

聚焦中国AI模型

LLM指南：全球普及的智能临界点 (摘要)

从去年的DeepSeek时刻（成本效益）到今年的智谱GLM时刻（模型智能）

我们认为，中国的AI开源/开放权重模型在智能表现上相对于全球闭源模型而言正达到一个临界点，在国内企业和全球中小企业中的采用率显著提升，这将形成一个正向的数据飞轮，推动模型进一步改进。我们评估了这些模型如何以低成本/有限的计算资源实现了这样的性能；它们为何采用开源/开放权重方式以及如何实现货币化；随着全球企业从“token用量最大化”转向“投资回报率优先”，关键的潜在市场在哪里（在此我们重点关注中国模型的两个优势ARR象限）；以及在我们的竞争定位框架下，谁将成为潜在的长期赢家。我们指出，中国最前沿模型出海限制的任何收紧、西方市场对中国模型的政策、以及模型训练中对高端算力的获取，是我们对中国AI模型token/收入增长轨迹预测面临的三项关键不确定因素/风险。

我们推出基于定价能力、成本优势和财务实力的中国AI模型竞争定位框架，并结合token规模 and 市场份额的进展进行分析，认为智谱（参见首次覆盖报告）和深度求索（DeepSeek，未上市）在基础模型领域处于最有利位置，而字节跳动（未上市）在多模态领域处于最有利位置。

姜朗霆, CFA

+852-2978-0856 |
ronald.keung@gs.com
高盛（亚洲）有限责任公司

Eric Sheridan

+1(917)343-8683 |
eric.sheridan@gs.com
高盛集团

张博凯

+852-2978-2930 |
allen.k.chang@gs.com
高盛（亚洲）有限责任公司

孔令侃, CFA

+852-2978-6603 |
lincoln.kong@gs.com
高盛（亚洲）有限责任公司

Timothy Zhao

+852-2978-2673 |
timothy.zhao@gs.com
高盛（亚洲）有限责任公司

谢元滔

+852-2978-1398 | damian.xie@gs.com
高盛（亚洲）有限责任公司

Iris Xiao

+852-2978-0202 | iris.xiao@gs.com
高盛（亚洲）有限责任公司

邱品儒

+852-2978-2672 | steve.qiu@gs.com
高盛（亚洲）有限责任公司

周璐晴

+852-3465-4207 | luqing.zhou@gs.com
高盛（亚洲）有限责任公司

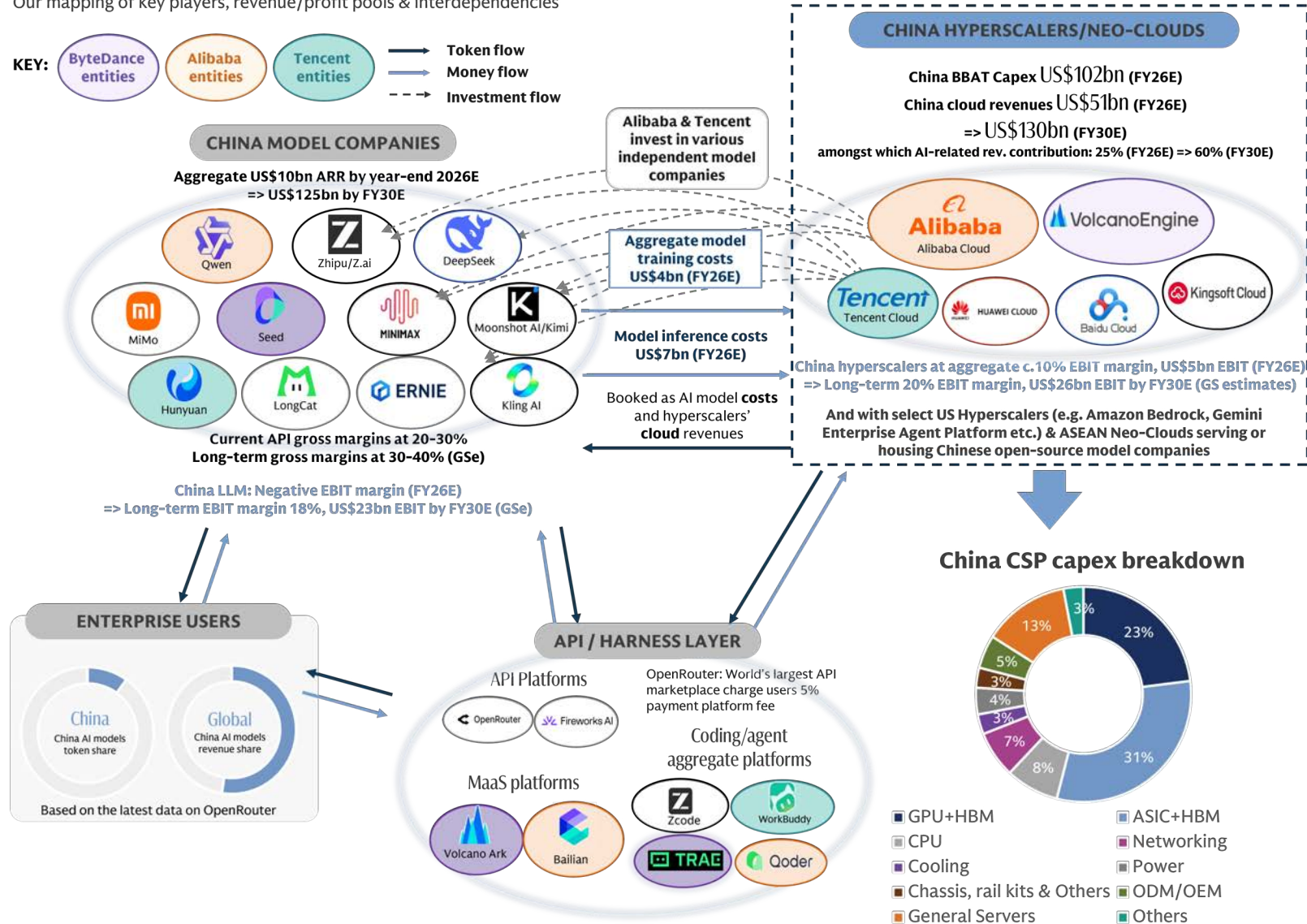
高盛与其研究报告所分析的企业存在业务关系，并且继续寻求发展这些关系。因此，投资者应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为作出投资决策的唯一因素。有关分析师的申明和其他重要信息，见信息披露附录，或参阅www.gs.com/research/hedge.html。由非美国附属公司聘用的分析师不是美国FINRA的注册/合格研究分析师。

目录

Dissecting China open-source models unit economics today and path to profitability	4
PM Summary	5
Comparing China' s key LLM players	12
Decoding China' s AI model tokens & our forecasts on token/revenue share	13
How do Chinese models achieve competitive performance at low costs/tight computing resources?	14
Why are Chinese models pursuing an open source/open weight approach, and ways to monetize?	22
What are the key addressable markets, domestically and internationally, and key risks?	25
Who are best positioned to be the long term winners? Introducing our Competitive Positioning framework	28
Foundation models	28
Multi-modal generation models	34
Key upside and downside risks	39
China' s key AI players at a glance	41
Price Target Methodology & Risks	48
信息披露附录	50

CHINA AI MODEL & HYPERSCALERS ECOSYSTEM

Our mapping of key players, revenue/profit pools & interdependencies



1) The calculation of China cloud revenues includes Alibaba, Tencent, ByteDance, Baidu, and Kingsoft Cloud. 2) China CSP capex breakdown is based on GS Greater China Tech team estimates; 3) Companies included are for illustrative purposes and may not be an exhaustive list.

