent skills. This positions Wukong to support diverse business functions—from e-commerce storefront design and supplier management to payment processing and cloud infrastructure orchestration. The platform also supports third-party agent skill integrations.

AI Innovation Business Unit

The AI innovation Business Unit leads the role to actively explore new AI application. It will in-charge of incubating cutting-edge AI products and explore the commercialization and implementation of the products and the unit will be led by Group CEO Eddie Wu.

Tongyi Lingma (AI Coding Assistant)

Tongyi Lingma, initially launched in 2023, is an intelligent coding assistant developed by Alibaba Cloud to leverage Ai to enhance programming experience. It offers automatic code continuation at line and function levels, unit test generation, converting natural language into code, generating comments, single-line completions and explaining code. It is often time referred to be comparable to GitHub Copilot or Cursor. It is an AI "coding copilot + agent" embedded inside the

www.reboard.com.cn

computer IDE, it helps developers write, edit, debug, and build apps automatically using natural language prompts. Lingma is equipped with comprehensive codebase awareness and advanced tool-calling capabilities and powered by SOTA models. In Ask mode, Lingma delivers instant and accurate solutions. In Agent mode, Lingma autonomously devises the objectives, breaks it down to manageable To-dos according user requirement and dynamically adjust the To-dos throughout the process to accomplish coding tasks end-to-end. Examples of customers adoption stories include China Insurance, FinVolution Group, GaiaWorks and Hello Group.

Foundation model – Qwen series

- **2023 -** Alibaba published its first open-source model, the 7bn parameter LLM Qwen-7B and Qwen-7B-Chat, through ModelScope and Hugging Face in August 2023. The models' code, model weights, and documentation were made freely accessible to academics, researchers, and commercial institutions worldwide. Subsequently in late Aug 2023, Ali launched two more open-source model, large vision language models (LVLM), Qwen-VL and Qwen-VL-Chat. Qwen-VL is the multi-modal version of Qwen-7B, and capable of understanding both image inputs and text prompts in English and Chinese.

- **2024** - In June 2024, Ali released Qwen2 series, and in Sept, it unveiled 100 new open-source Qwen2.5 models ranging from 0.5 to 7.2 bn parameters. In Nov, it released a reasoning Ai model QwQ (Qwen with Questions), a 32B-Preview. First reasoning AI model to make open source.

- **2025** – Latest visual-language open model, Qwen2.5-VL was released in January. This is a multimodal model that serves as a visual agent to facilitate the execution of simple tasks on computers and mobile services. In April, Qwen 3 family featuring MoE architecture was released. Qwen3-Max, Qwen3-VL and Qwen3-Next were released in Sept. In Nov, Alibaba rebrand its Tongyi app to Qwen app and Qwen app reaches over 10mn downloads in one week. In Dec 2025, Qwen3-TTS was upgraded and open-source Qwen-Image-Layered and Qwen-Image-2512.

- **2026** – In Jan, Alibaba announced that Qwen app integrates with Taobao and Alipay and Qwen3-Max-Thinking was release. On Feb 16, 2026, BABA released Qwen3.5 and Qwen3.5-Plus (open-weight) and Qwen3.5-Pax-Preview was released in March. On April 2, Qwen3.6-Plus focusing on agentic coding and reasoning was released. On April 15, Qwen3.6-35B-A3B was released as open source. April 27, 2026, BABA launched HappyHorse-1.0, an image-to-video model.

As of late 2025, there were over 600mn downloads and 170,000+ derivative models on Hugging Face that are based on Qwen model family, making it one of the most popular open-source models. In summary, Qwen model consists of Text (Qwen2.5, Qwen3), Reasoning (QwQ, Qwen3-Max-Thinking), Code (Qwen-Coder, Qwen3-Coder), Vision-Language (Qwen-VL, Qwen2.5-VL), Audio (Qwen-Audio), Video (Wan series), Math (Qwen2-Math).

Specifically, according an article on Towards Ai on April 12, 2026, Qwen3.6-Plus was reportedly hit 1 trillion tokens in a single day on OpenRouter.

Per an article by AI Base dated Sep 1, 2025, citing a report by Frost & Sullivan that the enterprise-level large model market in China experienced explosive growth with 1H25 daily total token consumption of enterprise-level large models in China

www.futoudaifxin.cn

reached 10.2 trillion, or 363% increase compared to 2H24. It noted that Alibaba Tongyi Qwen models lead with 17.7% market share as the most commonly chosen general-purpose large model by Chinese enterprises. This is followed by Bytedance Doubao at 14.1%, DeepSeek 10.3%.

Figure 17. Qwen model series development

Model series	Release / open-source date	Key model sizes (parameters)	Context length	Special Features / Breakthroughs
Qwen	11-Apr-23	1.8B, 7B, 14B, 72B	8,192 tokens (extended from 2,048)	First major open-weight LLM from Alibaba; strong bilingual (Eng/Chi) performance; introduced Qwen-Chat variants
Qwen1.5	6-Feb-24	1.8B, 7B, 32B, 72B, 110B	32,768 tokens	Improved alignment & multilingual support (~30 languages); added 110B flagship; better instruction following
Qwen2	7-Jun-24	0.5B, 1.5B, 7B, 57B-A14B (MoE), 72B	131,072 tokens	Introduced MoE architecture (57B-A14B); trained on 7T+ tokens; strong coding & math uplift; 30-language support
Qwen2.5	19-Sep-24	0.5B, 1.5B, 3B, 7B, 14B, 32B, 72B	Up to 1,010,000 tokens (1M variant)	Trained on 18T tokens; specialized sub-series: Qwen2.5-Coder, Qwen2.5-Math, Qwen2.5-VL; 1M context window introduced
QwQ-32B-Preview	28-Nov-24	32B	32,768 tokens	Dedicated reasoning/thinking model; chain-of-thought focus; outperformed OpenAI o1 on select benchmarks
Qwen2.5-1M	27-Jan-25	7B, 14B	1,010,000 tokens	World-class long-context model; 100% accuracy on 1M Passkey Retrieval; RULER score 93.1 (surpassed GPT-4's 91.6)
Qwen3	29-Apr-25	0.6B, 1.7B, 4B, 8B, 14B, 32B (dense); 30B-A3B, 235B-A22B (MoE)	32K native / 128K with YaRN (dense); 256K (MoE)	Hybrid Thinking Mode (switch between reasoning & instruct); 119-language support; MoE reintroduced at scale; strong agentic/tool-use capabilities
Qwen3-Coder	23-Jul-25	~32B+ (Coder-Plus)	128K+ tokens	Specialized coding agent; SWE-Bench Verified: 69.6% – surpassed Claude 3.5 Sonnet (49%) & GPT-4 Turbo (43.8%)
Qwen3-Max	24-Sep-25	>1 Trillion (proprietary API)	1,000,000 tokens	Largest Qwen model; pretrained on 36T tokens; 100% on AIME'25 & HMMT with tool use; LiveCodeBench v6: 74.8
Qwen3-VL	24-Sep-25	4B, 8B, 235B-A22B	131,072 tokens	Multimodal vision-language; "Thinking" variant for visual reasoning; strong OCR and document understanding
Qwen3.5	16-Feb-26	0.8B, 2B, 4B, 9B, 27B, 35B-A3B, 122B-A10B, 397B-A17B (MoE)	262K native / up to 1M	Ultra-large MoE flagship (397B-A17B); broad size range for edge-to-cloud deployment; continued hybrid thinking
Qwen3.6	2-Apr-26	27B (dense), 35B-A3B (MoE), Plus (API)	262K native / up to 1M	Latest generation; Qwen3.6-Plus offers 1M context window; ultra-efficient sparse MoE (only 3B activated out of 35B)

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

www.futoucaixin.cn

Figure 18. Qwen sub model series highlights

Sub Series	Base Series	Focus Area	Key Highlight
Qwen-VL	Qwen1/2/2.5/3	Vision-Language (multimodal)	Combines vision transformer + LLM; supports image/video understanding
Qwen-Audio	Qwen2/2.5	Audio / Speech	Speech recognition, synthesis, and audio understanding
Qwen2.5-Coder	Qwen2.5	Code Generation	Matched GPT-4o on coding benchmarks (Nov 2024)
Qwen2.5-Math	Qwen2.5	Mathematics	Specialized for complex mathematical reasoning
QwQ	Qwen2.5/3	Reasoning / Thinking	Chain-of-thought reasoning; outperformed OpenAI o1 on select tasks
Qwen3-Coder	Qwen3	Coding Agents	SWE-Bench Verified 69.6%— best open-source coding agent (Nov 2025)
Qwen-Image	Qwen3	Image Generation	Advanced text rendering, image editing, multi-style generation

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports, Citi Research

Figure 19. Qwen model performance comparison

Model	License	Context Window	Input / Output ($/1M)	Artificial Analysis			Other benchmarks				
				Intelligence Index	Speed (output token/s)	Blended Price ($/1M token)	GDPval-AA	τ²-Bench Telecom	Humanity's Last Exam	GPQA Diamond	MMMU-Pro
Qwen3.6											
Qwen3.6 Max Preview (Reasoning)	Proprietary	256K	$1.30 / $7.80	52	37	2.9	50%	96%	29%	89%	–
Qwen3.6 Plus (Reasoning)	Proprietary	1M	$0.50 / $3.00	50	53	1.1	43%	98%	26%	88%	78%
Qwen3.6 27B (Reasoning)	Open	262K	$0.60 / $3.60	46	66	1.4	46%	94%	22%	84%	75%
Qwen3.6 27B (Non-Reasoning)	Open	262K	$0.60 / $3.60	37	69	1.4	45%	94%	14%	83%	72%
Qwen3.6 35B-A3B (reasoning)	Open	262K	$0.248 / $1.485	43	197	0.6	40%	95%	20%	84%	75%
Qwen3.5											
Qwen3.5 397B-A17B (Reasoning)	Open	262K	$0.60 / $3.60	45	53	1.4	35%	96%	27%	89%	77%
Qwen3.5 122B-A10B (Reasoning)	Open	262K	$0.40 / $3.20	42	148	1.1	31%	94%	23%	86%	75%
Qwen3.5 35B-A3B (Reasoning)	Open	262K	$0.25 / $2.00	37	182	0.7	27%	94%	–	–	–
Qwen3.5 27B (Reasoning)	Open	262K	$0.30 / $2.40	42	91	0.8	33%	94%	22%	86%	75%
Qwen3											
Qwen3 Max (Thinking)	Proprietary	256K	$1.20 / $6.00	40	48	2.4	32%	84%	26%	86%	–

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company Reports, Artificial Analysis, Citi Research

Qwen model series & benchmark ranking comparison

Qwen3.6-Max

The latest Qwen model – Qwen3.6-Max-Preview was released on April 20, 2026 which highlighted it took first place on SWE-bench Pro, Terminal Bench2.0, Skills Bench, SciCode and QwenClawBench and QwenWebBench. While this model is top of the line, it is not an open-weight model, it is proprietary and text-only AI model, and only be able to access through Alibaba Cloud Model Studio, Qwen Chat interface and Open Router. It is designed for advanced coding, reasoning and instruction following. It does outperform Qwen3.6-Plus in agentic coding (SWE-bench Pro), but it only features 256k context window, compared to 1mn context window for Qwen3.6-Plus.

Qwen3.6-Plus

Released on April 2, 2026, Qwen3.6-Plus is a 1m context window model with significantly improved agentic coding capability and with better multimodal perception and reasoning ability. Qwen3.6-Plus receives comprehensive improvements in coding agents, general agents and tool usage by deeply integrating reasoning, memory and execution capabilities. In the coding agents, Qwen3.6-Plus demonstrates strong practical engineering performances. It also achieves top results in multiple challenging long-horizon planning tasks and evaluating STEM reasoning and precise information extraction from ultra-long

www.tutoucaixin.cn

contexts. The model supports text, image and video input, output text. According to Artificial Analysis, the model scores a 50 on AA Intelligence Index, ranking at 19 out of 145. Pricing is US$0.50 per 1M input tokens and US$3.00/M output tokens.

Qwen 3.5

With the release of Qwen3.5 series announced on Feb 16, 2026, Alibaba also introduce the open-weight version of 3.5 with Qwen3.5-397B-A17B. It is a native vision-language model achieving outstanding results across full range of benchmark evaluation including reasoning, coding, agent capabilities and multimodal understanding. The model is built on an innovative hybrid architecture that fuses linear attention with a sparse MoE and achieve remarkable inference efficiency.

Figure 20. Qwen3.6 Max (Preview) and Qwen3.6 Plus benchmark comparison

Benchmark	Context	Qwen 3.6 Max (Preview)	Qwen 3.6 Plus	Qwen 3.5 Plus	Claude 4.5 Opus	GLM 5.1
SuperGPQA	Garduate-Level Knowledge	73.9	71.6	70.4	70.6	68.0
AA-Omniscience Index	Knowledge Reliability and Hallucination	10.0	3.0	-30.0	13.0	2.0
GDPval-AA	Real-World Valuable Task	51.0	43.0	35.0	48.0	52.0
QwenChineseBench	Chinese Real-World Knowledge	84.0	78.7	78.8	69.0	81.2
QwenClawBench	Real-World Agent	59.0	57.2	51.8	52.3	58.7
SkillsBench	Agent Skills	55.6	45.7	30.0	45.3	53.1
ToolcallFormatIFBench	Real-World Toolcall Following	86.1	83.3	80.1	84.2	60.1
QwenWebBench (Elo Rating)	Artifacts	1532	1495	1182	1530	1558
SciCode	Research Coding	47.0	40.7	42.0	49.5	43.8
NL2Repo	Long-Horizon Coding	42.9	37.9	32.2	43.2	42.7
Terminal-Bench 2.0 (Terminus-2)	Agentic Terminal Coding	65.4	61.6	52.5	59.3	63.5
SWE-Bench Pro	Agentic Coding	57.3	56.6	50.9	57.1	58.4

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Alibaba, Citi Research

Figure 21. Qwen3.6-27B (dense) and Qwen3.6-35B-A3B (MoE) benchmark comparison

Benchmark	Context	Dense			MoE		
		Qwen3.6-27B	Qwen3.5-27B	Gemma4-31B	Qwen3.6-35B-A3B	Qwen3.5-397B-A17B	Claude 4.5 Opus
Terminal-Bench 2.0	Agentic Terminal Coding	59.3	41.6	42.9	51.5	52.5	59.3
SWE-Bench Pro	Agentic Coding	53.5	51.2	35.7	49.5	50.9	57.1
SWE-Bench Verified	Agentic Coding	77.2	75.0	52.0	73.4	76.2	80.9
SWE-Bench Multilingual	Multilingual Agentic Coding	71.3	69.3	51.7	67.2	69.3	77.5
QwenClawBench	Real-World Agent	53.4	52.2	41.7	52.6	51.8	52.3
QwenWebBench (Elo Rating)	Artifacts	1487	1068	1197	1397	1186	1536
NL2Repo	Long-Horizon Coding	36.2	27.3	15.5	29.4	32.2	43.2
SkillsBench	Agent Skills	48.2	27.6	23.6	28.7	30.0	45.3
Claw-Eval (pass^3)	Real-World Agent	60.6	46.2	25.0	50.0	48.1	59.6
GPQA Diamond	Graduate-Level Reasoning	87.8	85.5	84.3	86.0	88.4	87.0
MMMU	Multimodal Reasoning	82.9	82.3	80.4	81.7	85.0	80.7
RealWorldQA	Image Reasoning	84.1	83.7	72.3	85.3	83.9	77.0

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Alibaba, Citi Research

www.futoucaixin.cn

Figure 22. Artificial Analysis Intelligence Index leaderboard (May 7, 2026)

Lab	Leading Model	Artificial Analysis Intelligence Index	Speed (Output Tokens / Second)	Blended Price (US$ / 1M Tokens)	Context Window (1K Words)	Token usage (mn)
OpenAI	GPT-5.5 (xhigh)	60	67	11.25	922	300
Anthropic	Claude Opus 4.7 (max)	57	45	10.94	1000	480
Kimi	Kimi K2.6	54	30	1.71	256	470
Xiaomi	MiMo-V2.5-Pro	54	58	1.50	1000	280
Meta	Muse Spark	52	na	na	262	190
Alibaba	Qwen3.6 Max Preview	52	36	2.92	256	290
DeepSeek	DeepSeek V4 Pro (Max)	52	34	2.17	1000	430
Z.ai	GLM-5.1	51	54	2.15	200	150
MiniMax	MiniMax-M2.7	50	44	0.52	205	320
Tencent	Hy3-preview	42	85	0.00	256	na
Mistral	Mistral Medium 3.5	39	148	3.00	256	na
StepFun	Step 3.5 Flash 2603	38	176	0.00	256	na
Nvidia	Nvidia Nemotron 3 Super	36	167	0.41	1000	320
Amazon	Nova 2.0 Pro Preview (medium)	36	143	3.44	256	120

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Artificial Analysis, Citi Research

Figure 23. GDPval-AA ELO leaderboard (May 7, 2026)

Lab	Leading Model	GDPval-AA ELO	Token usage (mn)
OpenAI	GPT-5.5 (xhigh)	1774	190
Anthropic	Claude Opus 4.7 (max)	1753	340
Xiaomi	MiMo-V2.5-Pro	1573	130
DeepSeek	DeepSeek V4 Pro (Reasoning)	1558	190
Z.ai	GLM-5.1 (reasoning)	1535	na
MiniMax	MiniMax-M2.7	1508	180
Alibaba	Qwen3.6 Max Preview	1506	170
Kimi	Kimi K2.6	1483	na
Meta	Muse Spark	1422	83
Tencent	Hy3-preview (Reasoning)	1237	na
Mistral	Mistral Medium 3.5	1170	na

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Artificial Analysis, Citi Research

Figure 24. SuperCLUE composite rank as of Mar 2026 (score breakdown)

Rank	Model	Institution	Open-source?	Composite score	Math reasoning	Vision control	Sci reasoning	Instruct follow	Code generation	Agent	Geography	Deployment
1	Claude-Opus-4.6 (max)	Anthropic	No	77.02	85.71	82.95	85.37	47.57	71.15	89.35	Overseas	API
2	GPT-5.4 (xhigh)	OpenAI	No	72.48	88.89	85.43	84.15	44.32	52.05	80.04	Overseas	API
3	Doubao-Seed-2.0-pro-260215 (high)	ByteDance	No	71.53	84.87	79.41	80.49	39.46	63.93	81.04	Domestic	API
4	Kimi-K2.5-Thinking	Moonshot	Yes	64.60	81.51	77.61	68.29	16.22	65.50	78.44	Domestic	API
5	Qwen3.5-397B-A17B-Thinking	Alibaba	Yes	64.48	84.87	84.39	75.61	19.46	51.04	71.52	Domestic	API
6	GLM-5	Z.ai	Yes	64.27	73.95	86.65	75.00	24.86	58.32	66.64	Domestic	API
7	DeepSeek-V3.2-Thinking	DeepSeek	Yes	61.92	78.15	77.23	73.17	25.95	60.43	56.62	Domestic	API
8	MiMo-V2-Pro	Xiaomi	No	60.67	84.03	73.80	74.39	16.22	59.61	55.97	Domestic	API
9	Tencent HY2.0 Think	Tencent	No	59.16	76.47	76.46	70.73	14.05	57.58	59.68	Domestic	API
10	Qwen3.5-122B-A10B-Thinking	Alibaba	Yes	58.53	82.35	70.50	69.51	13.51	50.18	65.15	Domestic	API
11	LongCat-Flash-Thinking-2601	Meituan	Yes	57.47	77.31	66.31	69.51	11.89	51.86	67.94	Domestic	API
12	GPT-OSS-120b (high)	OpenAI	Yes	57.07	79.83	54.88	73.17	20.54	52.89	61.12	Overseas	API
13	Step-3.5-Flash	StepFun	Yes	56.23	80.67	63.94	68.29	9.73	50.71	64.06	Domestic	API
14	MiniMax-M2.5	MiniMax	Yes	55.96	73.95	67.41	65.85	7.57	55.33	65.64	Domestic	API
15	MiniMax-M2.7	MiniMax	No	55.68	78.15	55.61	64.63	17.30	58.74	59.68	Domestic	API
16	Spark-X2	iFlyTek	No	52.79	78.15	62.23	71.95	1.08	41.19	62.13	Domestic	API
17	MiMo-V2-Flash	Xiaomi	Yes	49.97	69.75	63.45	53.66	2.16	54.46	56.37	Domestic	API
18	Mistral Large 3	Mistral AI	Yes	41.06	48.74	51.67	54.88	4.32	42.24	44.51	Overseas	API
19	Llama-4-Maverick-17B-128E-Instruct	Meta	Yes	36.70	38.66	66.74	47.56	3.78	39.67	23.70	Overseas	API

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: SuperCLUE, Citi Research

www.futoucaixin.cn

Figure 25. Artificial Analysis text-to-video leaderboard (no audio, May 7, 2026)

Rank	Creator	Model	ELO	Samples	Release date	API pricing
1	Alibaba-ATH	HappyHorse-1.0	1,355	8,342	Apr-26	$14.40 / min
2	ByteDance Seed	Dreamina Seedance 2.0 720p	1,271	8,658	Mar-26	na
3	Kling AI	Kling 3.0 1080p (Pro)	1,250	5,802	Feb-26	$13.44 / min
4	Kling AI	Kling 3.0 Omni 1080p (Pro)	1,233	5,218	Feb-26	$13.44 / min
5	Video Rebirth	Bach-1.0 Preview	1,225	3,508	Apr-26	na
6	Kling AI	Kling 3.0 Omni 720p (Standard)	1,224	5,241	Feb-26	$10.08 / min
7	Vidu	Vidu Q3 Pro	1,223	6,583	Jan-26	$9.60 / min
8	PixVerse	PixVerse V6	1,221	9,142	Mar-26	$5.40 / min
9	PixVerse	PixVerse V5.6	1,221	5,940	Feb-26	$9.00 / min

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Artificial Analysis, Citi Research

Figure 26. Artificial Analysis image-to-video leaderboard (no audio, May 7, 2026)

Rank	Creator	Model	ELO	Samples	Release date	API pricing
1	Alibaba-ATH	HappyHorse-1.0	1,397	7,987	Apr-26	na
2	ByteDance Seed	Dreamina Seedance 2.0 720p	1,348	4,836	Mar-26	na
3	PixVerse	PixVerse V6	1,324	8,762	Mar-26	$5.40 / min
4	Vidu	Vidu Q3 Pro	1,288	6,359	Jan-26	$9.60 / min
5	Kling AI	Kling 2.5 Turbo 1080p	1,284	4,463	Sep-25	$4.20 / min
6	Kling AI	Kling 3.0 1080p (Pro)	1,282	5,740	Feb-26	$13.44 / min
7	Kling AI	Kling 3.0 Omni 1080p (Pro)	1,281	5,174	Feb-26	$13.44 / min
8	PixVerse	PixVerse V5.6	1,281	5,284	Feb-26	$9.00 / min
9	Kling AI	Kling 2.6 Standard (January)	1,271	5,172	Jan-26	na

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Artificial Analysis

AI Application integration with Ali's ecosystem

Agentic AI development for AliCloud

Qwen integration with Alibaba ecosystem

Back on Jan 15 (Securities Times CN dated 15 Jan 2026, link), Alibaba officially announced that Qwen app has integrated with Alibaba ecosystem including Taobao, Alipay, Taobao Shangou and Amap for AI functionalities supporting food delivery, online shopping, flight ticket booking, route planning etc. The integration feature is open for all users for testing, together with beta testing for Qwen app mission assistant.

Users who downloaded the latest version of Qwen app at that time can enable application authorization for Taobao, Taobao Shangou, Fliggy, Alipay and Alipay AI to deploy their missions. At the same time, Qwen app also upgraded with over 400 AI office functions.

Alibaba also shared that Qwen app will leverage real world knowledge and the transactional and service data from Alibaba to enhance its model training in order to facilitate objective and accurate decision-making for AI shopping functions. Looking ahead, Ali will focus on developing agentic AI products such as AI shopping for Taobao marketplace.

www.futoucaixin.cn

Later reported by 36Kr dated 9 Feb 2026 (link), Qwen app has launched a Rmb3bn marketing campaign during Chinese New Year starting Feb 6 to subsidize users to place food delivery orders (particularly tea beverages) on Qwen app at a delivery fee of Rmb0.01. This also includes Rmb2bn for delivery fee waiver and Rmb1bn for cash envelope subsidies. According to Qwen, the campaign generated over 10mn orders in the first 9 hours with Qwen once becoming the most downloaded app on China App Store.

Agentic products

On Apr 23, AliCloud officially launched an enterprise-level agent platform **JVS Crew** which enable enterprises to integrate with existing apps, SaaS and smart gadgets to acquire zero-barrier enterprise-level agentic capabilities. JVS Crew is embedded with 3 layers of security walls including identity wall, content wall and execution wall to support safe execution together with full-managed conversational management, cost management and channel integration. JVS Crew also adopted Harness engineering to separate reasoning and execution.

On Apr 28, Ali announced to upgrade its **QoderWork** with 10 expert modules covering six verticals including finance, legal, marketing, taxation, consulting and products. For example, the financial analysis modules support data research, financial information analysis, peer analysis and presentation generation. Within marketing module, AI can deploy market research, industry analysis, peer product analysis, strategic implementation and data output. Going forward, QoderWork aims to expand into more modules such as data and HR sectors.

On Apr 30, Alibaba officially released two agentic products, including digital employee **QoderWake** and **Qoder** mobile to fully cover enterprise and consumer use cases. QoderWake is positioned as a safe, self-evolutionary productive digital employee product to act as software engineer, operators and analyst roles. Currently QoderWake has opened for beta registration in which individuals and enterprises to apply for one or multiple digital employees and customize their features based on business or individual needs.

QoderWake adopts the latest Harness-First framework which can summarize and memorize its experience into memory, skills, strategies verification and workflows with embedded anti-rot governance. Currently the digital programmer feature of QoderWake has been launched within Alibaba which significantly improved productivity. Going forward, QoderWake will further release new features such as digital analyst, digital customer management, digital content editor and digital workflow staff.

Through Qoder mobile app, users can remotely manipulate its Qoder PC products to complete any missions. In contrast to agentic product on instant messenger app, Qoder mobile can directly display the chain of thought, workflow, pop-ups and user confirmation which significantly improve user experience. Qoder mobile has fully integrated with Qoder CLI capabilities and will soon integrate with full agentic products such as Qoder IDE, QoderWork, QoderWake.

Announcement of Qwen integrating to Taobao

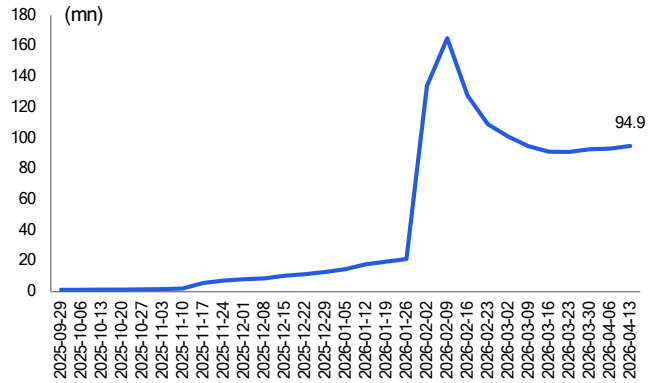
On May 10, 2026, according to various news media including Reuters (10 May 2026), reported that Alibaba is working to deepen integration of Qwen with its Taobao marketplace, enabling users to browse, compare and purchase items via Qwen app by chatting with the AI agent, rather than manually navigating product listings. According to the report, Qwen app will gain access to the entire Taobao and Tmall catalog of over 4bn products, which is backed a “skills library” capable of

www.reuterxrm.cn

managing logistics and post-sales services. It will also offer shopping recommendation based on users' order history and shopping preferences.

In addition, a Qwen-powered AI shopping assistant will be launched on Taobao app, offering tools including virtual try-ons and 30-day price tracking.

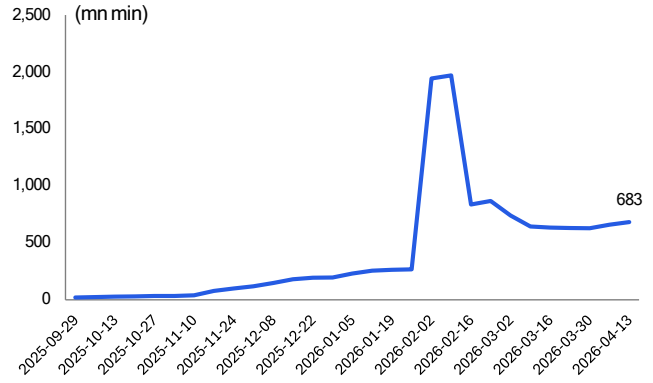
Figure 27. Qwen WAU from Oct'25 to Apr'26



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Questmobile, Citi Research

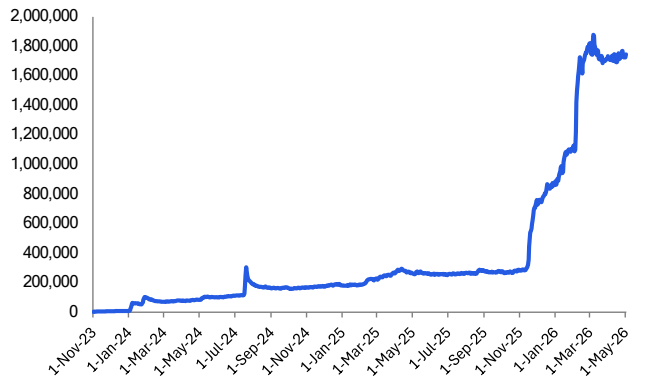
Figure 28. Qwen weekly time spent from Oct'25 to Apr'26



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Questmobile, Citi Research

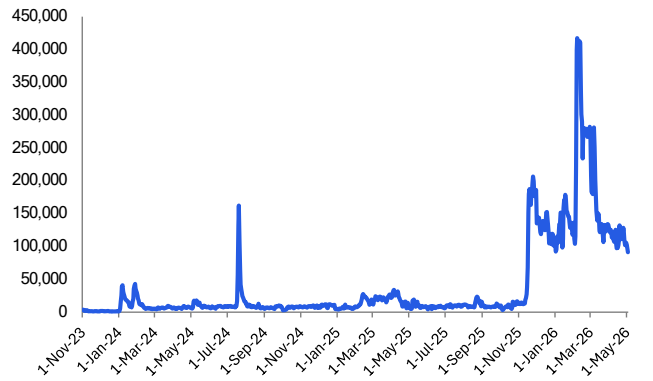
Figure 29. Qwen DAU trend on iOS in China



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: SensorTower, Citi Research

Figure 30. Qwen daily download trend on iOS in China

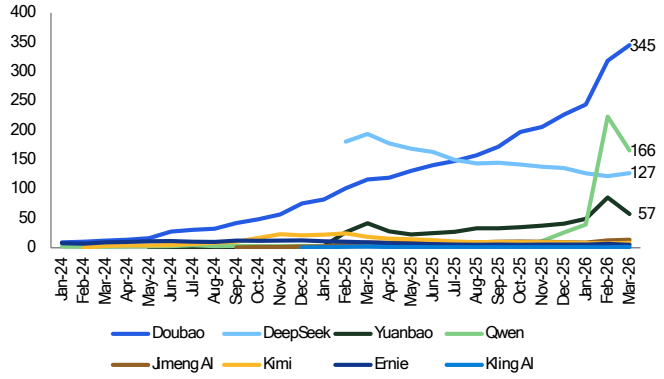


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: SensorTower, Citi Research

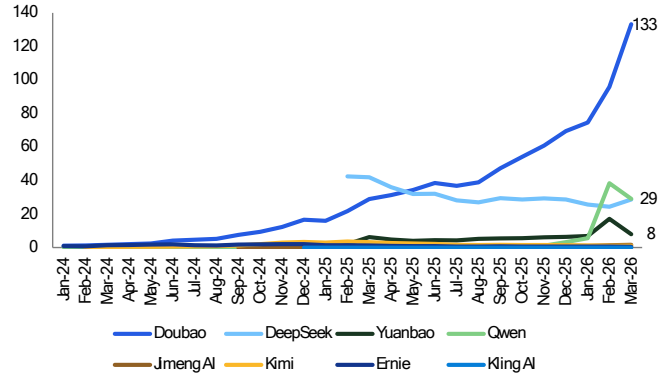
www.futoucaixin.cn

Figure 31. China AIGC app MAU (mn)



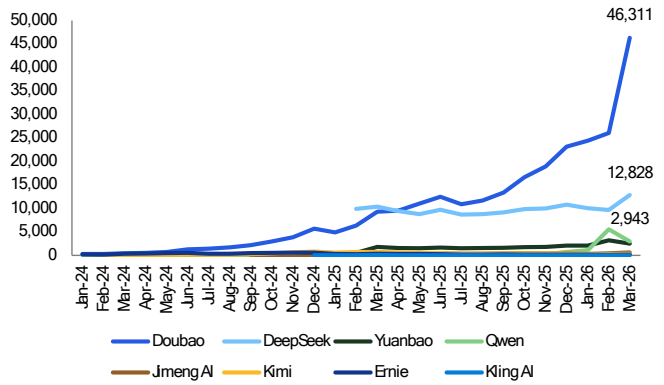
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Questmobile, Citi Research

Figure 32. China AIGC app DAU (mn)



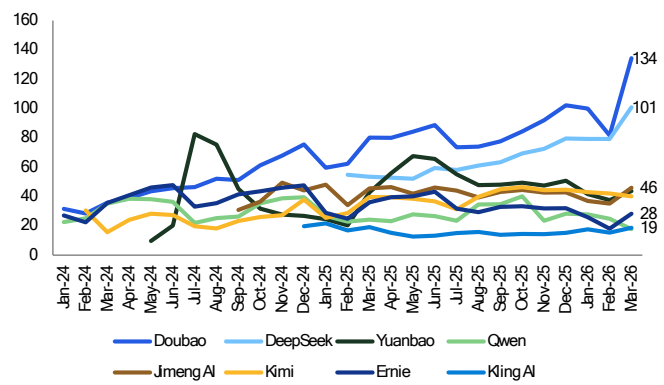
© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Questmobile, Citi Research

Figure 33. China AIGC app total time spent (mn min)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Questmobile, Citi Research

Figure 34. China AIGC app average time spent (min)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Questmobile, Citi Research

Token Economies and usage scale

As we noted in our MiniMax Initiation report, we wrote that as highlighted by Nvidia on its blog (dated Mar 17, 2025), tokens are units of data processed by AI models during training and inference, which enable prediction, generation and reasoning. The faster tokens can be processed, the faster models can learn and respond. Tokenization is the process of converting data into standardized units called tokens. Transformer AI models use this method to process text, images, audio clips, or videos, and translate the information into token format. As highlighted by the explanation, **AI factories are designed to generate intelligence** on a massive scale by processing data in the form of tokens and turning them into valuable, monetizable insights. As a result, many **AI services are now adopting pricing models based on token consumption**, where the cost is determined by the number of tokens a model processes for both input and output.