资料来源：公司数据，高盛全球投资研究部

Dissecting China open-source models unit economics today and path to profitability

图表 1: Unit economics for China's value-for-money agentic model and top performing coding model, vs. hypothetical economics for global SOTA model

UNIT ECONOMICS OF TWO-TIERED CHINA AI MODELS (PER 1M BLENDED TOKENS)

	Current (Year-end 26E)			Long-term	
	China's value-for-money agentic model	China's best performing coding model	Hypothetical economics for 'most frontier' SOTA model	China's value-for-money agentic model	China's best performing coding model
API Pricing	\$0.2	\$1	\$3.9	\$0.2	\$2
Inference cost as % of ARR	-74%	-70%	-26%	-57%	-51%
Gross Profit (ARR - inference cost)	0.03	0.18	2.70	0.07	0.80
% Gross margin	10-15%	15-20%	70%	30-40%	30-40%
Training as % of ARR	-30%	-40%	-49%	-13%	-13%
EBIT (Gross profit - Training cost & OpEx)	-0.06	-0.39	0.00	0.03	0.44
% EBIT margin	-30%	-39%	0%	14%	22%

CHINA'S VALUE FOR MONEY MODELS BLENDED PRICE: US\$0.2 per blended 1M tokens

(A)

(B)

= (A) - (B)

(C)

= (A) - (B) - (C)

CHINA'S TOP PERFORMING CODING MODELS BLENDED PRICE: US\$1 per blended 1M tokens



IN ABSOLUTE US\$ BILLIONS

	Current (Year-end 26E)			Long-term	
	China's value-for-money agentic model	China's best performing coding model	Hypothetical economics for 'most frontier' SOTA model	China's value-for-money agentic model	China's best performing coding model
ARR, US\$	\$0.8bn	\$1.5bn	\$55bn	\$10.9bn	\$30.6bn
Inference cost as % of ARR	-74%	-70%	-26%	-57%	-51%
Gross Profit (ARR - inference cost)	\$0.1bn	\$0.3bn	\$39bn	\$3.6bn	\$12.2bn
% Gross margin	10-15%	15-20%	70%	30-40%	30-40%
Training as % of ARR	-30%	-40%	-49%	-13%	-13%
EBIT (Gross profit - Training cost & OpEx)	-\$0.2bn	-\$0.6bn	-\$0.1bn	\$1.5bn	\$6.7bn
% EBIT margin	-30%	-39%	0%	14%	22%

We estimate aggregate of US\$10bn ARR for China AI model industry by year-end FY26E; we estimate to reach US\$125bn for FY30E

US\$7bn inference costs for China AI model industry in FY26E, we estimate US\$55bn by FY30E, which translates into hyperscalers'/neo-clouds' revenues

US\$4bn training costs for China AI model industry in FY26E, US\$20bn by FY30E, as model sizes continue to scale up

Negative EBIT margin for China AI model companies in FY26E; we estimate to reach 18% EBIT margin/US\$23bn profit pool by FY30E

资料来源：公司数据，高盛全球投资研究部

PM Summary

From DeepSeek's moment last year (on cost efficiency) to Zhipu's GLM moment this year (on model intelligence)

China's AI open-source/open-weight models are reaching a critical point of intelligence performance vs. global proprietary models, with a significant ramp up in domestic enterprise adoption and a proliferation of global consumer and SME demand. In this report, we evaluate How these models achieve such performance at low costs/tight computing resources; Why they pursue an open source/open weight approach and how they monetize; What the key addressable markets are, where we highlight two favoured ARR quadrants and risk factors; and Who the potential long-term winners will be under our Competitive Positioning framework.

We believe Chinese models are reaching a critical 'good enough' stage for agentic tasks/specific coding scenarios, and rising fragmentation in China's AI model landscape (but the strong will get stronger). While pricing power remains strong for models with frontier performance/multi-modal, the lower-end segment is in a price war; nevertheless, Agentic AI is driving explosive demand for these value-for-money models at the lower-end. Access to computing will be a swing factor, where US/China regulations, balance sheet and inference efficiency are key.

Accordingly, we introduce our Chinese AI model Competitive Positioning framework based on pricing power, cost advantage and financial strength, overlaying with token scale and market share progressions, and identify Knowledge Atlas (Zhipu, initiation) and DeepSeek (private) as the strongest positioned in foundation models, and Bytedance (private) in multi-modal.

- How do Chinese models achieve competitive performance at low costs/tight computing resources? By leveraging smaller-sized parameter models for equivalent benchmark performance, and Mixture-of-Expert and new architectural innovations. Chinese AI models are bifurcating into a two-tiered market where performance and time to market are key to pricing power, thereby leading to two 'ARR maximising' quadrants based on token adoption and pricing. A positive flywheel is taking effect for top Chinese AI models driven by increasing actual real world coding adoption, and reducing their reliance on model distillation practices.
- Why are Chinese models pursuing an open source/open weight approach, and ways to monetize? Open source allows for greater flexibility in model training/deployment, and allows for the widest adoption and an open community. We believe open source models' disclosed ARR are likely understating total deployment and revenue potentials, and we expect more shifts to open weight (with Community License, i.e., commercial terms if for commercial use) among Chinese AI models down the road.
- What are the key addressable markets, domestically and internationally, and key risks? We highlight two favoured

ARR quadrants and estimate China/China AI models market to see token growth of 25X by 2030E; the coding landscape to consolidate while the agentic/low-end segment could remain fragmented. International (going global) should be a key upside, with the potential for higher pricing and global proliferation, especially in non-US markets as global enterprises increasingly pivot from a token-maxxing to a ROI-first model that prioritizes clear task boundaries, number of agents per day, back-end process automation and actual output over pure computational token volume. Key risk factors are market access/anti-distillation and regulations, including any stricter access to China's future most frontier models for overseas markets, access to highest-end leased computing equipment (used for training), market access to western markets, further restriction lists/entity list designations (but this could be positive for China's path to further AI self-sufficiency across software/CPU/ASICs), and competition from SLM/threat from AI architectures.

- Who are best positioned to be the long term winners within China's AI model companies? We expect players with the largest ARR scale with a gross margin advantage + financial strength to be the long-term winners. We highlight independent AI model companies mostly stand out in our Competitive Positioning framework in pricing power + cost advantage (in aggregate represent over US\$200bn in implied valuations, based on latest market cap/funding rounds), where Zhipu and DeepSeek are the most strongly positioned in text based foundation models, while ByteDance leads in multi-modal capabilities based on our framework.

Signposts we are looking for into 2H 2026

- Harness/agentic applications: We expect China AI model companies to increasingly focus on positioning their harness/agentic applications as key entry points, especially in coding (e.g. Zhipu's ZCode, Tencent's Workbuddy, and Alibaba's Qoder which are aggregate platforms that support the full suite of AI models) as model companies attempt to close the loop in capturing more real-life coding and agentic data for scaling. Co-work and industry expert agentic products could be the next priorities. Enterprises will increasingly focus on overall cost per task instead of headline pricing per token, with an openness to using different models (multiple-models approach).
- Multiple large parameter high-end coding model launches and potentially stricter access to the most advanced Chinese AI models outside of China: We anticipate multiple new Chinese AI model launches over 2H26 with significantly larger total parameter sizes of 2-5 trillion across Chinese AI model players. Coding/programming segment competition will intensify as Chinese models try to challenge Zhipu GLM's leadership via training on high-quality real-life coding data (where available) and scaling to larger parameter model sizes. We also expect a potential shift from an open-source to an open-weight approach for best-performing models (i.e. from free-for-all use cases to requiring revenue sharing/a take rate for commercial use). The addition of multi-modal/visual understanding will be the next upgrades amongst Chinese AI foundation models (e.g. for GLM, DeepSeek), while MiniMax M3 already excels in these given the multi-modal focus of MiniMax from the start. That said, we note press

reports on potential future restrictions on overseas access to China's most advanced AI models (both closed and open source if at the frontier level) as cutting-edge artificial intelligence is increasingly being seen as a critical national asset.

- Continued suppressed API pricing for the lower end agentic-focused segment: While DeepSeek recently announced an increase in peak-hour pricing from mid-July, we expect API pricing and therefore gross margins to remain under pressure at the lower-end pricing segment (around US\$0.1-0.2 per 1M tokens) into the second half, as China's AI model players have significant cash buffers post-fund raising to subsidize competitive pricing at zero/negative gross margins in the near term. As a result, we see financial strength as one of the three most important metrics (alongside pricing power and cost efficiencies) in assessing a Chinese AI model company, where cash on hand, net cash as % of assets and valuation multiples will be the critical financial strength metrics for long-term success. This said, we are Buy-rated on MiniMax as the company stands out on cost efficiency/cost advantage metrics under our Competitive Positioning framework. With its M3 model well positioned in the favored ARR maximizing quadrant (attractive pricing + high token volumes), alongside its discounted valuation at 13X P/2026E year-end ARR (vs. China/global peers which command multiples several times higher at similar ARR stage), we believe risk-reward is skewed to the upside and reiterate our Buy rating. Key swing factors for MiniMax to improve its overall competitive positioning will hinge on its pricing power and financial strength. We expect its H3 video generation model performance (imminent launch in a more favorable industry landscape vs. text models) and time-to-market for its next M3 updates (focusing on further coding intelligence level uplift from post-training/reinforcement learning, we estimate over July-Aug, and a larger parameter size M3 model later in 2H 2026) will be the next key drivers.
- Multi-modal/video-generation models to see further ARR ramp-up from global adoption: We expect continued healthy industry pricing and gross margins within video generation (unlike foundation text) where key players ByteDance's SeeDance, Kuaishou's (Buy) Kling and MiniMax's Hailuo/upcoming H3 models to enjoy healthy growth over 2H 2026 amid new functionality breakthroughs (combination of video-generation with LLM) and tight computing resources where demand significantly outpaces capacity. According to China news reports like LatePost and 36Kr, ByteDance's Seedance gross margins have been at a healthy 70% at its latest US\$2bn+ ARR run-rate.
- We also expect increasing domestic ASICs supply, tighter access to overseas computing resources and potential market access limitations (could mirror TikTok's trajectory where rapid expansion in western markets was followed by more regulations/focus on ensuring data security where computing will have to be conducted within local jurisdictions).

Our assessment of AI model companies: ARR scale x gross margin advantage + financial strength

- Largest ARR scale (Token scale x pricing power)

- Gross margin advantage (Training & Inference efficiency, Technology)
- Financial strength (Balance sheet, Access to computing)

Accordingly, we lay out our Competitive Positioning framework for AI model players based on quantifiable metrics, 1) Pricing Power (amongst which, based on Time-to-market of model launch, Arena score based on actual usage cases and pricing level), 2) Cost advantage (based on token volume scale, throughput/cache hit rate, parameter sizes/activation ratio and our estimate of inference gross margins), and 3) Financial strength (based on cash on-hand, net cash as % of assets, and valuation multiples).

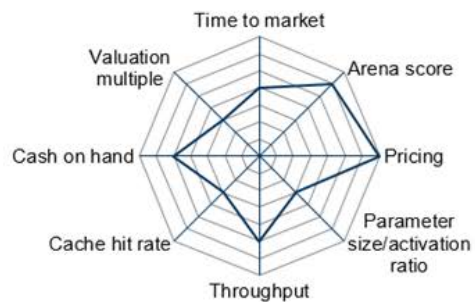
图表 2: Competitive positioning framework for AI model players

Key themes	Chinese models reaching critical stage of intelligence for wide adoption across agentic tasks/specific coding scenarios			Fragmentation between China AI models but strong to get stronger; pricing power to come from frontier performance/multi-modal			Agentic AI driving higher demand for value-for-money models/SLMs		Access to computing, where balance sheet and inference efficiency are key	
Drivers of corporate performance	Competitive positioning									
CP category	Pricing Power			Cost advantage				Financial strength		
CP metrics	Time to market	Arena score	Pricing	Token volume	Throughput /Cache hit rate	Parameter size/ activation ratio	Inference GPM	Cash on hand	Net cash as % of asset	Valuation multiples

资料来源：高盛全球投资研究部

图表 3: Overview of competitive analysis for key LLM labs' positioning

Alibaba Qwen: Full-stack AI player with multi-year AI/cloud upside, leading position of most areas



Tencent Hunyuan: Strong contender after AI strategic revamp; potential in Weixin AI agent



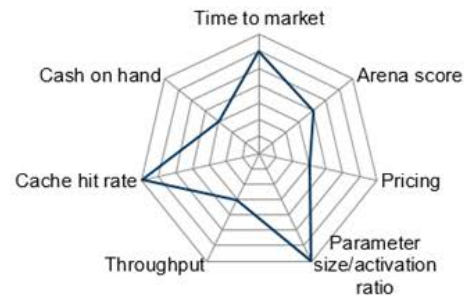
MiniMax M-series: Unique full-modal capabilities among independent players; cost-efficient computing architectures



Zhipu GLM: Leading model capabilities in coding/agentic scenarios



DeepSeek: Extreme cost efficiency to sustain scalable token usage on competitive pricing



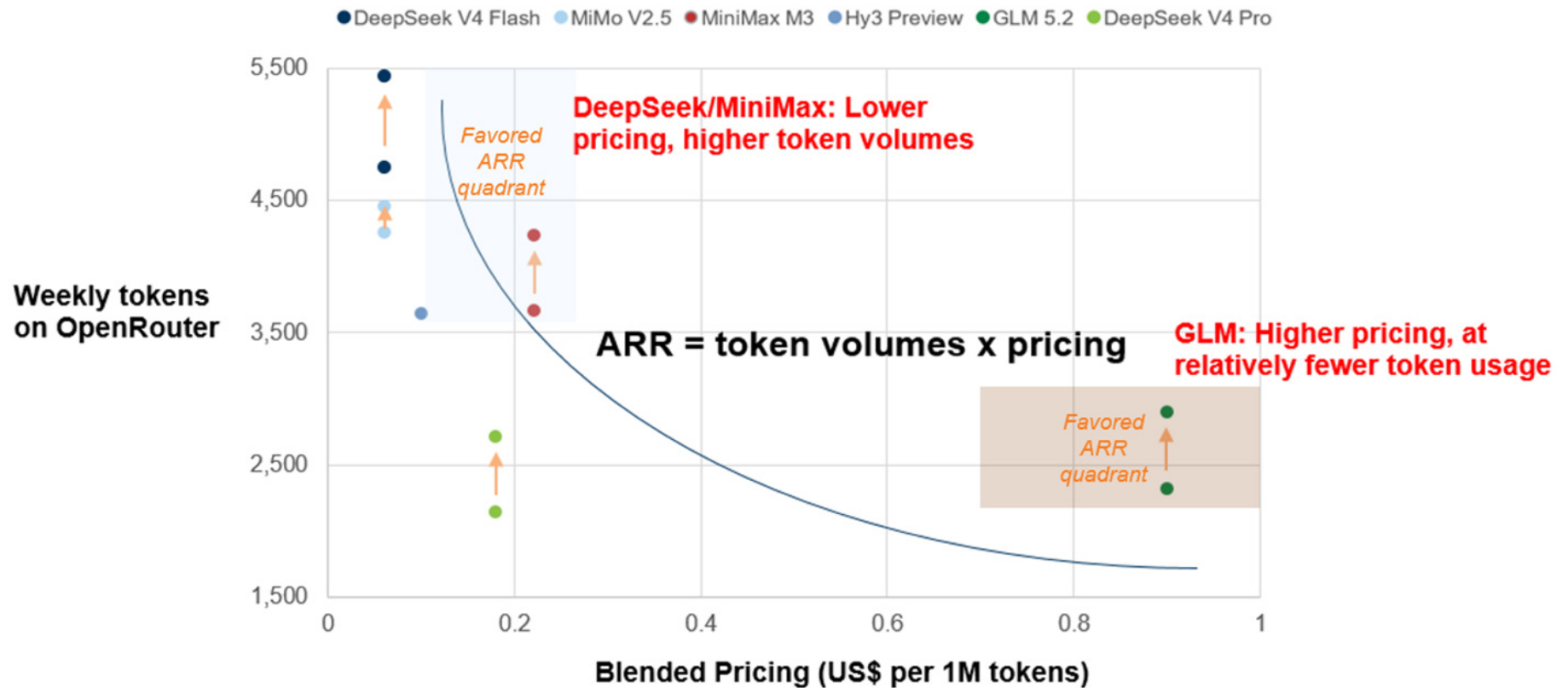
ByteDance Seedance: leading in model performance with strong financial/ecosystem back-up for quick adoption & commercialization



N/A for DeepSeek/ByteDance valuation multiple

资料来源: 高盛全球投资研究部

图表 4: We identify two favorable ARR quadrants (in maximizing ARR) which is a combination of maximizing token volumes and/or pricing level



资料来源：公司数据, OpenRouter, 高盛全球投资研究部

Related research

[Americas Technology: Decoding the Agentic Economy - The Coming Inflection in AI Usage and Margins](#), 5 May 2026, by James Schneider, Eric Sheridan and team

[Global Server: Raising both AI and General servers; better mix and higher memory costs drive CSP spending](#), 24 June 2026, by Allen Chang and team

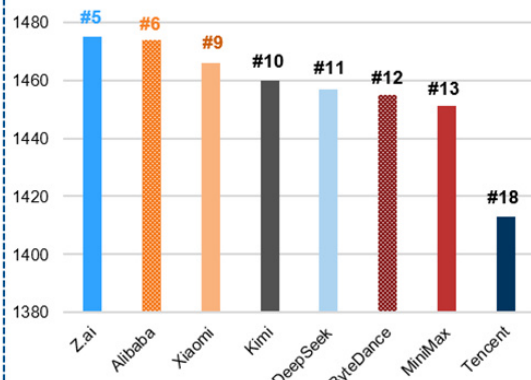
[Knowledge Atlas Technology \(2513.HK\): Initiate with Neutral at US\\$110bn valuation; China's top enterprise/coding AI model](#), July 10, 2026

[MiniMax Group \(0100.HK\): Key focuses post-share unlock; upside skew to risk-reward; Buy](#), July 10, 2026