Most AI model providers charge on the per-million tokens basis (M tokens). For any API call to the model, users are charged on input and output price. Typically, most models charge cheaper on **input tokens** which are primary prompts entered by users and models "reading" user instructions and context. **Output** prices are

www.tutodaixin.cn

usually more expensive (3-5x more than input) and it covers intensive compute required for the model to "write" the response.

Comparing China vs global token usage

The latest proliferation of agentic AI and demand of high-speed video generation has led to a surge in token consumption in China, which has recently surpassed the US as the largest token consumption country globally.

Sina news reported on Mar 30th, 2026 that according to the latest data from the National Data Administration, as of Mar 2026, China's average daily AI token usage has exceeded 140trn, an increase of >1,400 times from the 100 billion at the beginning of 2024. The data was disclosed by Mr. Liu Liehong, director of the National Data Administration, during the 2026 annual meeting of the China Development High-level Forum. The article also highlighted the growth trend from 10bn in early 2024 to 30bn in June 2025 and later to 100bn in Dec 2025 and surpassing 140bn in Mar 2026. The article also highlighted that usage from enterprises is also showing rapid growth. In 2H25, average daily call volume for enterprise large model APIs reached 37trn tokens, an increase of 263% from 10.2trn in 1H25. The report also highlighted that in Feb 2026 China's AI model weekly token calls surpassed those from the US for the first time. Global Times reported on Mar 29, 2026 that the recent surge in token usage drove up computing power and related services, and prompted major cloud providers in China to announce an average price increase of 30%+ for their AI cloud products.

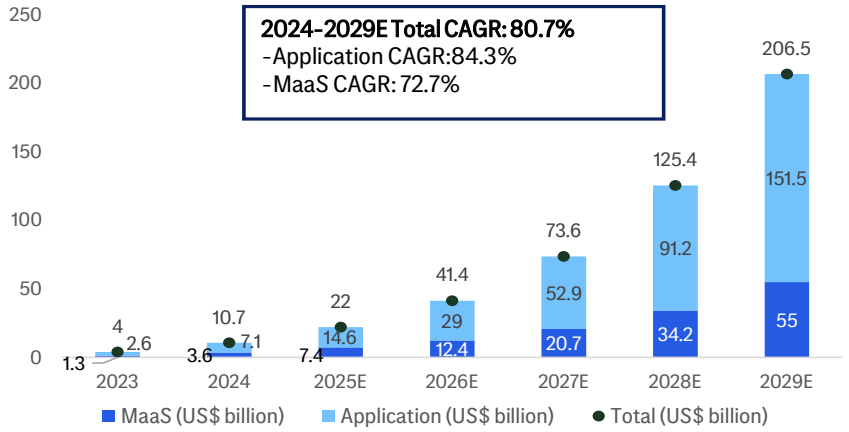
Global foundation model market size

According to CIC, the global foundation model will reach US$206.5bn, representing a 2024-2029 CAGR of 80.7%. Within that, MaaS market is expected to grow at a 72.7% CAGR, from US$7.4bn to US$55bn, while the application market size is forecast to grow from US$14.6bn to US$151.5bn or an 84.3% CAGR.

MaaS is a business model where revs are charged on cloud-based APIs (Application Programming Interfaces) and licensing agreements. Developers and enterprises can then integrate these powerful model functions directly into their own products, applications, or internal systems. The revenue for MaaS is generated on a usage basis, meaning customers pay for processing or the number of API calls they make. Application-based revenue is generated by offering complete, end-user-facing software products that are built on top of a foundational AI model. These applications are sold directly to both individual consumers and enterprise clients, most commonly through a subscription model.

www.futoucaixin.cn

Figure 35. Global foundation model market size in terms of model-based rev

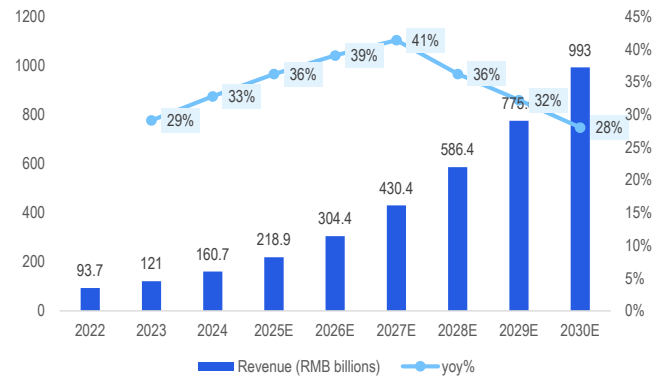


© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, CIC

China AI market TAM

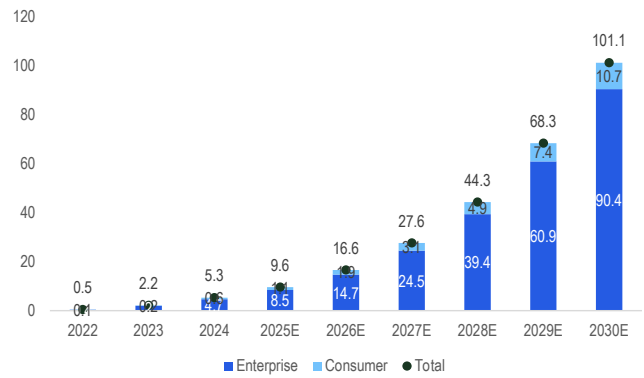
According to Frost & Sullivan and China Academy of Information and Communication Technology (CAICT), China's AI market reached Rmb160.7bn or US$23bn in 2024 and they had an estimate of +36% yoy to Rmb219bn (US$31.3bn) for 2025. The AI market is expected to reach Rmb993bn or US$141.9bn in 2030, suggesting a 2024-30 CAGR of 35.5%, per their estimates.

Figure 36. Market Size of AI Market in China in terms of revenue (RMB billions)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Frost & Sullivan, China Academy of Information and Communications Technology (CAICT)

Figure 37. Size of the LLMs Market in China in terms of revenue (RMB billions)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Citi Research, Frost & Sullivan, CAICT

Knowledge Atlas (Zhipu)'s prospectus quoting Frost & Sullivan and CAICT said that the size of the LLM market in China reached Rmb5.3bn in 2024, of which enterprise-based revs contributed Rmb4.7bn or 89% and consumer-based revs only Rmb0.6bn (11%). Th prospectus further said that the total LLM market will reach Rmb101.1bn in 2030, representing a 2024-30 CAGR of 63.5%. Within that, the enterprise-based market is expected to reach Rmb90.4bn by 2030 or a 63.7% CAGR while the consumer-based market size is forecast to reach Rmb10.7bn by 2030 or a 61.6% CAGR.

www.futuresixin.com

OpenRouter model comparison for token usage

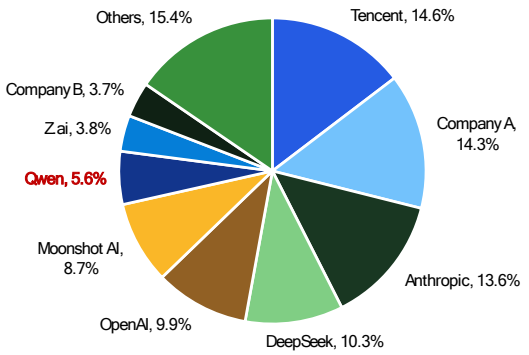
As noted in the below charts by Open Router usage tracking, as of May 4, 2026, top five token usage companies are Tencent at 14.6% (mainly driven by its latest release of HY3 Preview which is free currently), Anthropic 13.6%, DeepSeek 10.3% (following the release of V4 recently), OpenAI 9.9% and BABA's Qwen usage is at 5.6%.

Figure 38. Weekly usage of models across OpenRouter (week of May 4, 2026, trn)

Rank	Lab	Model	Token used (trn)
1	Tencent	HY3 Preview (free)	1.83
2	Moonshot	Kimi K2.6	0.95
3	Anthropic	Claude Sonnet 4.6	0.74
4	Anthropic	Claude Opus 4.7	0.49
5	DeepSeek	DeepSeek V4 Flash	0.45
6	DeepSeek	DeepSeek V3.2	0.38
7	DeepSeek	DeepSeek V4 Pro	0.37
Others			7.30
Total			12.50

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: OpenRouter, Citi Research

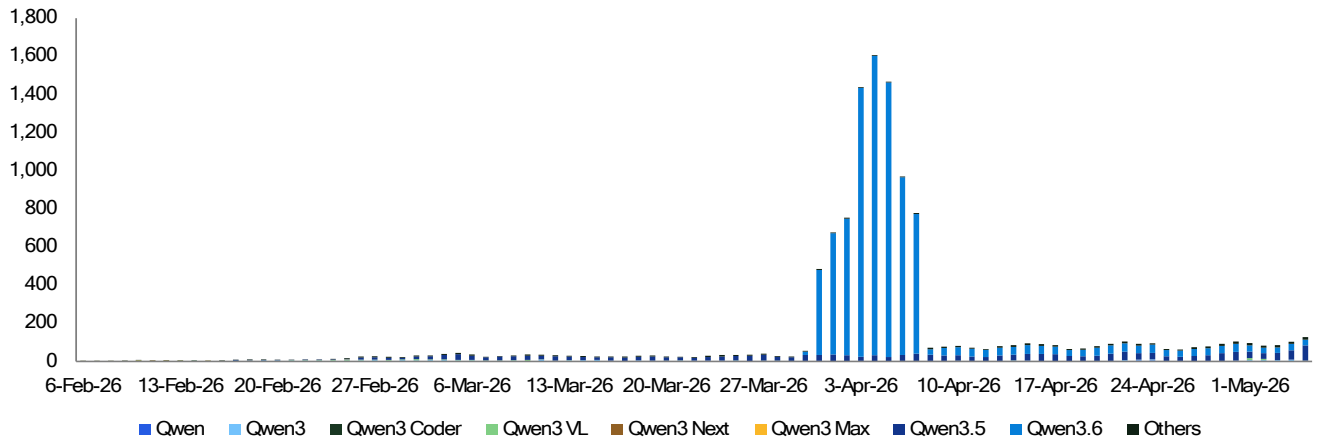
Figure 39. Token usage share across OpenRouter by model author (week of May 4, 2026)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: OpenRouter, Citi Research

The below chart shows that since the release of Qwen3.6, the token usage of Qwen models has jumped significantly in April before tapering off quickly.

Figure 40. AliCloud tokens processed on OpenRouter by products (bn)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: OpenRouter, Citi Research

According to several news articles including Towards AI dated April 12, 2026 that Qwen3.6 Plus became the first model in history to process over 1 trillion tokens in a single day on Open Router (reaching 1.4 trillion token or 711% surge (by Biggo.com dated April 7) in call volume on the platform), following the release of the model on April 2, 2026 where it started as free with 1mn-token context window compared to US$5 per million input tokens charged by Claude Opus 4.6 at the time. We **note**

www.futuoa.ixin.com

that the token usage tracking by OpenRouter tend to fluctuate a lot depending on the promotion of the model usage by model companies especially when there is newer version of the model release.

Qwen API pricing

Figure 41. Qwen models – general language models token pricing (per 1M tokens)

Model	Version	Region	Input	Output	Free Quota
Qwen-Turbo	7/15/2025	Chinese Mainland	¥0.30	¥0.60	1M tokens (90 days)
Qwen-Turbo (Thinking)	7/15/2025	Chinese Mainland	¥0.30	CoT+Response ¥3.00	1M tokens (90 days)
Qwen-Flash	7/28/2025	Chinese Mainland (≤128K)	¥0.15	¥1.50	1M tokens (90 days)
Qwen-Flash	7/28/2025	Chinese Mainland (≤256K)	¥0.60	¥6.00	1M tokens (90 days)
Qwen-Flash	7/28/2025	Chinese Mainland (≤1M)	¥1.20	¥12.00	1M tokens (90 days)
Qwen-Plus	12/1/2025	Chinese Mainland (≤128K)	¥0.80	¥2.00 (non-thinking) / ¥8.00 (thinking)	1M tokens (90 days)
Qwen-Plus	12/1/2025	Chinese Mainland (≤256K)	¥2.40	¥20.00 (non-thinking) / ¥24.00 (thinking)	1M tokens (90 days)
Qwen-Plus	12/1/2025	Chinese Mainland (≤1M)	¥4.80	¥48.00 (non-thinking) / ¥64.00 (thinking)	1M tokens (90 days)
Qwen-Long	1/25/2025	Chinese Mainland	¥0.50	¥2.00	✕ No free quota
Qwen3-Max	1/23/2026	Chinese Mainland (≤32K)	¥2.50	¥10.00	1M tokens (90 days)
Qwen3-Max	1/23/2026	Chinese Mainland (≤128K)	¥4.00	¥16.00	1M tokens (90 days)
Qwen3-Max	1/23/2026	Chinese Mainland (≤256K)	¥7.00	¥28.00	1M tokens (90 days)
Qwen3.5-Flash	2/23/2026	Chinese Mainland (≤128K)	¥0.20	¥2.00	1M tokens (90 days)
Qwen3.5-Flash	2/23/2026	Chinese Mainland (≤256K)	¥0.80	¥8.00	1M tokens (90 days)
Qwen3.5-Flash	2/23/2026	Chinese Mainland (≤1M)	¥1.20	¥12.00	1M tokens (90 days)
Qwen3.6-Flash	4/16/2026	Chinese Mainland (≤256K)	¥1.20	¥7.20	1M tokens (90 days)
Qwen3.6-Flash	4/16/2026	Chinese Mainland (≤1M)	¥4.80	¥28.80	1M tokens (90 days)

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company report, Citi Research

Coding pricing plan

Figure 42. Qwen3-Coder-Plus (v2025-09-23) Context: 1M

Token Range	Mainland Input	Mainland Output	Global Input	Global Output
0 – 32K	¥4.00	¥16.00	$0.57	$2.29
32K – 128K	¥6.00	¥24.00	$0.86	$3.43
128K – 256K	¥10.00	¥40.00	$1.43	$5.71
256K – 1M	¥20.00	¥200.00	$2.86	$28.60

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company report, Citi Research

Figure 43. Qwen3-Coder-Flash (v2025-07-28) Context:1M

Token Range	Mainland Input	Mainland Output	Global Input	Global Output
0 – 32K	¥1.00	¥4.00	$0.14	$0.57
32K – 128K	¥1.50	¥6.00	$0.21	$0.86
128K – 256K	¥2.50	¥10.00	$0.36	$1.43
256K – 1M	¥5.00	¥25.00	$0.71	$3.57

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company report, Citi Research

www.futoucaixin.cn

Figure 44. Qwen3-Coder-Next (v2026-02-20)

Token Range	Input (¥/1M)	Output (¥/1M)
0 – 32K	¥1.00	¥4.00
32K – 128K	¥1.50	¥6.00
128K – 256K	¥2.50	¥10.00

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company report Citi Research

Figure 45. Qwen models – Multimodal & Specialized Models

Model	Version	Region	Pricing	Currency	Free Quota
Qwen-VL-OCR	—	Global	Input 0.04/output 0.07 per 1M tokens	USD	1M tokens (90 days)
Qwen-VL-OCR	—	Chinese Mainland	Input ¥0.30 / Output ¥0.50 per 1M tokens	RMB	1M tokens (90 days)
Qwen-Image-2.0-Pro	3/3/2026	International	$0.075 per image	USD	100 images (90 days)
Qwen3-TTS-Flash	11/27/2025	Chinese Mainland	¥0.80 per 10K characters	RMB	10K characters
Qwen3-TTS-VC-Realtime	11/27/2025	Chinese Mainland	¥0.80 per 10K characters	RMB	10K characters
Qwen3-ASR-Flash-Realtime	10/27/2025	Chinese Mainland	¥0.00033 per second	RMB	10 hours
Fun-ASR	11/7/2025	Chinese Mainland	¥0.00022 per second	RMB	10 hours
Qwen3.5-Omni-Plus	3/15/2026	Chinese Mainland	Text In ¥7 / Audio In ¥53 / Image+Video In ¥40 / Text Out ¥213 per 1M tokens	RMB	1M tokens (90 days)
Qwen3.5-Omni-Flash	3/15/2026	Chinese Mainland	Text In ¥2.20 / Audio In ¥18 / Image+Video In ¥13.30 / Text Out ¥72 per 1M tokens	RMB	1M tokens (90 days)
QVQ-Max	—	International	Input 1.26/ Output 5.03 per 1M tokens	USD	1M tokens (90 days)
QVQ-Max	—	Chinese Mainland	Input ¥8.00 / Output ¥32.00 per 1M tokens	RMB	1M tokens (90 days)
QVQ-Plus	—	Chinese Mainland	Input ¥2.00 / Output ¥5.00 per 1M tokens	RMB	1M tokens (90 days)

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company report, Citi Research

Figure 46. Coding Subscription Plan (Pro)

Feature	Detail
Price	$50 / month (USD)
Request Quota	6,000 req / 5 hrs · 45,000 req / week · 90,000 req / month
Recommended Models	Qwen3.5-Plus, Kimi-K2.5, GLM-5, MiniMax-M2.5, Qwen3-Max, Qwen3-Coder-Next, Qwen3-Coder-Plus, GLM-4.7
Lite Plan	✕ Discontinued (no new subscriptions from March 20, 2026; existing users retain access)
Quota Usage	Simple tasks: ~5–10 requests · Complex tasks: ~10–30+ requests

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Company report Citi Research

www.futoucaixin.cn

Figure 47. Knowledge Base (RAG) Resource Packages

RAG Standard Edition

Package	Duration	Included	Price (¥ RMB)
Starter	1 month	1 knowledge base (720 hours)	¥20
Enterprise	1 year	100 knowledge bases (876,000 hours)	¥18,999

RAG Flagship Edition

Package	Duration	Included	Price (¥ RMB)
Starter	1 month	1 RCU (720 RCU·hours)	¥139
Enterprise	1 year	50 RCUs (438,000 RCU·hours)	¥65,999

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company report Citi Research

Recent price adjustment by China cloud vendors

Pricing signal: Alibaba Cloud raised AI compute prices 5–34% effective April 18, 2026, suggesting strong demand and improved pricing power.