Comparing China's key LLM players

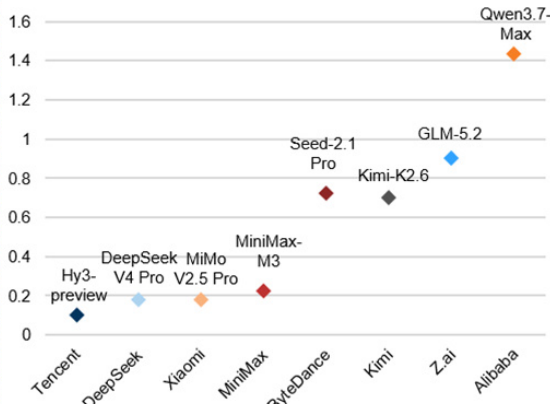
Performance: China LLM labs keep climbing up the Text Arena ranking, with flagship models hitting new high

Ranking by company (total 54) on LMArena, Text Arena
Dot for closed-source models, else open-source models



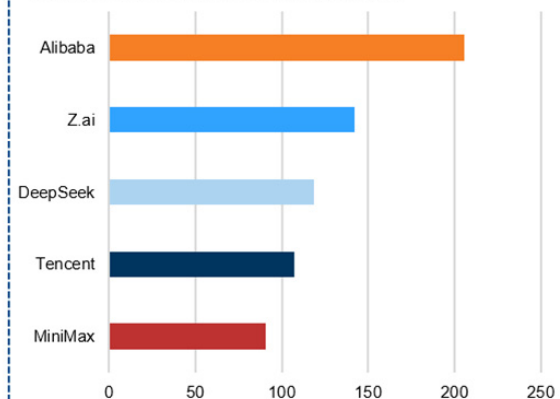
Two-tier price: premium pricing on model intelligence vs. low cost for price-sensitive demands

Blended price per 1M tokens of flagship models



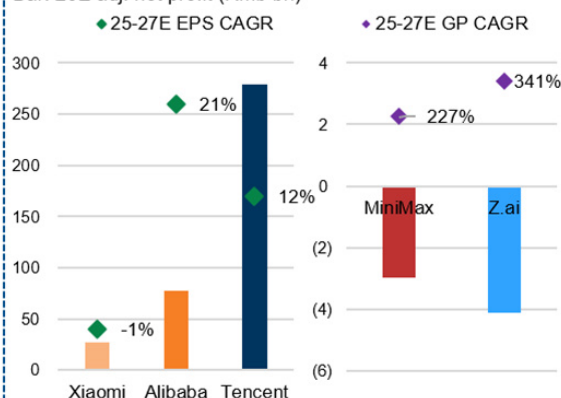
Speed: first-party API performance to reflect inference efficiency and infra capabilities

Output tokens per second of flagship models



P&L: mega-caps AI funding from stable profits while independents still scaling in commercialization

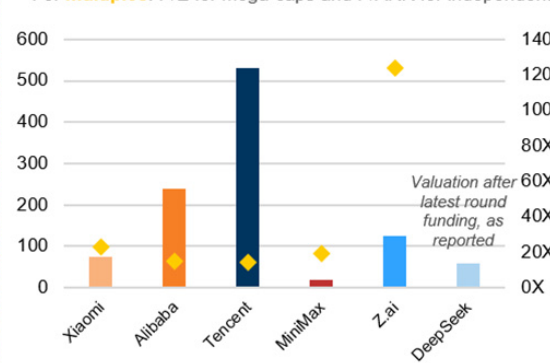
Bar: 26E adj. net profit (Rmb bn)



Valuation: latest market caps and reported private valuation for LLM players

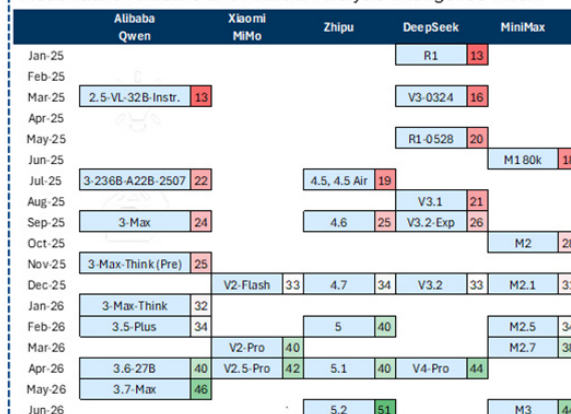
■ Market caps (US\$ bn) ◆ Valuation

For multiples: P/E for mega-caps and P/ARR for independents



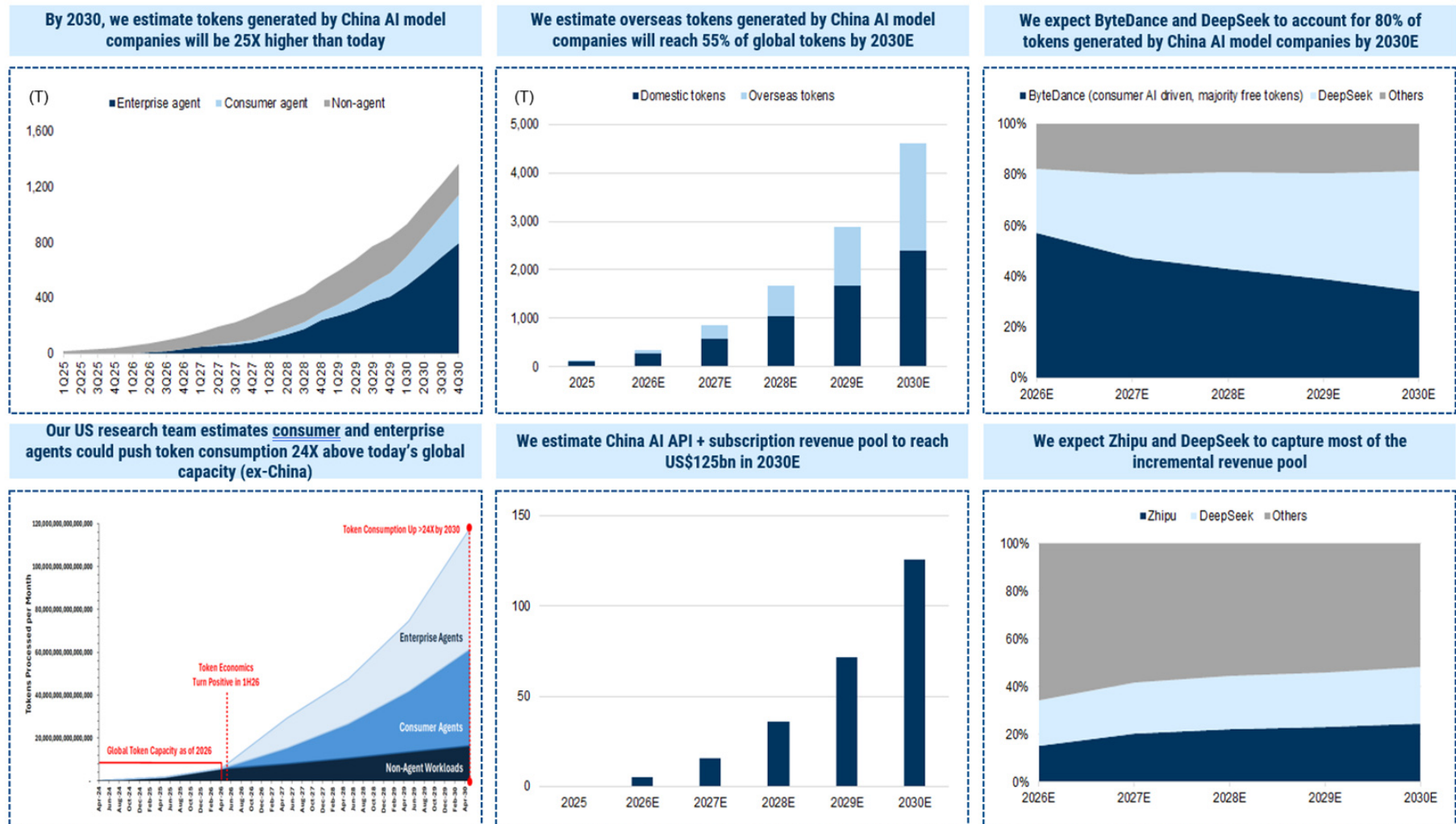
Model updates into 2026: faster iteration with rising intelligence level

Model launch timeline and Artificial Analysis Intelligence Index



资料来源：公司数据，高盛全球投资研究部，WSJ, Artificial Analysis

Decoding China's AI model tokens & our forecasts on token/revenue share



DeepSeek/ByteDance/Zhipu's (Not Covered) relevant metrics (e.g. token consumption/revenue) are calculated by analyzing the industry and companies that we cover, and then extrapolating them to DeepSeek, ByteDance and Zhipu.

资料来源：高盛全球投资研究部

How do Chinese models achieve competitive performance at low costs/tight computing resources?

Smaller-sized parameters models for equivalent benchmark performances, Mixture-of-Expert and new architectural innovations

Compared with a year ago, Chinese models' coding and agentic capabilities have reached a critical level in being able to complete more coding and autonomous agentic tasks with higher success rates as context windows have been expanded to 1 million tokens. The smaller parameter model sizes of Chinese models (spanning from 200bn to 1.6T parameters, at 2-10% of leading SOTA models, due to constrained access to high-end computing), and highly efficient architectures (MoE, Sparse Attention, OCR etc., at 3-5% activated parameters only vs. total parameter sizes) all contribute to the much lower training and inference costs for Chinese models vs. leading US models. We would attribute the recent step improvement of Chinese models in coding according to Arena.ai to data curation, distillation techniques and reinforcement learning post training, despite their relatively small parameter model sizes (1.6T for DeepSeek V4 Pro, 0.7T for Zhipu's GLM5.2 and 0.4T for MiniMax's M3). On June 27, DeepSeek introduced DSpark, a speculative decoding framework that makes existing DeepSeek-V4 models serve faster. DSpark has already been deployed in DeepSeek-V4 Flash / Pro online serving, improving per-user DeepSeek-V4 generation 60-85% faster on V4-Flash and 57-78% faster on V4 Pro without changing the model's weights or output quality.

Chinese AI models bifurcating into two-tiered market (where performance and time to market are key to pricing power), with two 'ARR maximising' quadrants based on token adoption and pricing

We are seeing pricing power for the highest performing Chinese AI models, e.g. Zhipu's GLM5.2 model and Alibaba's Qwen3.7 Max models at around US\$1 per blended 1M tokens, at 5X that of low-end Chinese AI models. The reported tighter US processes in allowing most SOTA model access has also opened new arenas for China's top performing coding models for enterprises and SMEs. Smaller parameter and activated ratios allow China's top performing models to be priced at US\$1, 10-25% vs. US SOTA models at US\$4-8 per blended 1M tokens, while generating double digit 10-20% gross margins (GSe) that are lower than global SOTA models due to relatively lower pricing power.

In the lower end segment, agentic focused models are priced at US\$0.06-0.2 per blended 1M tokens, which are enabling these models to tap into new global TAMs for price sensitive SMEs and one-man companies. MiniMax generates 60-70% revenues from overseas.

We also note DeepSeek announced that its V4 official version is set for launch in mid-July, alongside the introduction of peak/off-peak API pricing to better allocate resources and improve service stability. V4 Pro/Flash non-peak pricing

remains unchanged, while peak hour (9am-12pm/2pm-6pm China time) will be charged at 2X non-peak rates, implying blended pricing of US\$0.35/US\$0.12 per 1M tokens due to strong Chinese AI model demand that is increasingly causing significant compute tightness in work/productivity scenarios.

Positive flywheel is taking effect for top Chinese AI models from increasing actual real world coding adoption, with less reliance on model distillation ahead

Compared with learning and distillation tactics from global SOTA models in the past, we believe top Chinese models like GLM5 are reaching a critical stage of adoption by China's major enterprises and global users. As per LatePost, AI-generated code has increased to as high as 90% at some China mega-cap companies, up from 20-30% in 2H25, which we believe will enable a positive data flywheel effect of further improvements via reinforcement learning with actual user data (both successful and unsuccessful coding cases) and post training. We think these have underpinned the step improvements in GLM5.2 from GLM5.1 in just over the course of a few months in 2026, and we expect to see further step improvements to Chinese AI models over the next 6-12 months.

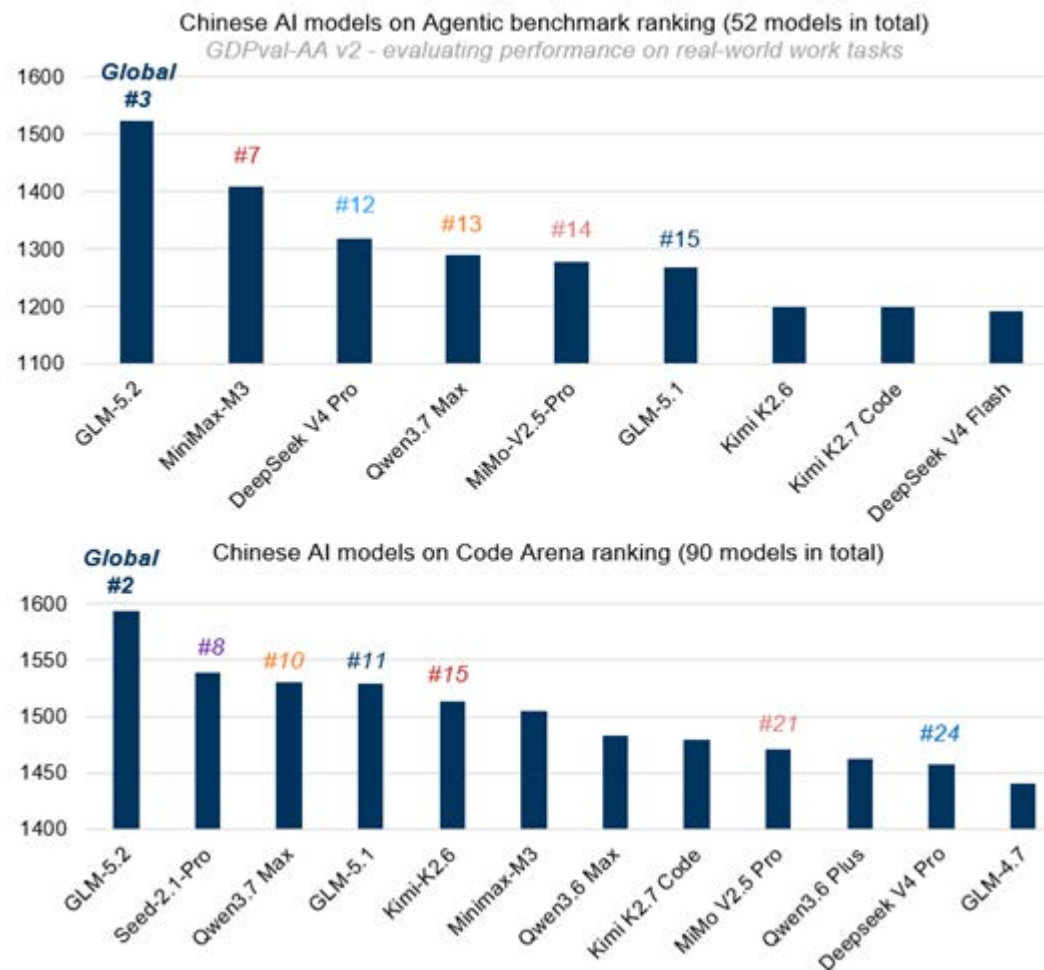
图表 5: China’s model count is rising in top rank bands across usage (text, coding and agent)
LMArena ranking of Text, WebDev (Coding) and Agent Arena



Baseline date for Text/WebDev (Coding) Arena ranking = 3/18/2026, and current date refers to latest date available, i.e. 6/16/2026 & 6/20/2026. Agent Arena ranking started at June 5, 2026, with current date = 6/30/2026