Figure 48. China cloud vendors' recent price adjustment

Cloud providers	Announcement date	Effective date	Product category	Products	Applied regions	Price adjustment
AliCloud	18-Mar-26	18-Apr-26	AI compute	T-Head Zhenwu 810Eand related products	All regions	+5-34%
			AI storage	Cloud Parallel File Storage (CPFS)	All regions	+30%
Baidu AI Cloud	18-Mar-26	18-Apr-26	AI compute	AI compute related products and services	na	+5-30%
			AI storage	Cloud Parallel File Storage (CPFS)	na	+30%
MiniMax	18-Mar-26	18-Mar-26	AI large model	MiniMax M2.7	na	Higher effective pricing than M2.5
Volcano Engine (ByteDance)	16-Mar-26	17-Mar-26	AI coding	Ark Coding Plan	na	Discount cancellation for first time users
Knowledge Atlas (Zhipu AI)	15-Mar-26	16-Mar-26	AI large model	GLM-5-Turbo	na	API +20% vs. GLM-5
	12-Feb-26	12-Feb-26	AI large model	GLM-5	na	Coding Plan and API +50% vs. GLM-4.7
Tencent Cloud	9-Apr-26	9-May-26	AI compute	CVM-GPU, HCC-GPU, HCC-THPC	All regions	+5%
			Container service	TKEnative node	All regions	+5%
			Elasticity	Elastic MapReduce	All regions	+5%
	11-Mar-26	13-Mar-26	AI large model	GLM 5, MiniMax 2.5, Kimi 2.5	na	Official commercialization from free beta earlier
				HY2.0 Instruct	na	Import: Rmb0.0008 to Rmb 0.004505 per K tokens Export: Rmb0.002 to Rmb 0.01113 per K tokens
				HY2.0 Think	na	Import: Rmb0.001 to Rmb 0.0053 per K tokens Export: Rmb0.004 to Rmb 0.0212 per K tokens

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports, Citi Research

Cloud industry market size and market share

Figure 49. Global public cloud market size

Segments	2024 Spending (US$ bn)	2025E Spending (US$ bn)	2026E Spending (US$ bn)
IaaS (Infrastructure as a Service)	182	212	272
PaaS (Platform as a Service)	176	209	259
SaaS (Software as a Service)	244	299	356
BPaaS (Business Process as a Service)	73	78	86
Cloud Management & Security	na	44	52
DaaS (Desktop as a Service)	3	4	4
Total Public Cloud Market	679	845	1,029

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

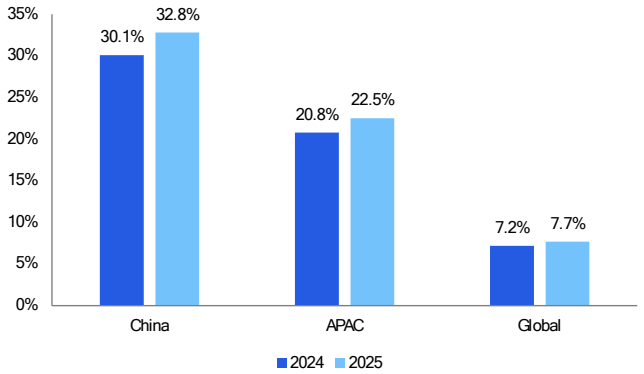
Source: Gartner, Citi Research Estimates

www.futoucaixin.cn

Alibaba maintains its number 1 position in China IaaS public cloud market with 33% market share in 2025, per Gartner. Against the broader APAC (#1) and global (#4) context, BABA market share was at 22.5% and 7.7% in 2025 respectively.

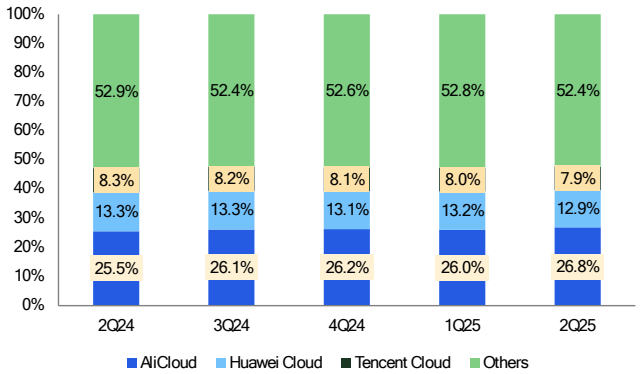
According to IDC, Alibaba Cloud has 27% China public cloud IaaS market share in 2Q25, followed by Huawei at 12.9% while Tencent captured 7.9% share.

Figure 50. AliCloud IaaS public cloud market share



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Gartner (Apr 10, 2026), Citi Research

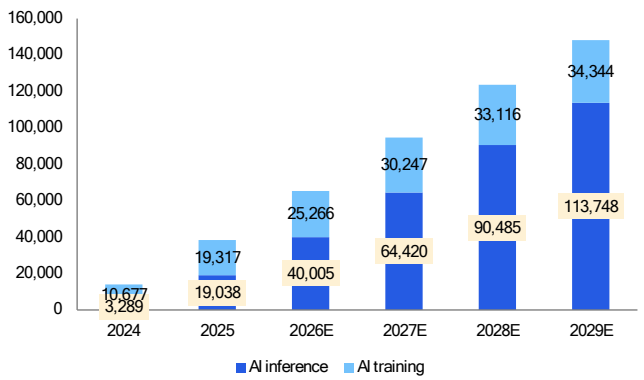
Figure 51. China public cloud IaaS market share by quarter



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: IDC (Oct 31, 2025), Citi Research

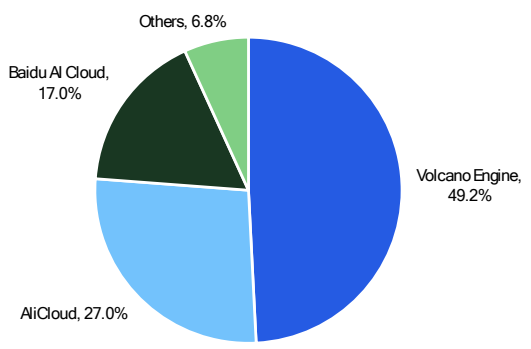
In terms of public cloud LLM service market share, IDC estimate that Alibaba Cloud has 27% share as of 1H25, # 2 behind Bytedance's Volcano Engine at 49.2% while Baidu AI cloud captured 17% market share.

Figure 52. China GenAI IaaS market size forecast (Rmb mn)



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: IDC (Oct 20, 2025), Citi Research

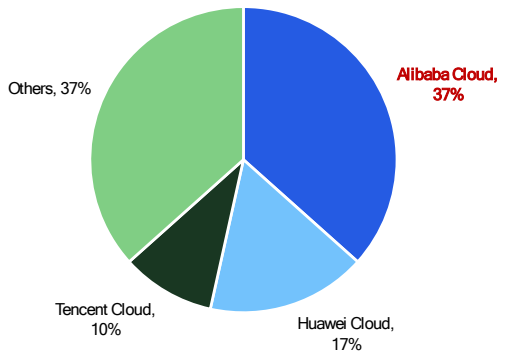
Figure 53. China public cloud LLM service market share in 1H25



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: IDC (Sep 17, 2025), Citi Research

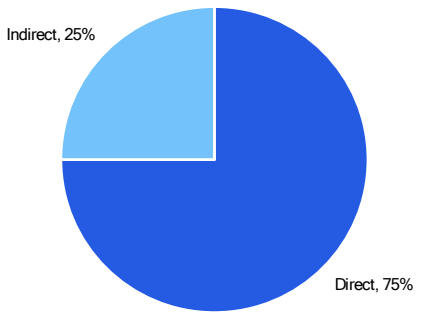
www.futoucaixin.cn

Figure 54. China cloud infrastructure spending market share in 4Q25



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Omdia Cloud Channels Analysis, Citi Research

Figure 55. China cloud infrastructure spending by channel in 4Q25



© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.
Source: Omdia Cloud Channels Analysis, Citi Research

Token usage forecast by IDC

As reported by Sina (dated May 7, 2026) citing latest IDC report that where IDC noted that China's enterprise-level MaaS market share (measured by call volume) reached 114 trillion in 2024 and grew by 16x to reach 1,944 trillion tokens by 2025 which drove the China's public cloud MaaS market reached Rmb3.07bn in 2025. IDC estimates that total token consumption for 2026 will be ~40,000 trillion, suggesting an increase of 20x from 2025 level, driven by accelerated growth and large-scale implementation of Agent-like applications, essentially expanding from "text generation" to "multimodal understanding and automated execution" which led to significant increase in the number of tokens consumed per interaction.

IDC noted that currently main users of public cloud MaaS are concentrated in broader internet industry (gaming, entertainment, education), smart office sector, smart hardware sector (smart cars, mobile phones, smart glasses), and the mass consumer sector. Common application scenarios include role-playing, short drama generation, marketing, search, data processing, data analysis, and document processing. In private deployment projects for traditional government and enterprise clients, the application scenarios are more focused, mainly concentrating on smart office, data processing and analysis, and marketing.

Players in MaaS market

IDC noted that when measured by call volume in public cloud MaaS market, Bytedance's Volcano Engine captured nearly half of the market share in 2025, followed by Alibaba Cloud, Baidu AI Cloud and Silicon-based Mobility. And when measured by revs, Volcano Engine holds >40% market share with Alibaba Cloud, Baidu AI Cloud and Zhipu Ai rounding out the top four.

Our forecast of China public cloud MaaS size

By forecasting token consumption in China's MaaS market to reach 40,000 trillion in 2026, IDC estimate revs scale will reach Rmb18.6bn. IDC forecast that token consumption will grow at CAGR of 723% from 2024-2030E to reach 35.5 quintillion with revs size to grow at CAGR 1156% from 2024-2030E.

www.futoucaixin.cn

Based on the "actual" 2025 estimated token and revs quoted by IDC, we perform some sensitivity analysis on China MaaS market size with our base, bull and bear assumption for token assumption and token price.

- **Base case:** Our base case forecast that token consumption growth CAGR of 472% from 2024-2030 while unit price will only grow at 2% from 2027-2030 after dropping 71% yoy in 2026 (as estimated by IDC). Our MaaS market size is forecast to grow at 372% CAGR to reach Rmb2 trillion (or US$286bn) in 2030.

Figure 56. Sensitivity analysis on China public cloud MaaS market size

Base case (2030E market size at Rmb2trn)	2024	2025	2026	2027	2028	2029	2030	2024-30 CAGR
China public cloud MaaS token consumption (quintillion)	0.000114	0.001941	0.040	0.260	0.780	1.950	3.978	*472%*
yoy growth %		*1603%*	*1961%*	*550%*	*200%*	*150%*	*104%*	
Unit token price per 1M tokens (Rmb)	1.582	1.582	0.465	0.474	0.484	0.493	0.503	*-17%*
yoy growth %			*-71%*	*2%*	*2%*	*2%*	*2%*	
China public cloud MaaS market size (Rmb bn)	0.18	3.07	18.60	123	377	962	2,002	*372%*
yoy growth %		*1603%*	*506%*	*563%*	*206%*	*155%*	*108%*	
Bull case (more robust MaaS adoption)	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**	**2024-30 CAGR**
China public cloud MaaS token consumption (quintillion)	0.000114	0.001941	0.048	0.317	0.982	2.553	5.490	*503%*
yoy growth %		*1603%*	*2373%*	*560%*	*210%*	*160%*	*115%*	
Unit token price per 1M tokens (Rmb)	1.582	1.582	0.484	0.509	0.534	0.561	0.589	*-15%*
yoy growth %			*-69%*	*5%*	*5%*	*5%*	*5%*	
China public cloud MaaS market size (Rmb bn)	0.18	3.07	23.25	161	524	1,432	3,232	*412%*
yoy growth %		*1603%*	*657%*	*593%*	*226%*	*173%*	*126%*	
Bear case (potential price war)	**2024**	**2025**	**2026**	**2027**	**2028**	**2029**	**2030**	**2024-30 CAGR**
China public cloud MaaS token consumption (quintillion)	0.000114	0.001941	0.036	0.230	0.668	1.604	3.127	*449%*
yoy growth %		*1603%*	*1755%*	*540%*	*190%*	*140%*	*95%*	
Unit token price per 1M tokens (Rmb)	1.582	1.582	0.439	0.430	0.422	0.413	0.405	*-20%*
yoy growth %			*-72%*	*-2%*	*-2%*	*-2%*	*-2%*	
China public cloud MaaS market size (Rmb bn)	0.18	3.07	15.81	99	282	663	1,267	*338%*
yoy growth %		*1603%*	*415%*					

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: IDC (May 7, 2026), Citi Research

- **Bull case –** Our bull case assumption forecast token consumption usage to grow at 503% CAGR with token prices increasing 5% per year from 2027-2030 after declining by 69% in 2026. Our MaaS market size is estimated to grow at 412% CAGR to reach Rmb3.23 trillion by 2030.

- **Bear case –** Our bear case assumption forecast token consumption usage to grow slower at 449% CAGR and token price to further decline by 2% each year from 2027-2030 after dropping 72% in 2026 and our market size is forecast to reach Rmb1.27 trillion or 338% CAGR.

Rationale on assumption

We tend to agree that with the accelerated demand and proliferation of agentic workflow, token consumption growth will continue to grow at robust speed in China. Nevertheless, given the intense competitive environment in China, we have a more conservative view on token prices. Once supply constraints ease off and if Chinese players no longer face capacity constraints while the demand slowing down, we are cautious if it could trigger massive pricing cut on token price in the outer years to fight for market share gain and industry consolidation could also materialize.

www.futoucaixin.cn

Figure 57. Earnings revisions – detail

	FY2026E			FY2027E			FY2028E		
RMB mn	Old	New	% Chg	Old	New	% Chg	Old	New	% Chg
Alibaba China E-commerce Group	**553,434**	**553,434**	**0.0%**	**572,544**	**572,544**	***0.0%***	**594,112**	**594,112**	***0.0%***
E-commerce	**449,637**	449,637	**0.0%**	**456,612**	456,612	***0.0%***	**471,783**	471,783	***0.0%***
-CMR	**342,858**	**342,858**	***0.0%***	**348,232**	**348,232**	***0.0%***	**361,777**	**361,777**	***0.0%***
yoy growth	*6.4%*	*6.4%*	**0.0%**	*1.6%*	*1.6%*	***0.0%***	*3.9%*	*3.9%*	***0.0%***
Alibaba International Digital Commerce Group	**143,380**	**143,380**	**0.0%**	**151,845**	**151,845**	**0.0%**	**161,530**	**161,530**	**0.0%**
yoy growth rate %	*8.4%*	*8.4%*	*0.0%*	*5.9%*	*5.9%*	*0.0%*	*6.4%*	*6.4%*	*0.0%*
Cloud Intelligence Group	**158,985**	**158,985**	**0.0%**	**227,110**	**227,110**	**0.0%**	**317,090**	**319,531**	**0.8%**
yoy growth rate %	*34.7%*	*34.7%*	*0.0%*	*42.9%*	*42.9%*	*0.0%*	*39.6%*	*40.7%*	*1.1%*
All others	**252,970**	**252,970**	**0.0%**	**252,897**	**252,897**	**0.0%**	**253,861**	**253,861**	**0.0%**
Consolidated revenue	**1,022,651**	**1,022,651**	**0.0%**	**1,117,143**	**1,117,143**	**0.0%**	**1,247,265**	**1,249,706**	**0.2%**
yoy growth rate %	*2.6%*	*2.6%*	*0.0%*	*9.2%*	*9.2%*	*0.0%*	*11.6%*	*11.9%*	*0.2%*
Non-GAAP Operating Income	**75,434**	**75,434**	**0.0%**	**114,461**	**114,461**	**0.0%**	**171,624**	**171,986**	**0.2%**
Non-GAAP OPM	*7.4%*	*7.4%*	*0.0%*	*10.2%*	*10.2%*	*0.0%*	*13.8%*	*13.8%*	*0.0%*
Non-GAAP Adjusted Net Income	**73,695**	**73,695**	**0.0%**	**114,981**	**114,981**	**0.0%**	**158,200**	**158,489**	**0.2%**
Non-GAAP NM	*7.2%*	*7.2%*	*0.0%*	*10.3%*	*10.3%*	*0.0%*	*12.7%*	*12.7%*	*0.0%*
yoy growth rate %	*-53.3%*	*-53.3%*	*0.0%*	*56.0%*	*56.0%*	*0.0%*	*37.6%*	*37.8%*	*0.3%*
Adjusted EBITA	**75,434**	**75,434**	**0.0%**	**114,461**	**114,461**	**0.0%**	**171,624**	**171,986**	**0.2%**
Adjusted EBITA Margin	*7.4%*	*7.4%*	*0.0%*	*10.2%*	*10.2%*	*0.0%*	*13.8%*	*13.8%*	*0.0%*
yoy growth rate %	*-56.4%*	*-56.4%*	*0.0%*	*51.7%*	*51.7%*	*0.0%*	*0.0%*	*0.0%*	*0.0%*
Alibaba China E-commerce Group EBITA	**106,572**	**106,572**	**0.0%**	**150,229**	**150,229**	**0.0%**	**172,291**	**172,291**	**0.0%**
Adjusted Taobao and Tmall EBITA Margin	*19.3%*	*19.3%*	*NA*	*26.2%*	*26.2%*	*NA*	*29.0%*	*29.0%*	*NA*
Cloud Intelligence Group EBITA	**14,292**	**14,292**	**0.0%**	**20,440**	**20,440**	**0.0%**	**28,855**	**29,077**	**0.8%**
Adjusted Cloud EBITA Margin	*9.0%*	*9.0%*	*0.0%*	*9.0%*	*9.0%*	*0.0%*	*9.1%*	*9.1%*	*0.0%*
All Others EBITA	**-34,757**	**-34,757**	**NM**	**-49,646**	**-49,646**	**NM**	**-23,993**	**-23,993**	**NM**