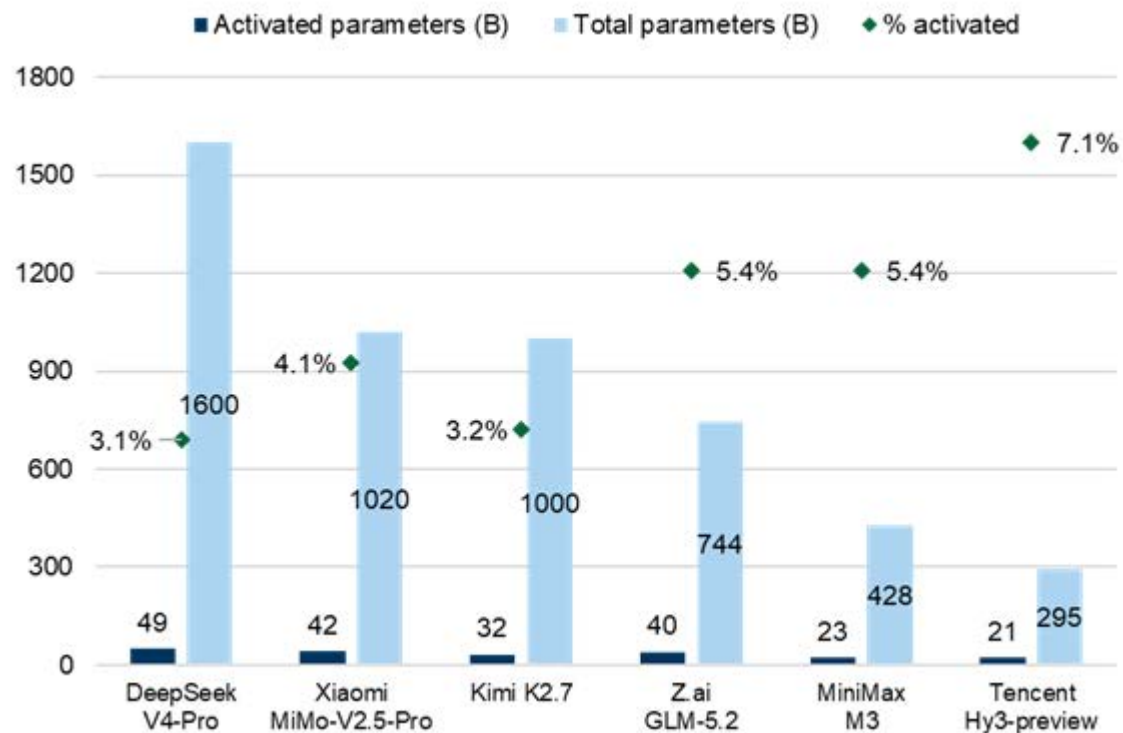
资料来源：LMArena, Data compiled by Goldman Sachs Global Investment Research

图表 6: Notably in coding/agent tasks, Chinese players are reaching global top-tier positions
As of Jun 20, 2026



资料来源: GDPval-AA, LMArena, Data compiled by Goldman Sachs Global Investment Research

图表 7: Compared with global SOTA leading models of several tens of trillions, China open source models are mostly around or below 1 trillion parameter in size, and adopt a MoE structure, with low activated to total parameter ratios for higher inference efficiency



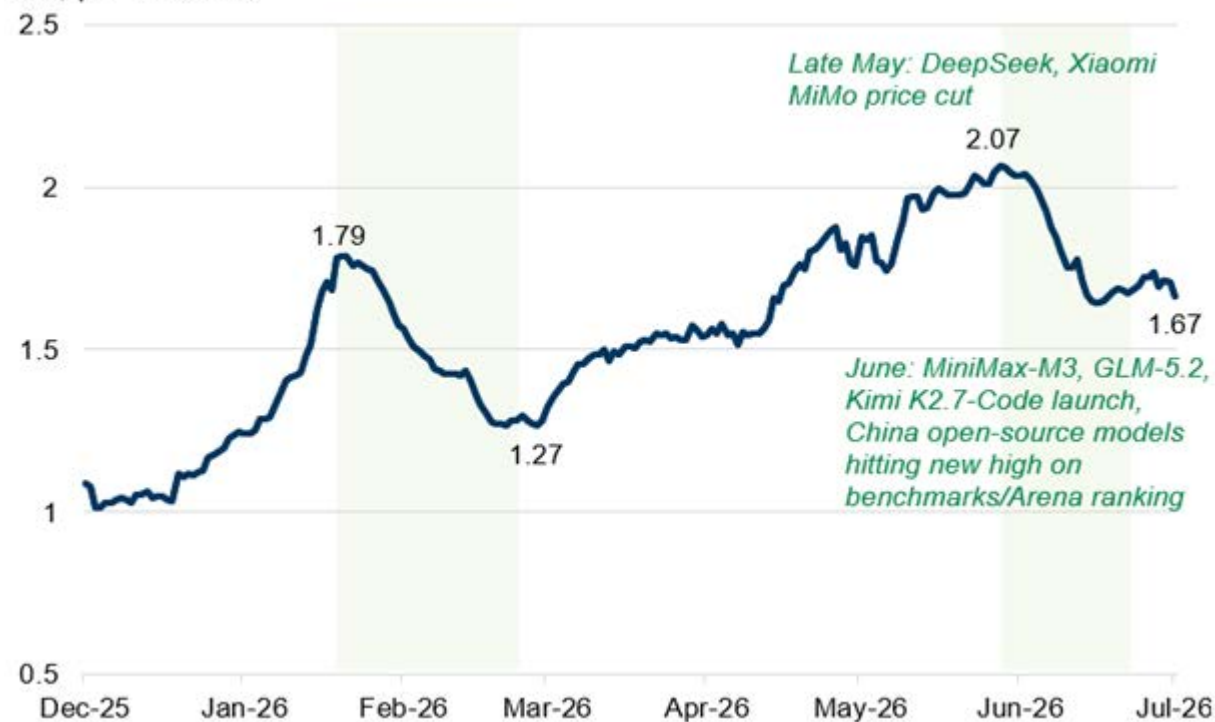
A Mixture-of-Experts (MoE) structure is a sparsely activated neural architecture, where it dynamically selects a small subset of expert networks to process each input, allowing the model to increase its total capacity without activating all parameters during inference.

资料来源：公司数据, Data compiled by Goldman Sachs Global Investment Research

图表 8: Blended LLM token price (SDLLMTK) has been declining since early June, potentially driven by rising adoption of cost effective China models

US\$ per 1M tokens

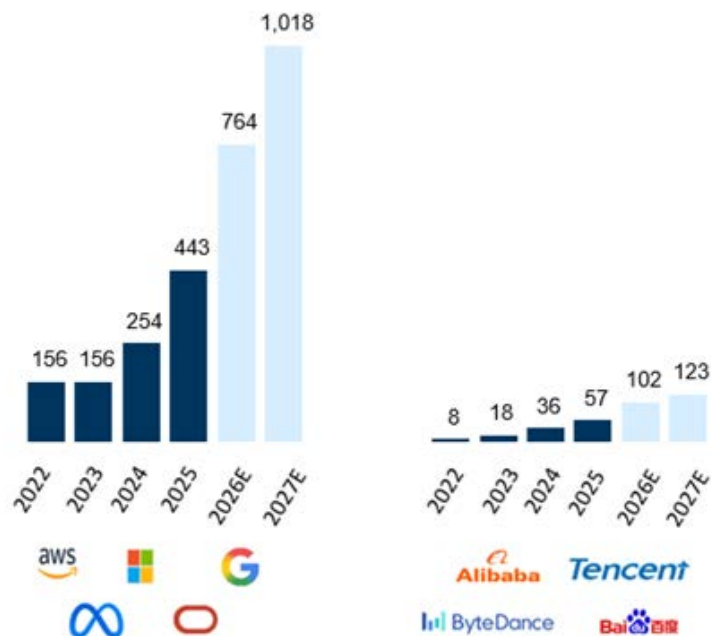
US\$ per 1M tokens



The Silicon Data LLM Token Expenditure Index (SDLLMTK) is a daily blended measure of what the market actually pays to run LLM inference across frontier and open-weight models.

资料来源: Silicon Data, 彭博, Data compiled by Goldman Sachs Global Investment Research

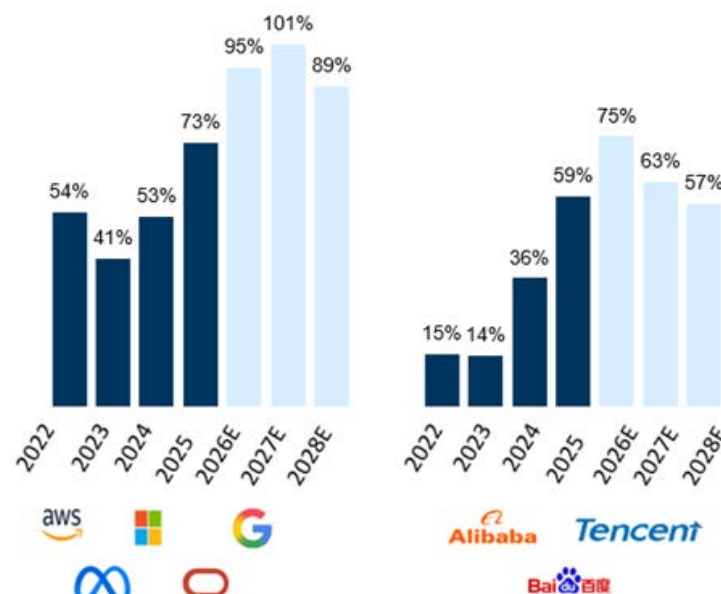
图表 9: Historical and projected capex for major US & China cloud service providers
US\$bn



We calculate ByteDance's (Not Covered) relevant operating metrics by analyzing the industry and companies that we cover, and then extrapolating them to ByteDance.