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Citi Research

www.futoucaixin.cn

Figure 58. Income statement of Alibaba

	1Q26	2Q26	3Q26	4Q26E	FY2026E	FY2027E	FY2028E
Alibaba China E-commerce Group	**140,072**	**132,578**	**159,347**	**121,437**	**553,434**	**572,544**	**594,112**
E-commerce	118,577	102,933	131,583	96,544	449,637	456,612	471,783
-CMR	89,252	78,927	102,664	72,015	342,858	348,232	361,777
-Direct sales, logistics and others	29,325	24,006	28,919	24,529	106,779	108,380	110,006
Quick commerce	14,784	22,906	20,842	18,758	77,290	87,976	93,255
China commerce wholesale	6,711	6,739	6,922	6,135	26,507	27,956	29,074
Alibaba International Digital Commerce Group	**34,741**	**34,799**	**39,201**	**34,639**	**143,380**	**151,845**	**161,530**
-commerce retail	28,395	28,068	32,351	28,155	116,969	123,136	130,524
-commerce wholesale	6,346	6,731	6,850	6,484	26,411	28,709	31,006
Cloud Intelligence Group	**33,398**	**39,824**	**43,284**	**42,479**	**158,985**	**227,110**	**319,531**
All others	58,599	62,969	67,340	64,062	252,970	252,897	253,861
Unallocated	519	577	603	600	2,299	2,000	2,000
Inter-segment elemination	(19,677)	(22,952)	(24,932)	(20,855)	(88,416)	(89,254)	(81,328)
Consolidated revenue	**247,652**	**247,795**	**284,843**	**242,361**	**1,022,651**	**1,117,143**	**1,249,706**
yoy%	*1.8%*	*4.8%*	*1.7%*	*2.5%*	*2.6%*	*9.2%*	*11.9%*
Cost of revenue	136,429	150,781	169,534	147,598	604,342	640,953	707,473
Gross profit	111,223	97,014	115,309	94,763	418,309	476,190	542,233
Gross margin %	*44.9%*	*39.2%*	*40.5%*	*39.1%*	*40.9%*	*42.6%*	*43.4%*
Product development expenses	15,001	17,095	15,480	16,965	64,541	81,831	89,599
Sales and marketing expenses	53,178	66,496	71,934	69,073	260,681	261,339	259,163
General and administrative expenses	7,398	7,380	8,355	7,756	30,889	32,770	36,133
Amortization of intangible assets	807	826	841	841	3,315	3,364	3,364
Impairment of goodwill and intangible assets	0	0	8,054	0	8,054	0	0
Total costs and expenses	**76,235**	**91,649**	**104,664**	**94,635**	**367,183**	**379,304**	**388,259**
Operating income (GAAP)	**34,988**	**5,365**	**10,645**	**128**	**51,126**	**96,886**	**153,974**
GAAP Op margin	*14.1%*	*2.2%*	*3.7%*	*0.1%*	*5.0%*	*8.7%*	*12.3%*
Net income (loss) - GAAP	**42,382**	**20,612**	**15,631**	**6,059**	**84,684**	**93,742**	**136,413**
Net income (loss) attributable to noncontrolling interests	(1,733)	407	752	852	278	3,908	4,308
Net income (loss) attributable to Alibaba Group Holding Limited	**40,649**	**21,019**	**16,383**	**6,911**	**84,962**	**97,650**	**140,721**
Non-GAAP adjusted net income	**33,510**	**10,352**	**21,569**	**10,051**	**75,482**	**111,317**	**154,425**
Non-GAAP Net margin	13.5%	4.2%	7.6%	4.1%	7.4%	10.0%	12.4%
Non-GAAP adjusted net income attributable to shareholders	**35,291**	**10,450**	**17,112**	**10,842**	**73,695**	**114,981**	**158,489**
Non-GAAP Net margin (shareholder level)	14.3%	4.2%	6.0%	4.5%	7.2%	10.3%	12.7%
Non-GAAP EPADS—Diluted	**14.75**	**4.36**	**7.09**	**4.51**	**30.69**	**48.24**	**67.29**
Consolidated adjusted EBITA	**38,844**	**9,073**	**23,397**	**4,120**	**75,434**	**114,461**	**171,986**
Adj. EBITA margin	16%	4%	8%	2%	7%	10%	14%
Alibaba China E-commerce Group	**38,389**	**10,497**	**34,613**	**23,073**	**106,572**	**150,229**	**172,291**
Adj. EBITA margin	27%	8%	22%	19%	19%	26%	29%
Cloud Intelligence Group	**2,954**	**3,604**	**3,911**	**3,823**	**14,292**	**20,440**	**29,077**
Adj. EBITA margin	8.8%	9.0%	9.0%	9.0%	9.0%	9.0%	9.1%

© 2026 Citigroup Inc. No redistribution without Citigroup's written permission.

Source: Company Reports and Citi Research Estimates

www.futoucaixin.cn

Companies Mentioned:
Alibaba Group Holding (BABA.N; US$140.06; 1; 08 May 26; 16:00)
Alibaba (9988.HK; HK$139.0; 1; 08 May 26; 16:10)
Advanced Micro Devices (AMD.O; US$455.19; 2; 08 May 26; 16:00)
Airbnb Inc (ABNB.O; US$141.49; 1; 08 May 26; 16:00)
Amazon.com, Inc. (AMZN.O; US$272.68; 1; 08 May 26; 16:00)
Arm Holdings PLC (ARM.O; US$213.27; 1; 08 May 26; 16:00)
Baidu.com (BIDU.O; US$141.05; 1; 08 May 26; 16:00)
Broadcom Inc (AVGO.O; US$430.0; 1; 08 May 26; 16:00)
FinVolution (FINV.N; US$5.17; 1H; 08 May 26; 16:00)
GoTo Gojek Tokopedia (GOTO.JK; Rp50.0; 1H; 08 May 26; 16:00)
Hello Group Inc (MOMO.O; US$6.23; 2H; 08 May 26; 16:00)
Iflytek (002230.SZ; Rmb49.91; 2; 08 May 26; 15:00)
Intel Corp (INTC.O; US$124.92; 1; 08 May 26; 16:00)
Kuaishou (1024.HK; HK$52.95; 1; 08 May 26; 16:10)
Marvell Technology Inc (MRVL.O; US$170.13; 1; 08 May 26; 16:00)
Meituan (3690.HK; HK$84.05; 1H; 08 May 26; 16:10)
Meta Platforms Inc (META.O; US$609.63; 1; 08 May 26; 16:00)
Microsoft Corp. (MSFT.O; US$415.12; 1; 08 May 26; 16:00)
MiniMax (0100.HK; HK$742.5; 1H; 08 May 26; 16:10)
Nokia OYJ (NOKIA.HE; €10.95; 3; 08 May 26; 18:30)
NVIDIA Corp (NVDA.O; US$215.2; 1; 08 May 26; 16:00)
NXP Semiconductors NV (NXPI.O; US$294.75; 1; 08 May 26; 16:00)
Salesforce Inc (CRM.N; US$181.82; 2; 08 May 26; 16:00)
Samsung Electronics (005930.KS; W268500.0; 1; 08 May 26; 15:45)
Tencent Holdings (0700.HK; HK$471.4; 1; 08 May 26; 16:10)
Vodafone Group PLC (VOD.L; £1.19; 2; 08 May 26; 16:30)
Xiaomi (1810.HK; HK$31.68; 1; 08 May 26; 16:10)
XPeng (XPEV.N; US$15.62; 1; 08 May 26; 16:00)
XPeng (9868.HK; HK$61.3; 1; 08 May 26; 16:10)
Yuanbao (YB.O; US$15.64; 2; 08 May 26; 16:00)

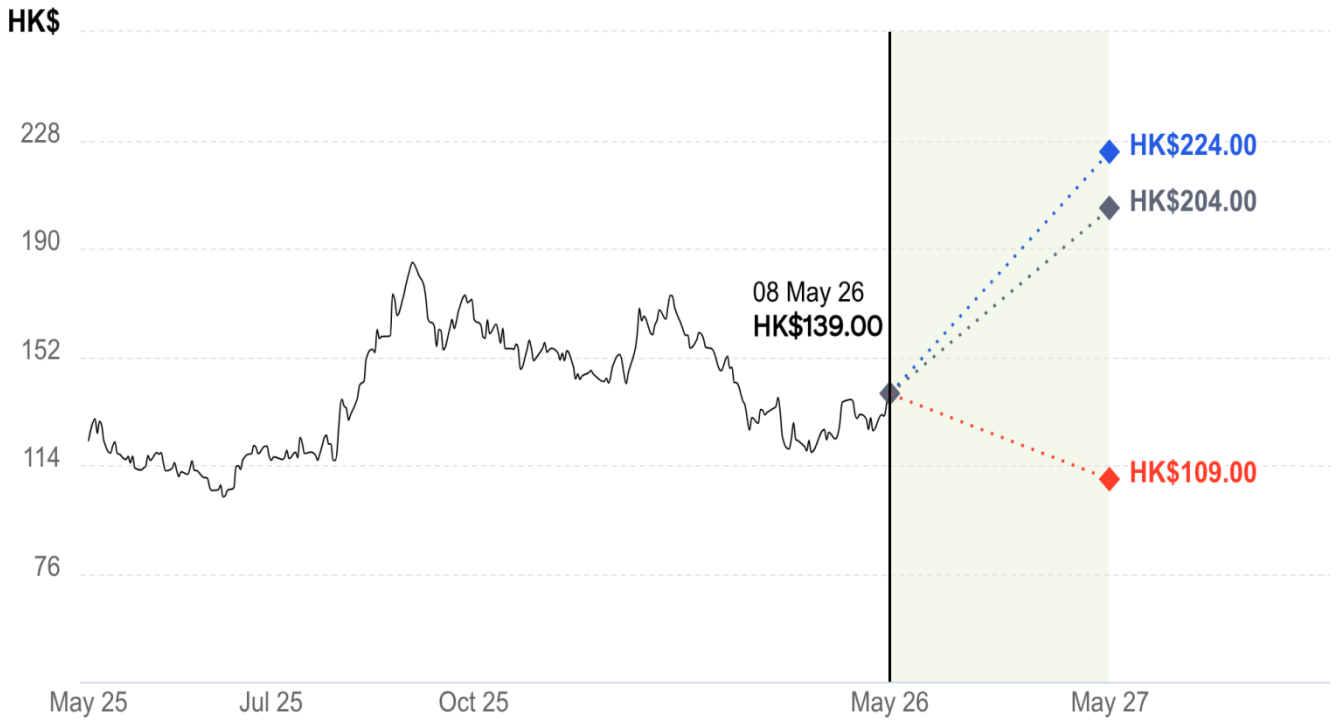
www.futoucaixin.cn

Bull/Bear: Alibaba (9988.HK)

HK$224.00 ▲61% Upside

HK$204.00 ▲47% Upside

HK$109.00 ▼22% Downside



Spread 83pp
Current Price and expected returns (upside/downside) as of 08 May 2026

BULL Assumptions

- 12x P/E on FY2027E Ecommerce Group net profit, 7x P/S multiple on FY2027E Cloud Intelligence Group revenue

BASE Assumptions

- 10x P/E on FY2027E Ecommerce Group net profit, 6.5x P/S multiple on FY2027E Cloud Intelligence Group revenue

BEAR Assumptions

- 5x P/E on FY2027E Ecommerce net profit, 2x P/S multiple on FY2027E Cloud Intelligence Group revenue

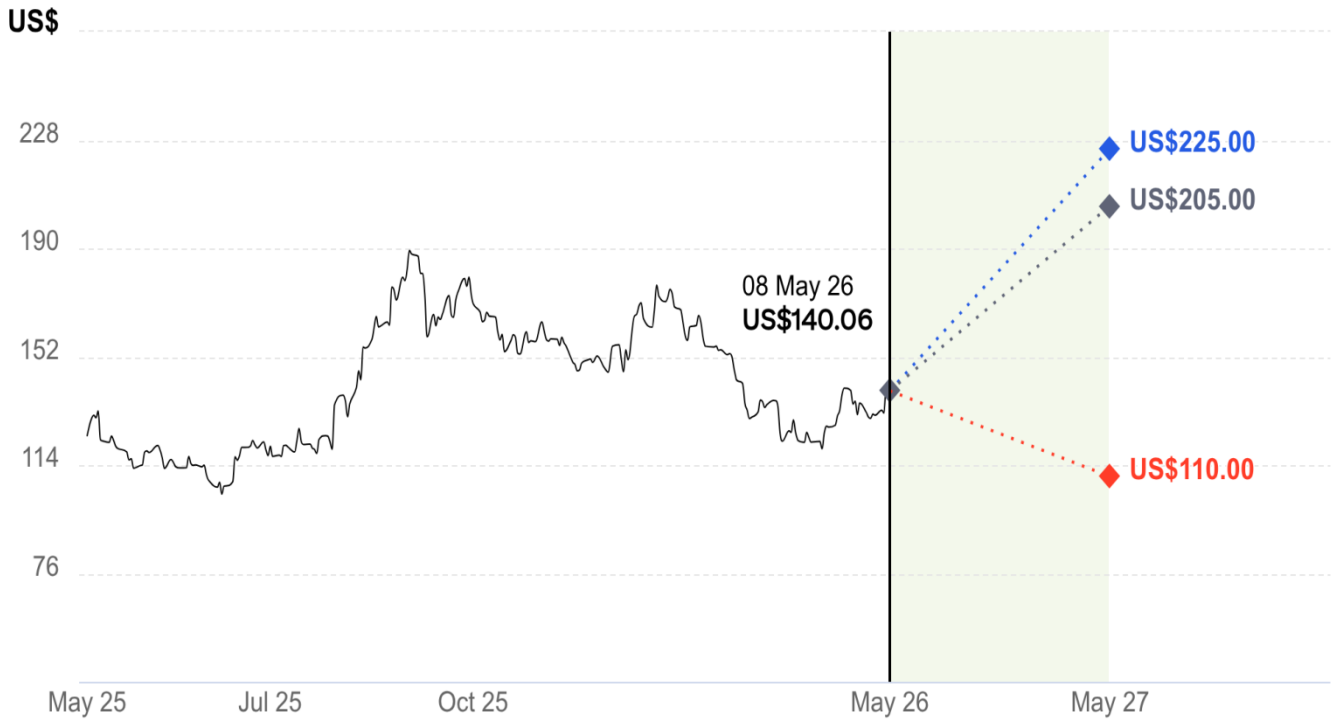
www.futoucaixin.cn

Bull/Bear: Alibaba Group Holding (BABA.N)

US$ 225.00 ▲61% Upside

US$ 205.00 ▲46% Upside

US$ 110.00 ▼21% Downside



Spread 82pp
Current Price and expected returns (upside/downside) as of 08 May 2026

BULL Assumptions

- 12x P/E on FY2027E Ecommerce Group net profit, 7x P/S on FY2027E Cloud Intelligence Group revenue

BASE Assumptions

- 10x P/E on FY2027E Ecommerce Group net profit, 6.5x P/S on FY2027E Cloud Intelligence Group revenue

BEAR Assumptions

- 5x P/E on FY2027E Ecommerce Group net profit, 2x P/S on FY2027E Cloud Intelligence Group revenue

www.futoucaixin.cn

Alibaba

Company description

Alibaba was founded in 1999. In the years since, Alibaba has created one of the world's leading internet e-commerce companies with an online ecosystem for buyers, sellers, third-party service providers, third-party payments, and strategic partners. The company was listed on the NYSE in September 2014 and completed its secondary listing in Hong Kong in November 2019.

Investment strategy

We rate Alibaba H-shares as Buy. We are positive on Alibaba given: i) it has the leading online commerce ecosystem in China (combined 80% market share in C2C and B2C segments); ii) its monetization rate has upside potential; iii) synergies can be harvested across marketplaces, leveraging big data analysis, and AliCloud infrastructure; and iv) growth opportunities beckon in China's rural cities and globally.

Valuation

Our target price for Alibaba H-shares of HK$204 is based on 10x P/E on FY2027E Ecommerce Group net profit, 6.5x P/S on FY2027E Cloud Intelligence Group revenue, 2.0x P/S on AIDC Group revenue, and 1.0x P/S on all other businesses. Our target multiples are benchmarked against comps. We also assign a 30% discount to net cash position, in-line with our SOTP for other internet peers. Our SOTP fair value comes to HK$204 per H-share.

Risks

Key downside risks that could prevent the shares from reaching our target price include: (i) failure in executing its new retail strategy; ii) investment spend and margins pressure become worse than expected; iii) slowdown of user traffic and online GMV and losing appeal to brands & merchants; iv) integration risks for newly acquired entities; v) slowdown of the Chinese and global economies and overhang of US-China or other trade disputes; and vi) regulatory risks on poor product quality and integrity of merchants.

Alibaba Group Holding

Company description

Alibaba was founded in 1999. In the years since, Alibaba has created one of the world's leading internet e-commerce companies with an online ecosystem for buyers, sellers, third-party service providers, third-party payments, and strategic partners. The company was listed on the NYSE in September 2014 and completed its secondary listing in Hong Kong in November 2019.

Investment strategy

We rate Alibaba N-shares as Buy. We are positive on Alibaba given: i) it has the leading online commerce ecosystem in China (combined 80% market share in C2C and B2C segments); ii) its monetization rate has upside potential; iii)

www.futouqaixin.cn

synergies can be harvested across marketplaces, leveraging big data analysis, and AliCloud infrastructure; and iv) growth opportunities beckon in China's rural cities and globally.

Valuation

Our target price for Alibaba N-shares of US$205 is based on 10x P/E on FY2027E Ecommerce group net profit, 6.5x P/S on FY2027E Cloud Intelligence Group revenue, 2.0x P/S on AIDC Group revenue, and 1.0x P/S on all other businesses. Our target multiples are benchmarked against comps. We also assign a 30% discount to net cash position, in-line with our SOTP for other internet peers. Our SOTP fair value comes to US$205 per share.

Risks

Key downside risks that could prevent the shares from reaching our target price include: (i) failure in executing its new retail strategy; ii) investment spend and margins pressure become worse than expected; iii) slowdown of user traffic and online GMV and losing appeal to brands & merchants; iv) integration risks for newly acquired entities; v) slowdown of the Chinese and global economies and overhang of US-China or other trade disputes; and vi) regulatory risks on poor product quality and integrity of merchants.

www.futoucaixin.cn

If you are visually impaired and would like to speak to a Citi representative regarding the details of the graphics in this document, please call USA 1-888-500-5008 (TTY: 711), from outside the US +1-210-677-3788

Appendix A-1

ANALYST CERTIFICATION

The research analysts primarily responsible for the preparation and content of this research report are either (i) designated by "AC" in the author block or (ii) listed in bold alongside content which is attributable to that analyst. If multiple AC analysts are designated in the author block, each analyst is certifying with respect to the entire research report other than (a) content attributable to another AC certifying analyst listed in bold alongside the content and (b) views expressed solely with respect to a specific issuer which are attributable to another AC certifying analyst identified in the price charts or rating history tables for that issuer shown below. Each of these analysts certify, with respect to the sections of the report for which they are responsible: (1) that the views expressed therein accurately reflect their personal views about each issuer and security referenced and were prepared in an independent manner, including with respect to Citigroup Global Markets Inc. and its affiliates; and (2) no part of the research analyst's compensation was, is, or will be, directly or indirectly, related to the specific recommendations or views expressed by that research analyst in this report.

IMPORTANT DISCLOSURES

Alibaba Group Holding (BABA)

Ratings and Target Price History
Fundamental Research

Analyst: Alicia Yap, CFA



	Date	Rating	Target Price	Closing Price
1	18-May-23 19:50:00	1	*144.00	84.37
2	09-Jul-23 17:07:52	1	*149.00	89.08
3	10-Aug-23 20:47:07	1	*151.00	97.59
4	08-Oct-23 17:09:43	1	*147.00	84.66
5	16-Nov-23 18:13:37	1	*142.00	77.82
6	09-Jan-24 13:40:44	1	*125.00	70.85
7	07-Feb-24 18:37:09	1	*126.00	72.44
8	09-Apr-24 18:25:26	1	*124.00	71.80

	Date	Rating	Target Price	Closing Price
9	15-May-24 04:34:45	1	*122.00	79.67
10	15-Aug-24 16:50:03	1	*116.00	78.91
11	29-Sep-24 22:45:36	1	*136.00	106.48
12	09-Oct-24 09:54:52	1	*135.00	107.04
13	17-Nov-24 16:15:27	1	*133.00	87.89
14	09-Jan-25 17:58:55	1	*138.00	83.03
15	20-Feb-25 16:33:13	1	*170.00	134.90
16	08-Apr-25 05:01:59	1	*169.00	98.59

	Date	Rating	Target Price	Closing Price
17	10-Jul-25 07:57:24	1	*148.00	106.64
18	31-Aug-25 17:30:14	1	*187.00	135.00
19	24-Sep-25 15:11:51	1	*217.00	176.44
20	08-Oct-25 07:59:47	1	*218.00	181.12
21	25-Nov-25 16:54:34	1	*225.00	157.01
22	08-Jan-26 09:37:26	1	*197.00	154.47
23	19-Mar-26 16:25:38	1	*200.00	124.90
24	08-Apr-26 06:24:46	1	*205.00	125.32

*Indicates Change

Rating/target price changes above reflect Eastern Time

www.futoucaixin.cn

Alibaba (9988.HK)

Ratings and Target Price History
Fundamental Research

Analyst: Alicia Yap, CFA



	Date	Rating	Target Price	Closing Price
1	18-May-23 19:50:00	1	*140.00	86.97
2	09-Jul-23 17:07:52	1	*145.00	83.55
3	10-Aug-23 20:47:07	1	*148.00	93.46
4	08-Oct-23 17:09:43	1	*144.00	82.16
5	16-Nov-23 18:13:37	1	*140.00	80.62
6	09-Jan-24 13:40:44	1	*122.00	69.13
7	07-Feb-24 18:37:09	1	*124.00	74.23
8	09-Apr-24 18:25:26	1	*122.00	69.87

	Date	Rating	Target Price	Closing Price
9	15-May-24 04:34:45	1	*120.00	81.91
10	15-Aug-24 16:50:03	1	*115.00	75.80
11	29-Sep-24 22:45:36	1	*135.00	101.70
12	09-Oct-24 09:54:52	1	*133.00	102.09
13	17-Nov-24 16:15:27	1	*132.00	86.52
14	09-Jan-25 17:58:55	1	*137.00	79.97
15	20-Feb-25 16:33:13	1	*166.00	119.95
16	08-Apr-25 05:01:59	1	*165.00	101.70

	Date	Rating	Target Price	Closing Price
17	10-Jul-25 07:57:24	1	*144.00	103.20
18	31-Aug-25 17:30:14	1	*183.00	115.70
19	24-Sep-25 15:11:51	1	*215.00	174.00
20	08-Oct-25 07:59:47	1	*216.00	177.60
21	25-Nov-25 16:54:34	1	*223.00	157.80
22	08-Jan-26 09:37:26	1	*195.00	142.60
23	19-Mar-26 16:25:38	1	*199.00	132.00
24	08-Apr-26 06:24:46	1	*204.00	126.50

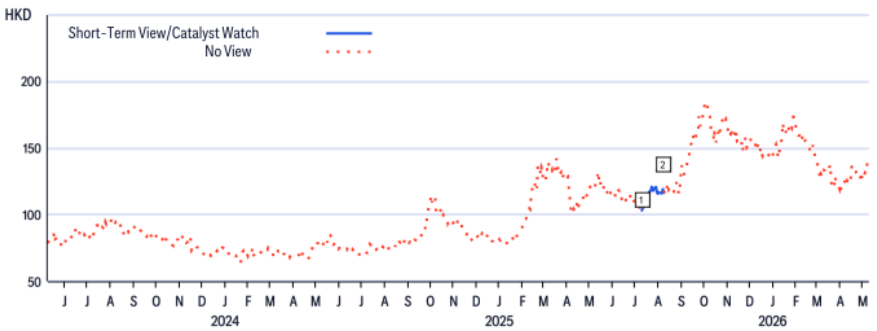
*Indicates Change

Rating/target price changes above reflect Eastern Time

Alibaba (9988.HK)

Short-Term View/Catalyst Watch Research

Analyst: Alicia Yap, CFA



	Date	Action	Expected Direction	Duration	Closing Price
1	10-Jul-25 03:57:24	Add STV	Downside	30 Days	103.20

	Date	Action	Expected Direction	Duration	Closing Price
2	08-Aug-25 14:18:01	Remove STV	Downside	30 Days	116.30

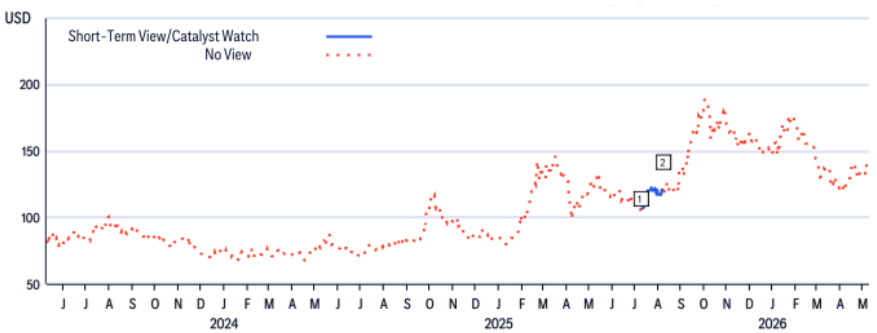
CW - Catalyst Watch , STV - Short-Term View

Rating/target price changes above reflect Eastern Time

Alibaba Group Holding (BABA)

Short-Term View/Catalyst Watch Research

Analyst: Alicia Yap, CFA



	Date	Action	Expected Direction	Duration	Closing Price
1	10-Jul-25 03:57:24	Add STV	Downside	30 Days	106.64

	Date	Action	Expected Direction	Duration	Closing Price
2	08-Aug-25 14:17:52	Remove STV	Downside	30 Days	120.36

CW - Catalyst Watch , STV - Short-Term View

Rating/target price changes above reflect Eastern Time

The Firm has made a market in the publicly traded equity securities of Meituan on at least one occasion since 1 Jan 2025.

The Firm has made a market in the publicly traded equity securities of MiniMax Group Inc on at least one occasion since 1 Jan 2025.

The Firm has made a market in the publicly traded equity securities of Xiaomi Corp on at least one occasion since 1 Jan 2025.

The Firm has made a market in the publicly traded equity securities of Kuaishou Technology on at least one occasion since 1 Jan 2025.

www.futoucaixin.cn

An employee of Citigroup Global Markets Inc or its affiliates is a Director of Microsoft Corp

The Firm has made a market in the publicly traded equity securities of Alibaba Group Holding Ltd on at least one occasion since 1 Jan 2025.

The Firm has made a market in the publicly traded equity securities of Baidu Inc on at least one occasion since 1 Jan 2025.

The Firm has made a market in the publicly traded equity securities of Tencent Holdings Ltd on at least one occasion since 1 Jan 2025.

Citigroup Global Markets Inc. or its affiliates beneficially owns 1% or more of any class of common equity securities of Salesforce Inc. This position reflects information available as of the prior business day.

Citigroup Global Markets Inc. or its affiliates has a net long position of 0.5% or more of any class of common equity securities of Marvell Technology Inc,NXP Semiconductors NV,Salesforce Inc.

Citigroup Global Markets Inc. or its affiliates beneficially owns 2.0% or more of any class of common equity securities of Salesforce Inc.

Within the past 12 months, Citigroup Global Markets Inc. or its affiliates has acted as manager or co-manager of an offering of securities of Airbnb Inc,Alibaba,Amazon.com, Inc.,Broadcom Inc,Intel Corp,Marvell Technology Inc,Meta Platforms Inc,Salesforce Inc,Samsung Electronics.

Citigroup Global Markets Inc. or its affiliates has received compensation for investment banking services provided within the past 12 months from Advanced Micro Devices,Airbnb Inc,Alibaba,Amazon.com, Inc.,Arm Holdings PLC,Broadcom Inc,Intel Corp,Kuaishou,Marvell Technology Inc,Meta Platforms Inc,Microsoft Corp.,NVIDIA Corp,NXP Semiconductors NV,Nokia OYJ,Salesforce Inc,Samsung Electronics,Tencent Holdings,Vodafone Group PLC,Xiaomi,Yuanbao.

Citigroup Global Markets Inc. or its affiliates expects to receive or intends to seek, within the next three months, compensation for investment banking services from Alibaba,Arm Holdings PLC,Intel Corp,Vodafone Group PLC.

Citigroup Global Markets Inc. or its affiliates received compensation for products and services other than investment banking services from Advanced Micro Devices,Airbnb Inc,Alibaba,Amazon.com, Inc.,Arm Holdings PLC,Baidu.com,Broadcom Inc,FinVolution,GoTo Gojek Tokopedia,Hello Group Inc,Intel Corp,Kuaishou,Marvell Technology Inc,Meituan,Meta Platforms Inc,Microsoft Corp.,MiniMax,NVIDIA Corp,NXP Semiconductors NV,Nokia OYJ,Salesforce Inc,Samsung Electronics,Tencent Holdings,Vodafone Group PLC,XPeng,Xiaomi in the past 12 months.

Citigroup Global Markets Inc. or its affiliates currently has, or had within the past 12 months, the following as investment banking client(s): Advanced Micro Devices,Airbnb Inc,Alibaba,Amazon.com, Inc.,Arm Holdings PLC,Broadcom Inc,Intel Corp,Kuaishou,Marvell Technology Inc,Meta Platforms Inc,Microsoft Corp.,NVIDIA Corp,NXP Semiconductors NV,Nokia OYJ,Salesforce Inc,Samsung Electronics,Tencent Holdings,Vodafone Group PLC,Xiaomi,Yuanbao.

Citigroup Global Markets Inc. or its affiliates currently has, or had within the past 12 months, the following as clients, and the services provided were non-investment-banking, securities-related: Advanced Micro Devices,Airbnb Inc,Alibaba,Amazon.com, Inc.,Arm Holdings PLC,Baidu.com,Broadcom Inc,FinVolution,GoTo Gojek Tokopedia,Hello Group Inc,Intel Corp,Kuaishou,Marvell Technology Inc,Meituan,Meta Platforms Inc,Microsoft Corp.,MiniMax,NVIDIA Corp,NXP Semiconductors NV,Nokia OYJ,Salesforce Inc,Samsung Electronics,Tencent Holdings,Vodafone Group PLC,XPeng,Xiaomi.

Citigroup Global Markets Inc. or its affiliates currently has, or had within the past 12 months, the following as clients, and the services provided were non-investment-banking, non-securities-related: Advanced Micro Devices,Airbnb Inc,Alibaba,Amazon.com, Inc.,Arm Holdings PLC,Baidu.com,Broadcom Inc,FinVolution,GoTo Gojek Tokopedia,Hello Group Inc,Intel Corp,Kuaishou,Marvell Technology Inc,Meituan,Meta Platforms Inc,Microsoft Corp.,MiniMax,NVIDIA Corp,NXP Semiconductors NV,Nokia OYJ,Salesforce Inc,Samsung Electronics,Tencent Holdings,Vodafone Group PLC,XPeng,Xiaomi.

Citigroup Global Markets Inc. and/or its affiliates has a significant financial interest in relation to Advanced Micro Devices,Airbnb Inc,Alibaba,Amazon.com, Inc.,Arm Holdings PLC,Baidu.com,Broadcom Inc,Intel Corp,Marvell Technology Inc,Meituan,Meta Platforms Inc,Microsoft Corp.,NVIDIA Corp,NXP Semiconductors NV,Nokia OYJ,Salesforce Inc,Samsung Electronics,Tencent Holdings,Vodafone Group PLC,XPeng,Xiaomi. (For an explanation of the determination of significant financial interest, please refer to the policy for managing conflicts of interest which can be found at www.citiVelocity.com.)

Citigroup Global Markets Inc. or its affiliates acts as a corporate broker to Vodafone Group PLC.

Analysts' compensation is determined by Citi Research management and Citigroup's senior management and is based upon activities and services intended to benefit the investor clients of Citigroup Global Markets Inc. and its affiliates (the "Firm"). Compensation is not linked to specific transactions or recommendations. Like all Firm employees, analysts receive compensation that is impacted by overall Firm profitability which includes investment banking, sales and trading, and principal trading revenues. One factor in equity research analyst compensation is arranging corporate access events between institutional clients and the management teams of covered companies. Typically, company management is more likely to participate when the analyst has a positive view of the company.