资料来源：公司数据, 高盛全球投资研究部

图表 10: Capex to operating cash flow ratio is still healthy for China hyperscalers



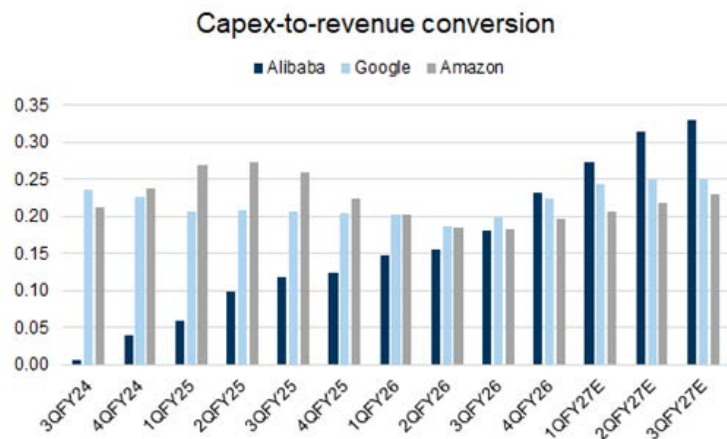
资料来源：公司数据, 高盛全球投资研究部

Case study on Meituan's LongCat 2.0: A milestone for China's domestic AI infrastructure

Released on June 30, 2026, we believe Meituan's LongCat 2.0 marks a major milestone as China's first official 1.6 trillion-parameter open-source Mixture-of-Experts (MoE) model trained and deployed entirely on a 50,000-card domestic compute cluster. Purpose-built for agentic coding and complex software engineering workflows, the model features a native 1-million-token context window enabled by LongCat Sparse Attention (LSA) and dynamically activates an average of 48 billion parameters per token to optimize inference costs. Implications for a more self-sustainable China AI model outlook that is less reliant on foreign high-end chips for model training: The successful end-to-end pre-training and inference of a trillion-parameter class model on Chinese silicon (reportedly utilizing Huawei Atlas-950 SuperPods) fundamentally drives a more sustainable China's AI model development outlook, in our view. While previous Chinese flagship models, such as DeepSeek V4-pro, mentioned domestic chips for inference, LongCat 2.0's ability to overcome

critical memory bottlenecks and distributed stability challenges during the compute-heavy pre-training phase proves the viability of a wider and more localized hardware stack for AI model training in the future.

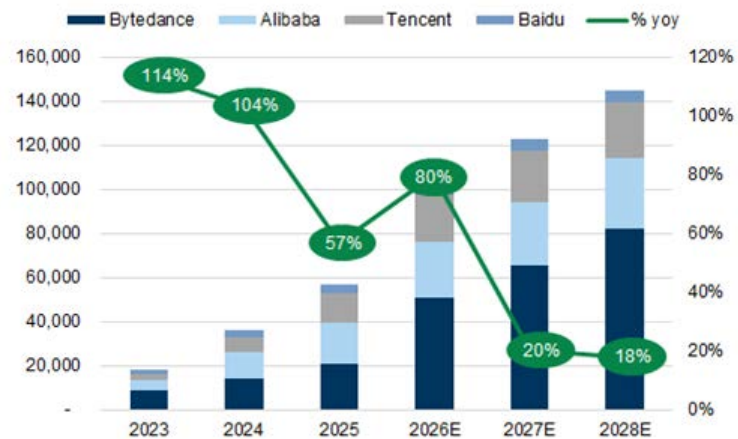
图表 11: Alibaba capex-to-revenue conversion ratio vs. Amazon and Google
Rolling 12m incremental cloud revenue/rolling 12m capex (lagged by 1 quarter)



We benchmark Alibaba external cloud revenue growth to Google and Amazon Cloud (lagged c.2 years)

资料来源: 公司数据, 高盛全球投资研究部

图表 12: China Internet yearly capex forecast
US\$mn



For ByteDance (Not Covered), we impute capex based on industry trends.

资料来源: 公司数据, 高盛全球投资研究部, 彭博

Why are Chinese models pursuing an open source/open weight approach, and ways to monetize?

Open source allows for higher flexibility in model training/deployment, and allows for the widest adoption and an open community

Alibaba's Qwen model family has long pursued an open source approach (before adopting a closed source for its largest and highest performing Qwen-Max models for better monetization), while other key Chinese AI model players have mostly pursued an open source/open weight approach including DeepSeek, Zhipu's GLM and MiniMax M3 series models with the exception of ByteDance's full closed proprietary approach for its Seed model. We believe the open source approach allows for the highest flexibility in terms of the locations of where models are trained (and thus where the models can be deployed both inside and outside of Mainland China after training). An open source approach also allows for the highest adoption amongst the AI community with full transparency of the model parameters/architecture for trust, and an open community in driving more user feedback and thus model iterations/improvements. The open source approach vs. world's leading closed/proprietary approach also provides an alternative choice for worldwide AI users when the best performing closed models have stricter user access and higher costs from their premium pricing, especially at a time when 'token-maxxing' has become a key cost consideration for many corporates.

Open source models' disclosed ARR are likely understating total deployment and revenue potentials

While open source model companies offer their own coding plans and a chargeable open platform API channel (where the model companies conduct their own model inference), the majority of open source models also allow individuals/third-party hyperscalers/neocloud providers to deploy the models without a charge even for commercial use (e.g. Alibaba Cloud's Bailian MaaS platform can house GLM5.2 open source model without needing to pay a fee/take rate to Zhipu). As a result, while Zhipu's last stated ARR target for year-end 2026 is at US\$1bn, we believe the actual deployment of GLM5.2 model worldwide is and will be multiple-fold higher vs. Zhipu's own API channel token volumes and revenue. There is also increasing post training of Chinese AI models that are re-branded by global companies (e.g. Composer 2 etc.) where the Chinese AI models do not necessarily receive any revenues given the open source spirit.

We expect more shifts to open weight (with Community License) approach among Chinese AI models down the road

While Zhipu's open source GLM model's MIT license allows for free for all use cases (regardless of revenue), MiniMax M series models have pursued a restricted license (where the industry terms it as an open weight with Community License model), which requires MiniMax's agreement and commercial terms (e.g. revenue sharing/a take rate) on

commercial use. We believe this will be the likely next path for other open source models in the Chinese AI model industry, in order for eventual gross profits of inference tokens to cover training costs and for each AI model company to achieve a sustainable path to returns.