For financial instruments recommended in the Product in which the Firm is not a market maker, the Firm is a liquidity provider in such financial instruments (and any underlying instruments) and may act as principal in connection with transactions in such instruments. The Firm is a regular issuer of traded financial instruments linked to securities that may have been recommended in the Product. The Firm regularly trades in the securities of the issuer(s) discussed in the Product. The Firm may engage in securities transactions in a manner inconsistent with the Product and, with respect to securities covered by the Product, will buy or sell from customers on a principal basis.

The Firm is a market maker in the publicly traded equity securities of Alibaba,Baidu.com,Kuaishou,Meituan,MiniMax,Tencent Holdings,Xiaomi.

www.futouqaixin.cn

Unless stated otherwise neither the Research Analyst nor any member of their team has viewed the material operations of the Companies for which an investment view has been provided within the past 12 months.

For important disclosures (including copies of historical disclosures) regarding the companies that are the subject of this Citi Research product ("the Product"), please contact Citi Research, 388 Greenwich Street, 6th Floor, New York, NY, 10013, Attention: Legal/Compliance [E6WYB6412478]. In addition, the same important disclosures, with the exception of the Valuation and Risk assessments and historical disclosures, are contained on the Firm's disclosure website at https://www.citivelocity.com/cvr/eppublic/citi_research_disclosures. Valuation and Risk assessments can be found in the text of the most recent research note/report regarding the subject company. Pursuant to the Market Abuse Regulation a history of all Citi Research recommendations published during the preceding 12-month period can be accessed via Citi Velocity (https://www.citivelocity.com/cv2) or your standard distribution portal. Historical disclosures (for up to the past three years) will be provided upon request.

Citi Research Equity Ratings Distribution

	12 Month Rating			Catalyst Watch		
Data current as of 01 Apr 2026	**Buy**	**Hold**	**Sell**	**Buy**	**Hold**	**Sell**
Citi Research Global Fundamental Coverage (Neutral=Hold)	61%	32%	8%	37%	47%	16%
% of companies in each rating category that are investment banking clients	38%	41%	28%	42%	37%	36%

Guide to Citi Research Fundamental Research Investment Ratings:
Citi Research stock recommendations include an investment rating and an optional risk rating to highlight high risk stocks.
Risk rating takes into account both price volatility and fundamental criteria. Stocks will either have no risk rating or a High risk rating assigned.
Investment Ratings: Citi Research investment ratings are Buy, Neutral and Sell. Our ratings are a function of analyst expectations of expected total return ("ETR") and risk. ETR is the sum of the forecast price appreciation (or depreciation) plus the dividend yield for a stock within the next 12 months. The target price is based on a 12 month time horizon. The Investment rating definitions are: Buy (1) ETR of 15% or more or 25% or more for High risk stocks; and Sell (3) for negative ETR. Any covered stock not assigned a Buy or a Sell is a Neutral (2). For stocks rated Neutral (2), if an analyst believes that there are insufficient valuation drivers and/or investment catalysts to derive a positive or negative investment view, they may elect with the approval of Citi Research management not to assign a target price and, thus, not derive an ETR. Citi Research may suspend its rating and target price and assign “Rating Suspended” status for regulatory and/or internal policy reasons. Citi Research may also suspend its rating and target price and assign "Under Review" status for other exceptional circumstances (e.g. lack of information critical to the analyst's thesis, trading suspension) affecting the company and/or trading in the company's securities. In both such situations, the rating and target price will show as “—“ and “-“ respectively in the rating history price chart. Prior to 11 April 2022 Citi Research assigned “Under Review” status to both situations and prior to 11 Nov 2020 only in exceptional circumstances. As soon as practically possible, the analyst will publish a note re-establishing a rating and investment thesis. Investment ratings are determined by the ranges described above at the time of initiation of coverage, a change in investment and/or risk rating, or a change in target price (subject to limited management discretion). At times, the expected total returns may fall outside of these ranges because of market price movements and/or other short-term volatility or trading patterns. Such interim deviations will be permitted but will become subject to review by Research Management. Your decision to buy or sell a security should be based upon your personal investment objectives and should be made only after evaluating the stock's expected performance and risk.

Catalyst Watch/Short Term Views (“STV”) Ratings Disclosure:
Catalyst Watch and STV Upside/Downside calls: Citi Research may also include a Catalyst Watch or STV Upside or Downside call to indicate the analyst expects the share price to rise (fall) in absolute terms over a specified period of 30 or 90 days in reaction to one or more specific near-term catalysts or events impacting the company or the market. A Catalyst Watch will be published when Analyst confidence is high that an impact to share price will occur; it will be a STV when confidence level is moderate. A Catalyst Watch or STV Upside/Downside call will automatically expire at the end of the specified 30/90 day period. The Catalyst Watch will also be automatically removed if share price performance (calculated at market close) exceeds 15% against the direction of the call (unless over-ridden by the analyst). The analyst may also remove a Catalyst Watch or STV call prior to the end of the specified period in a published research note. A Catalyst Watch/STV Upside or Downside call may be different from and does not affect a stock’s fundamental equity rating, which reflects a longer-term total absolute return expectation. For purposes of FINRA ratings-distribution-disclosure rules, a Catalyst Watch/STV Upside call corresponds to a buy recommendation and a Catalyst Watch/STV Downside call corresponds to a sell recommendation. Any stock not assigned to a Catalyst Watch Upside, Catalyst Watch Downside, STV Upside, or STV Downside call is considered Catalyst Watch/STV No View. For purposes of FINRA ratings distribution-disclosure rules, we correspond Catalyst Watch/STV No View to Hold in our ratings distribution table for our Catalyst Watch/STV Upside/Downside rating system. However, we reiterate that we do not consider No View to be a recommendation. For all Catalyst Watch/STV Upside/Downside calls, risk exists that the catalyst(s) and associated share-price movement will not materialize as expected.

RESEARCH ANALYST AFFILIATIONS / NON-US RESEARCH ANALYST DISCLOSURES
The legal entities employing the authors of this report are listed below (and their regulators are listed further herein). Non-US research analysts who have prepared this report (i.e., all research analysts listed below other than those identified as employed by Citigroup Global Markets Inc.) are not registered/qualified as research analysts with FINRA. Such research analysts may not be associated persons of the member organization (but are employed by an affiliate of the member organization) and therefore may not be subject to the FINRA Rule 2241 restrictions on communications with a subject company, public appearances and trading securities held by a research analyst account.

www.futoucaifxin.cn

Citigroup Global Markets Asia Limited Nelson Cheung; Alicia Yap, CFA

OTHER DISCLOSURES

Any price(s) of instruments mentioned in recommendations are as of the prior day's market close on the primary market for the instrument, unless otherwise stated.

The completion and first dissemination of any recommendations made within this research report are as of the Eastern date-time displayed at the top of the Product. If the Product references views of other analysts then please refer to the price chart or rating history table for the date/time of completion and first dissemination with respect to that view.

Regulations in various jurisdictions require that where a recommendation differs from any of the author's previous recommendations concerning the same financial instrument or issuer that has been published during the preceding 12-month period that the change(s) and the date of that previous recommendation are indicated. For fundamental coverage please refer to the price chart or rating change history within this disclosure appendix or the issuer disclosure summary at https://www.citivelocity.com/cvr/eppublic/citi_research_disclosures.

Citi Research has implemented policies for identifying, considering and managing potential conflicts of interest arising as a result of publication or distribution of investment research. A description of these policies can be found at https://www.citivelocity.com/cvr/eppublic/citi_research_disclosures.

The proportion of all Citi Research research recommendations that were the equivalent to "Buy","Hold","Sell" at the end of each quarter over the prior 12 months (with the % of these that had received investment firm services from Citi in the prior 12 months shown in brackets) is as follows: Q4 2025 Buy 33% (63%), Hold 44% (50%), Sell 23% (46%), RV 0.4% (91%); Q3 2025 Buy 33% (61%), Hold 44% (52%), Sell 23% (50%), RV 0.4% (80%); Q2 2025 Buy 33%(63%), Hold 44% (51%), Sell 23% (49%), RV 0.4% (86%); Q1 2025 Buy 33%(63%), Hold 44% (51%), Sell 23% (49%), RV 0.5% (87%). For the purposes of disclosing recommendations other than for equity (whose definitions can be found in the corresponding disclosure sections), "Buy" means a positive directional trade idea; "Sell" means a negative directional trade idea; and "Relative Value" means any trade idea which does not have a clear direction to the investment strategy.

European regulations require a 5 year price history when past performance of a security is referenced. CitiVelocity's Charting Tool (https://www.citivelocity.com/cv2/#go/CHARTING_3_Equities) provides the facility to create customisable price charts including a five year option. This tool can be found in the Data & Analytics section under any of the asset class menus in CitiVelocity (https://www.citivelocity.com/). For further information contact CitiVelocity support (https://www.citivelocity.com/cv2/go/CLIENT_SUPPORT). The source for all referenced prices, unless otherwise stated, is DataCentral, which sources price information from LSEG Data & Analytics. Past performance is not a guarantee or reliable indicator of future results. Forecasts are not a guarantee or reliable indicator of future performance.

Investors should always consider the investment objectives, risks, and charges and expenses of an ETF carefully before investing. The applicable prospectus and key investor information document (as applicable) for an ETF should contain this and other information about such ETF. It is important to read carefully any such prospectus before investing. Clients may obtain prospectuses and key investor information documents for ETFs from the applicable distributor or authorized participant, the exchange upon which an ETF is listed and/or from the applicable website of the applicable ETF issuer. The value of the investments and any accruing income may fall or rise. Any past performance, prediction or forecast is not indicative of future or likely performance. Any information on ETFs contained herein is provided strictly for illustrative purposes and should not be deemed an offer to sell or a solicitation of an offer to purchase units of any ETF either explicitly or implicitly. The opinions expressed are those of the authors and do not necessarily reflect the views of ETF issuers, any of their agents or their affiliates.

Citigroup Global Markets India Private Limited and/or its affiliates may have, from time to time, actual or beneficial ownership of 1% or more in the debt securities of the subject issuer.

Please be advised that pursuant to Executive Order 13959 as amended (the "Order"), U.S. persons are prohibited from investing in securities of any company determined by the United States Government to be the subject of the Order. This research is not intended to be used or relied upon in any way that could result in a violation of the Order. Investors are encouraged to rely upon their own legal counsel for advice on compliance with the Order and other economic sanctions programs administered and enforced by the Office of Foreign Assets Control of the U.S. Treasury Department.

This communication is directed at persons who are "Eligible Clients" as such term is defined in the Israeli Regulation of Investment Advice, Investment Marketing and Investment Portfolio Management law, 1995 (the "Advisory Law"). Within Israel, this communication is not intended for retail clients and Citi will not make such products or transactions available to retail clients or to non-Eligible Clients. The presenter is not licensed as investment advisor or investment marketer by the Israeli Securities Authority ("ISA") and this communication does not constitute investment or marketing advice. The information contained herein may relate to matters that are not regulated by the ISA. Any securities which are the subject of this communication may not be offered or sold to any Israeli person except pursuant to a security offering exemption according to the Israeli Securities Law, 1968and the public offering rules provided thereunder.

Citi Research broadly and simultaneously disseminates its research content to the Firm's institutional and retail clients via the Firm's proprietary electronic distribution platforms (e.g., Citi Velocity and various Global Wealth platforms). As a convenience, certain, but not all, research content may be distributed through third party aggregators. Clients may receive published research reports by email, on a discretionary basis, and only after such research content has been broadly disseminated. Certain research is made available only to institutional investors to satisfy regulatory requirements. The level and types of services provided by Citi Research analysts to clients may vary depending on various factors such as the client's individual preferences as to the frequency and manner of receiving communications from analysts, the client's risk profile and investment

www.tuoluoaixin.cn

focus and perspective (e.g. market-wide, sector specific, long term, short-term etc.), the size and scope of the overall client relationship with the Firm and legal and regulatory constraints.

Pursuant to Comissão de Valores Mobiliários Resolução 20 and ASIC Regulatory Guide 264, Citi is required to disclose whether a Citi related company or business has a commercial relationship with the subject company. Considering that Citi operates multiple businesses in more than 100 countries around the world, it is likely that Citi has a commercial relationship with the subject company.

Securities recommended, offered, or sold by the Firm: (i) are not insured by the Federal Deposit Insurance Corporation; (ii) are not deposits or other obligations of any insured depository institution (including Citibank); and (iii) are subject to investment risks, including the possible loss of the principal amount invested. The Product is for informational purposes only and is not intended as an offer or solicitation for the purchase or sale of a security. Any decision to purchase securities mentioned in the Product must take into account existing public information on such security or any registered prospectus. Although information has been obtained from and is based upon sources that the Firm believes to be reliable, we do not guarantee its accuracy and it may be incomplete and condensed. Note, however, that the Firm has taken all reasonable steps to determine the accuracy and completeness of the disclosures made in the Important Disclosures section of the Product. The Firm's research department has received assistance from the subject company(ies) referred to in this Product including, but not limited to, discussions with management of the subject company(ies). Firm policy prohibits research analysts from sending draft research to subject companies. However, it should be presumed that the author of the Product has had discussions with the subject company to ensure factual accuracy prior to publication. Statements and views concerning ESG (environmental, social, governance) factors are typically based upon public statements made by the affected company or other public news, which the author may not have independently verified. ESG factors are one consideration that investors may choose to examine when making investment decisions. All opinions, projections and estimates constitute the judgment of the author as of the date of the Product and these, plus any other information contained in the Product, are subject to change without notice. Prices and availability of financial instruments also are subject to change without notice. Notwithstanding other departments within the Firm advising the companies discussed in this Product, information obtained in such role is not used in the preparation of the Product. Although Citi Research does not set a predetermined frequency for publication, if the Product is a fundamental equity or credit research report, it is the intention of Citi Research to provide research coverage of the covered issuers, including in response to news affecting the issuer. For non-fundamental research reports, Citi Research may not provide regular updates to the views, recommendations and facts included in the reports. Notwithstanding that Citi Research maintains coverage on, makes recommendations concerning or discusses issuers, Citi Research may be periodically restricted from referencing certain issuers due to legal or policy reasons. Where a component of a published trade idea is subject to a restriction, the trade idea will be removed from any list of open trade ideas included in the Product. Upon the lifting of the restriction, the trade idea will either be re-instated in the open trade ideas list if the analyst continues to support it or it will be officially closed. Citi Research may provide different research products and services to different classes of customers (for example, based upon long-term or short-term investment horizons) that may lead to differing conclusions or recommendations that could impact the price of a security contrary to the recommendations in the alternative research product, provided that each is consistent with the rating system for each respective product.

Investing in non-U.S. securities, including ADRs, may entail certain risks. The securities of non-U.S. issuers may not be registered with, nor be subject to the reporting requirements of the U.S. Securities and Exchange Commission. There may be limited information available on foreign securities. Foreign companies are generally not subject to uniform audit and reporting standards, practices and requirements comparable to those in the U.S. Securities of some foreign companies may be less liquid and their prices more volatile than securities of comparable U.S. companies. In addition, exchange rate movements may have an adverse effect on the value of an investment in a foreign stock and its corresponding dividend payment for U.S. investors. Net dividends to ADR investors are estimated, using withholding tax rates conventions, deemed accurate, but investors are urged to consult their tax advisor for exact dividend computations. Investors who have received the Product from the Firm may be prohibited in certain states or other jurisdictions from purchasing securities mentioned in the Product from the Firm. Please ask your Financial Consultant for additional details. Citigroup Global Markets Inc. takes responsibility for the Product in the United States. Any orders by US investors resulting from the information contained in the Product may be placed only through Citigroup Global Markets Inc.

The Citigroup legal entity that takes responsibility for the production of the Product is the legal entity which the first named author is employed by.

The Product is made available in **Australia** through Citigroup Global Markets Australia Pty Limited. (ABN 64 003 114 832 and AFSL No. 240992), participant of the ASX Group and regulated by the Australian Securities & Investments Commission. Citigroup Centre, 2 Park Street, Sydney, NSW 2000. Citigroup Global Markets Australia Pty Limited is not an Authorised Deposit-Taking Institution under the Banking Act 1959, nor is it regulated by the Australian Prudential Regulation Authority.

The Product is made available in **Brazil** by Citigroup Global Markets Brasil - CCTVM SA, which is regulated by CVM - Comissão de Valores Mobiliários ("CVM"), BACEN - Brazilian Central Bank, APIMEC - Associação dos Analistas e Profissionais de Investimento do Mercado de Capitais and ANBIMA – Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais. Av. Paulista, 1111 - 14º andar(parte) - CEP: 01311920 - São Paulo - SP.

This Product is available in **Chile** through Banchile Corredores de Bolsa S.A., an indirect subsidiary of Citigroup Inc., which is regulated by the Comisión Para El Mercado Financiero. Enrique Foster Sur, 20, piso 6, Las Condes, Santiago, Chile.

Disclosure for investors in the Republic of Colombia - This communication or message does not constitute a professional recommendation to make investment in the terms of article 2.40.1.1.2 of Decree 2555 de 2010 or the regulations that modify,

www.reversetention.com

substitute or complement it. Para la elaboración y distribución de informes de investigación y de comunicaciones generales de que trata este artículo no se requiere ser una entidad vigilada por la Superintendencia Financiera de Colombia.

The Product is made available in **Germany** by Citigroup Global Markets Europe AG ("CGME"), which is regulated by the European Central Bank and the German Federal Financial Supervisory Authority (Bundesanstalt fur Finanzdienstleistungsaufsicht BaFin). Börsenplatz 9, 60313 Frankfurt am Main, Germany.

Unless otherwise specified, if the analyst who prepared this report is based in Hong Kong and it relates to "securities" (as defined in the Securities and Futures Ordinance (Cap.571 of the Laws of Hong Kong)), the report is issued in Hong Kong by Citigroup Global Markets Asia Limited. Citigroup Global Markets Asia Limited is regulated by Hong Kong Securities and Futures Commission. If the report is prepared by a non-Hong Kong based analyst, please note that such analyst (and the legal entity that the analyst is employed by or accredited to) is not licensed/registered in Hong Kong and they do not hold themselves out as such. Please refer to the section "Research Analyst Affiliations / Non-US Research Analyst Disclosures" for the details of the employment entity of the analysts.

The Product is made available in **India** by Citigroup Global Markets India Private Limited (CGM), which is regulated by the Securities and Exchange Board of India (SEBI), as a Research Analyst (SEBI Registration No. INH000000438). CGM is also actively involved in the business of merchant banking (SEBI Registration No. INM000010718) and stock brokerage ((SEBI Registration No. INZ000263033) in India, and is registered with SEBI in this regard. Registration granted by SEBI and certification from National Institute of Securities Markets (NISM) in no way guarantee performance of the intermediary or provide any assurance of returns to investors. CGM's registered office is at 1202, 12th Floor, First International Financial Centre (FIFC), G Block, Bandra Kurla Complex, Bandra East, Mumbai – 400098 & registered Tel: +91 22 61759999. Citi maintains robust policies, procedures,controls, and training to ensure continued compliance with all applicable rulesand regulations. All recommendations contained herein are made by dulyqualified research analysts. CGM's Corporate Identity Number is U99999MH2000PTC126657, and its Compliance Officer [Vishal Bohra] contact details are: Tel:+91-022-61759994, Fax:+91-022-61759851, Email: cgmcompliance@citi.com. The Investor Charter in respect of Research Analysts, the Compliance Audit Report and Complaints information can be found at https://www.citivelocity.com/cvr/eppublic/citi_research_disclosures. The grievance officer [Niraj Mody] contact details are Tel: +91-22-42775002, Email: EMEA.CR.Complaints@citi.com. Investment in securities market are subject to market risks. Read all the related documents carefully before investing. SEBI prescribed Client Terms & Conditions can be found at https://www.citivelocity.com/rendition/authfilelinksvcs/eppublic/V1/file?paramData=ZmlsZU5hbWU9L1MzL0dETVMvcHViRGlzY2xvc3VyZUh0bWxzL2dkbV9kaXNfc2ViaV9UZXJtc29mVXNlLmh0bWw

The Product is made available in **Indonesia** through PT Citigroup Sekuritas Indonesia. Citibank Tower 10/F, Pacific Century Place, SCBD lot 10, Jl. Jend Sudirman Kav 52-53, Jakarta 12190, Indonesia. Neither this Product nor any copy hereof may be distributed in Indonesia or to any Indonesian citizens wherever they are domiciled or to Indonesian residents except in compliance with applicable capital market laws and regulations. This Product is not an offer of securities in Indonesia. The securities referred to in this Product have not been registered with the Capital Market and Financial Services Authority (OJK) pursuant to relevant capital market laws and regulations, and may not be offered or sold within the territory of the Republic of Indonesia or to Indonesian citizens through a public offering or in circumstances which constitute an offer within the meaning of the Indonesian capital market laws and regulations.

The Product is made available in **Japan** by Citigroup Global Markets Japan Inc. ("CGMJ"), which is regulated by Financial Services Agency, Securities and Exchange Surveillance Commission, Japan Securities Dealers Association, Tokyo Stock Exchange and Osaka Securities Exchange. Otemachi Park Building, 1-1-1 Otemachi, Chiyoda-ku, Tokyo 100-8132 Japan. In the event that an error is found in an CGMJ research report, a revised version will be posted on the Firm's Citi Velocity website. If you have questions regarding Citi Velocity, please call (81 3) 6270-3019 for help.

The product is made available in the **Kingdom of Saudi Arabia** in accordance with Saudi laws through Citigroup Saudi Arabia, which is regulated by the Capital Market Authority (CMA) under CMA license (17184-31). 2239 Al Urubah Rd – Al Olaya Dist. Unit No. 18, Riyadh 12214 – 9597, Kingdom Of Saudi Arabia.

The Product is made available in **Korea** by Citigroup Global Markets Korea Securities Ltd. (CGMK), which is regulated by the Financial Services Commission, the Financial Supervisory Service and the Korea Financial Investment Association (KOFIA). The address of CGMK is Citibank Center, 50 Saemunan-ro, Jongno-gu, Seoul 03184, Korea. KOFIA makes available registration information of research analysts on its website. Please visit the following website if you wish to find KOFIA registration information on research analysts of CGMK. http://dis.kofia.or.kr/websquare/index.jsp?w2xPath=/wq/fundMgr/DISFundMgrAnalystList.xml&divisionId=MDIS03002002000000&serviceId=SDIS03002002000. The Product is made available in Korea by Citibank Korea Inc., which is regulated by the Financial Services Commission and the Financial Supervisory Service. Address is Citibank Center, 50 Saemunan-ro, Jongno-gu, Seoul 03184, Korea. This research report is intended to be provided only to Professional Investors as defined in the Financial Investment Services and Capital Market Act and its Enforcement Decree in Korea.

The Product is made available in **Malaysia** by Citigroup Global Markets Malaysia Sdn Bhd (Registration No. 199801004692 (460819-D)) ("CGMM") to its clients and CGMM takes responsibility for its contents as regards CGMM's clients. CGMM is regulated by the Securities Commission Malaysia. Please contact CGMM at Level 43 Menara Citibank, 165 Jalan Ampang, 50450 Kuala Lumpur, Malaysia in respect of any matters arising from, or in connection with, the Product.

The Product is made available in **Mexico** by Citi México Casa de Bolsa, S.A. de C.V., Grupo Financiero Citi México which is a wholly owned subsidiary of Citigroup Inc. and is regulated by Comision Nacional Bancaria y de Valores. Prolongación Reforma 1196, 24 floor, Colonia Santa Fe, Alcaldía Cuajimalpa de Morelos, C.P. 05348, Ciudad de México.

www.fulltextreports.com

The Product is made available in **Poland** by Biuro Maklerskie Banku Handlowego (DMBH), separate department of Bank Handlowy w Warszawie S.A. a subsidiary of Citigroup Inc., which is regulated by Komisja Nadzoru Finansowego. Biuro Maklerskie Banku Handlowego (DMBH), ul.Senatorska 16, 00-923 Warszawa.

The Product is made available in **Singapore** through Citigroup Global Markets Singapore Pte. Ltd. ("CGMSPL"), a capital markets services and Exempt Financial Advisor license holder, and regulated by Monetary Authority of Singapore. Please contact CGMSPL at 8 Marina View, 21st Floor Asia Square Tower 1, Singapore 018960, in respect of any matters arising from, or in connection with, the analysis of this document. This Product is intended for recipients who are accredited, expert and institutional investors as defined under the Securities and Futures Act 2001. For Citi Private Bank, the Product is made available in Singapore by Citi Private Bank through Citibank, N.A., Singapore Branch. Citibank N.A., Singapore Branch is a licensed bank in Singapore that is regulated by the Monetary Authority of Singapore. Please contact your Private Banker in Citibank N.A., Singapore Branch if you have any queries on or any matters arising from or in connection with this document. The Product is intended for recipients who are accredited, expert and institutional investors as defined under the Securities and Futures Act 2001. For Citibank Singapore Limited ("CSL"), the Product is distributed in Singapore by CSL to selected Citigold/Citigold Private Clients. CSL provides no independent research or analysis of the substance or in preparation of the Product. Please contact your Citigold//Citigold Private Client Relationship Manager in CSL if you have any queries on or any matters arising from or in connection with this document. The Product is intended for recipients who are accredited investors as defined under the Securities and Futures Act (Cap. 289).

Citigroup Global Markets (Pty) Ltd. is incorporated in the **Republic of South Africa** (company registration number 2000/025866/07) and its registered office is at 145 West Street, Sandton, 2196, Saxonwold. Citigroup Global Markets (Pty) Ltd. is regulated by JSE Securities Exchange South Africa, South African Reserve Bank and the Financial Services Board. The investments and services contained herein are not available to private customers in South Africa.

The Product is made available in the **Republic of China (Taiwan)** through Citigroup Global Markets Taiwan Securities Company Ltd. ("CGMTS"), 14F, 15F and 16F, No. 1, Songzhi Road, Taipei 110, Taiwan, subject to the license scope and the applicable laws and regulations in the Republic of China (Taiwan). CGMTS is regulated by the Securities and Futures Bureau of the Financial Supervisory Commission of Taiwan, the Republic of China (Taiwan). No portion of the Product may be reproduced or quoted in the Republic of China (Taiwan) by the press or any third parties [without the written authorization of CGMTS]. Pursuant to the applicable laws and regulations in the Republic of China (Taiwan), the recipient of the Product shall not take advantage of such Product to involve in any matters in which the recipient may have conflicts of interest. If the Product covers securities which are not allowed to be offered or traded in the Republic of China (Taiwan), neither the Product nor any information contained in the Product shall be considered as advertising the securities or making recommendation of the securities in the Republic of China (Taiwan). The Product is for informational purposes only and is not intended as an offer or solicitation for the purchase or sale of a security or financial products. Any decision to purchase securities or financial products mentioned in the Product must take into account existing public information on such security or the financial products or any registered prospectus.

The Product is made available in **Thailand** through Citicorp Securities (Thailand) Ltd., which is regulated by the Securities and Exchange Commission of Thailand. 399 Interchange 21 Building, 18th Floor, Sukhumvit Road, Klongtoey Nua, Wattana ,Bangkok 10110, Thailand.

Disclosure for investors in the Republic of Turkey: The Product is made available in **Turkey** through Citibank AS which is regulated by Capital Markets Board. Tekfen Tower, Eski Buyukdere Caddesi # 209 Kat 2B, 23294 Levent, Istanbul, Turkey. Under Capital Markets Law of Turkey (Law No: 6362), the investment information, comments and recommendations stated here, are not within the scope of investment advisory activity. Comments and recommendations stated here rely on the individual opinions of the ones providing these comments and recommendations. These opinions may not fit to your financial status, risk and return preferences. For this reason, to make an investment decision by relying solely to this information stated here may not bring about outcomes that fit your expectations. Furthermore, Citi Research is a division of Citigroup Global Markets Inc. (the "Firm"), which does and seeks to do business with companies and/or trades on securities covered in this research reports. As a result, investors should be aware that the Firm may have a conflict of interest that could affect the objectivity of this report, however investors should also note that the Firm has in place organisational and administrative arrangements to manage potential conflicts of interest of this nature.

In the **U.A.E**, these materials (the "Materials") are communicated byCitigroup Global Markets Limited, DIFC branch ("CGML"), an entityregistered in the Dubai International Financial Center ("DIFC") andlicensed and regulated by the Dubai Financial Services Authority ("DFSA",license #CL0221) to Professional Clients and Market Counterparties, as definedin DFSA regulations, only and should not be relied upon or distributed toRetail Clients. Financial products and/or services to which the Materialsrelate will only be made available to Professional Clients and MarketCounterparties. Citigroup Global Markets Limited DIFC Branch registered addressis Level 3, Gate District Building 02, Dubai International Financial Centre andcan be contacted on +971 4 509 97 90.

The Product is made available in **United Kingdom** by Citigroup Global Markets Limited, which is authorised by the Prudential Regulation Authority ("PRA") and regulated by the Financial Conduct Authority ("FCA") and the PRA. This material may relate to investments or services of a person outside of the UK or to other matters which are not authorised by the PRA nor regulated by the FCA and the PRA and further details as to where this may be the case are available upon request in respect of this material. Citigroup Centre, Canada Square, Canary Wharf, London, E14 5LB.

The Product is made available in **United States** and **Canada** by Citigroup Global Markets Inc., which is a member of FINRA and registered with the US Securities and Exchange Commission. 388 Greenwich Street, New York, NY 10013.

www.reportdiary.com

Unless specified to the contrary, within EU Member States, the Product is made available by Citigroup Global Markets Europe AG ("CGME"), which is regulated by the European Central Bank and the German Federal Financial Supervisory Authority (Bundesanstalt für Finanzdienstleistungsaufsicht-BaFin).

The Product is not to be construed as providing investment services in any jurisdiction where the provision of such services would not be permitted.

Subject to the nature and contents of the Product, the investments described therein are subject to fluctuations in price and/or value and investors may get back less than originally invested. Certain high-volatility investments can be subject to sudden and large falls in value that could equal or exceed the amount invested.The yield and average life of CMOs (collateralized mortgage obligations) referenced in this Product will fluctuate depending on the actual rate at which mortgage holders prepay the mortgages underlying the CMO and changes in current interest rates. Any government agency backing of the CMO applies only to the face value of the CMO and not to any premium paid. Certain investments contained in the Product may have tax implications for private customers whereby levels and basis of taxation may be subject to change. If in doubt, investors should seek advice from a tax adviser. The Product does not purport to identify the nature of the specific market or other risks associated with a particular transaction. Advice in the Product is general and should not be construed as personal advice given it has been prepared without taking account of the objectives, financial situation or needs of any particular investor. Accordingly, investors should, before acting on the advice, consider the appropriateness of the advice, having regard to their objectives, financial situation and needs. Prior to acquiring any financial product, it is the client's responsibility to obtain the relevant offer document for the product and consider it before making a decision as to whether to purchase the product.

Card Insights. Where this report references Card Insights data, Card Insights consists of selected data from a subset of Citi's proprietary credit card transactions. Such data has undergone rigorous security protocols to keep all customer information confidential and secure; the data is highly aggregated and anonymized so that all unique customer identifiable information is removed from the data prior to receipt by the report's author or distribution to external parties. This data should be considered in the context of other economic indicators and publicly available information. Further, the selected data represents only a subset of Citi's proprietary credit card transactions due to the selection methodology or other limitations and should not be considered as indicative or predictive of the past or future financial performance of Citi or its credit card business.

Citi Research product may source data from dataCentral. dataCentral is a Citi Research proprietary database, which includes the Firm's estimates, data from company reports and feeds from LSEG Data & Analytics. The source for all referenced prices, unless otherwise stated, is DataCentral. Past performance is not a guarantee or reliable indicator of future results. Forecasts are not a guarantee or reliable indicator of future performance. The printed and printable version of the research report may not include all the information <(e.g. certain financial summary information and comparable company data) that is linked to the online version available on the Firm's proprietary electronic distribution platforms.

Where included in this report, MSCI sourced information is the exclusive property of Morgan Stanley Capital International Inc. (MSCI). Without prior written permission of MSCI, this information and any other MSCI intellectual property may not be reproduced, redisseminated or used to create any financial products, including any indices. This information is provided on an "as is" basis. The user assumes the entire risk of any use made of this information. MSCI, its affiliates and any third party involved in, or related to, computing or compiling the information hereby expressly disclaim all warranties of originality, accuracy, completeness, merchantability or fitness for a particular purpose with respect to any of this information. Without limiting any of the foregoing, in no event shall MSCI, any of its affiliates or any third party involved in, or related to, computing or compiling the information have any liability for any damages of any kind. MSCI, Morgan Stanley Capital International and the MSCI indexes are services marks of MSCI and its affiliates. Where data is attributed to Morningstar that data is © 2026 Morningstar, Inc. All Rights Reserved. That information: (1) is proprietary to Morningstar and/or its content providers; (2) may not be copied or distributed; and (3) is not warranted to be accurate, complete or timely. Neither Morningstar nor its content providers are responsible for any damages or losses arising from any use of this information.

The Firm accepts no liability whatsoever for the actions of third parties. The Product may provide the addresses of, or contain hyperlinks to, websites. Except to the extent to which the Product refers to website material of the Firm, the Firm has not reviewed the linked site. Equally, except to the extent to which the Product refers to website material of the Firm, the Firm takes no responsibility for, and makes no representations or warranties whatsoever as to, the data and information contained therein. Such address or hyperlink (including addresses or hyperlinks to website material of the Firm) is provided solely for your convenience and information and the content of the linked site does not in any way form part of this document. Accessing such website or following such link through the Product or the website of the Firm shall be at your own risk and the Firm shall have no liability arising out of, or in connection with, any such referenced website.

© 2026 Citigroup Global Markets Inc. Citi Research is a division of Citigroup Global Markets Inc. Citi and Citi and Arc Design are trademarks and service marks of Citigroup Inc. and its affiliates and are used and registered throughout the world. All rights reserved. The research data in this report are not intended to be used for the purpose of (a) determining the price of or amounts due in respect of (or to value) one or more financial products or instruments and/or (b) measuring or comparing the performance of, or defining the asset allocation of a financial product, a portfolio of financial instruments, or a collective investment undertaking, and any such use is strictly prohibited without the prior written consent of Citi Research. Any unauthorized use, duplication, redistribution or disclosure of this report (the "Product"), including, but not limited to, redistribution of the Product by electronic mail, posting of the Product on a website or page, and/or providing to a third party a link to the Product, is prohibited by law and will result in prosecution. The information contained in the Product is intended solely for the recipient and may not be further distributed by the recipient to any third party.

www.fuyoudaili.com

ADDITIONAL INFORMATION IS AVAILABLE UPON REQUEST

www.futoucaixin.cn