图表 13: Comparison of different open source licenses

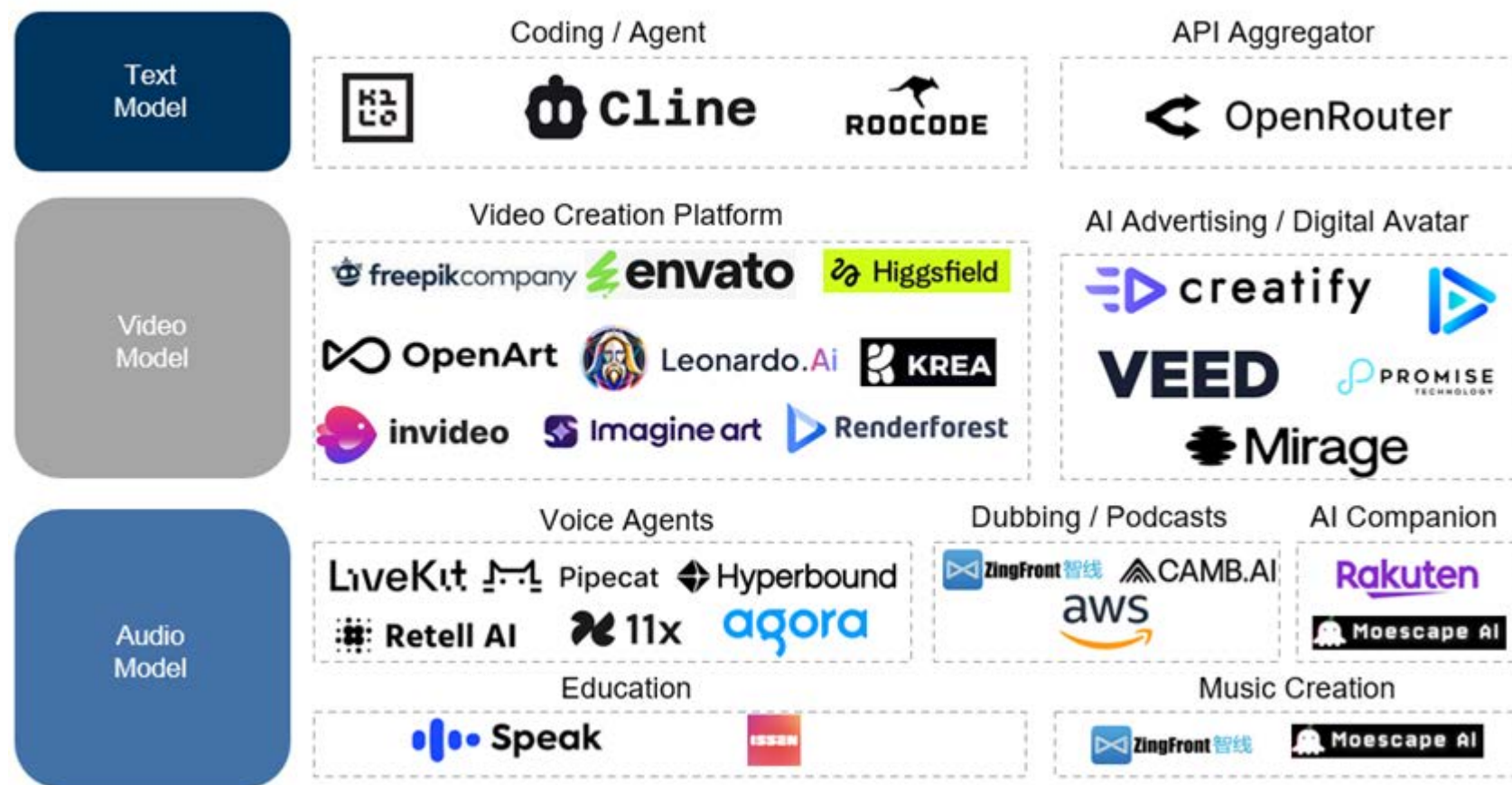
Feature	Zhipu GLM 5.2	MiniMax M3
License Type	MIT License (Permissive)	MiniMax Community License (Restricted)
Commercial Use	Unrestricted. Free for all users regardless of revenue	Restricted. Free only for entities with <US\$20M annual revenue
Approval Process	None. No permission needed to deploy commercially	Required. Must email api@minimax.io . <US\$20M revenue needs notice; >\$20M requires formal written license
Attribution	Standard copyright notice retention (MIT terms)	Mandatory "Built with MiniMax" label on products/UI
Prohibitions	Standard legal limitations apply	Explicit restrictions on military use, hate speech, and illegal content

资料来源：公司数据, Data compiled by Goldman Sachs Global Investment Research

What are API platform services?

Chinese AI models typically monetize via subscription plans or their in-house API platform services. API access is typically offered on third party platforms like OpenRouter, Cline or coding agent platforms. As we expect more open source models could shift to open weight, we believe the unit economics of Chinese AI model companies could improve as not only can official API channels be monetized, but also by revenue sharing agreements when these open weight models are deployed and sold on third-party neocloud/hyperscalers like AWS Bedrock, Alibaba Cloud Bailian. Such arrangements could become highly margin accretive as AI model companies will not need to conduct inference/bear computing costs themselves.

图表 14: API platform international enterprise customers through which customers can access Chinese AI models



资料来源：公司数据, Data compiled by Goldman Sachs Global Investment Research

What are the key addressable markets, domestically and internationally, and key risks?

We estimate China AI models' aggregate API+subscription revenue to increase from Rmb35bn in 2026E to Rmb879bn by 2030E from rising model intelligence, in particular with recent models like GLM5.2 reaching a critical point for global adoption and attractive pricing. Our revenue pool estimates for China AI models imply total Chinese AI model daily token consumption of 350T in 2026E to increase to 4,600T by 2030E.

We estimate domestic market to see token growth of 25X by 2030E; coding landscape to consolidate while agentic/low-end segment could remain fragmented

At 140tn daily token volumes for the country in March 2026 and several hundred trillions by June 2026 as per the National Bureau of Statistics, we estimate open source/open weight models have roughly a 30% token share (vs. 70% token market share by ByteDance alone, which is closed source and mainly driven by its Doubao app enterprise and individual user base, as the #1 used AI chatbot in China). We expect similar to the US, the coding segment (at a premium pricing level) will continue to be dominated by SOTA best performing models. Meanwhile, we expect the lower-end segment focused on agentic AI will remain fragmented with multiple players due to the financial strength of AI model companies/mega-caps which would sustain the lower-end segment price war for longer.

International (going global) to be the key upside; with potential for higher pricing and global proliferation, especially in non-US markets

Our US research team estimates ([link](#)) agentic AI will drive 24X growth in token consumption by 2030 (from 2026) to 120 quadrillion tokens per month (or 4 quadrillion tokens per day, from their estimate of 170 trillion daily tokens today), with the biggest driver at 55X from enterprise agents and 12X from consumer agents. The global (ex. China) landscape has seen significant token share gains from Chinese AI models, as rising model intelligence and attractive token costs have driven higher adoption across 24/7 Hermes/Claw/Productivity agents, and shifting global SME mindset on using Chinese models in managing token costs given Chinese models have reached a 'good enough' stage in terms of intelligence/performance.

Pivoting from 'token-maxxing' to ROI-focused metrics, e.g. Daily Active Agents/Agentic Work Units

We believe the AI token proliferation is undergoing a paradigm shift from 'token-maxxing' (an initial focus from late 2025 to early 2026 where enterprises equated high AI token consumption directly with organization productivity) towards an 'ROI-first' model that prioritizes clear task boundaries and output over raw computational volume.

- 'Token-maxxing' has been attributed to corporate inefficiencies and cost overruns: Data from a Jellyfish AI

Engineering trends study indicated heavy AI users at enterprises consumed 10X more tokens but only had a 2X increase in output. Meanwhile, we note multiple US Internet companies have commented back in April 2026 that either their engineering teams utilized a full year's AI budget in just four months using agentic products (promoting stricter monthly caps per tool) or changed their internal token utilization leaderboards that previously wrongly incentivized staff to launch inefficient/low value autonomous agent tasks.

- Enterprise framework is pivoting towards an ROI-first model that prioritizes clear task boundaries, number of agents per day, backend process automation and actual output over pure computational token volume. Besides a marked increase in adoption of Chinese AI models, global enterprises have been downgrading default models to cheaper flash models for more typical tasks (while reserving SOTA models for only the top most value creating tasks like coding/programming). We see a transition from just token tracking to metrics like Daily Active Agents (DAA) and Agentic Work Units (AWU), and believe the overall cost per task will become more relevant than price per token metrics.

Open source/open weight approach allows for the option for U.S. hyperscalers to host Chinese models, operated within U.S. cloud ecosystem

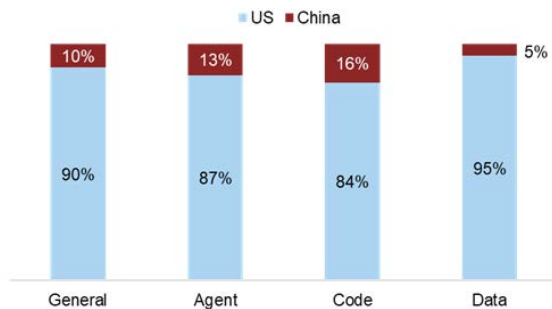
Alphabet and Amazon's (both covered by Eric Sheridan) respective cloud services Gemini Enterprise Agent Platform (via. its Model Garden) and AWS Bedrock already offer a broad selection of Chinese AI models including DeepSeek, MiniMax, Moonshot, GLM and Qwen which are fully managed by the US hyperscalers. Besides the model layer, into applications, worthy of note is Microsoft (covered by Gabriela Borges) CEO's recent remarks at a Wall Street Journal interview ([link](#)) where he noted Microsoft is considering hosting versions of DeepSeek on Copilot as an optional, cost-effective model which could give its customers access to cheaper choices alongside US proprietary models. Microsoft indicated that if it hosts DeepSeek, the model would operate within its cloud ecosystem, ensuring customer data stays inside Azure.

Noting key risks around any potential tighter geopolitical policies given Chinese AI models inroads into western markets.

Key risks to the 'going global' opportunity will hinge on end market access (especially in western countries, and with a focus on where computing is done/data is stored), access to high-end computing for AI model training (that could impact iteration pace and cost structure of Chinese models), and restrictions on Chinese AI model companies could impact supplier relationships/access to US companies and/or access to US capital. More in the key risks section.

图表 15: China AI models now only account for limited share of spend across different tasks, per OpenRouter

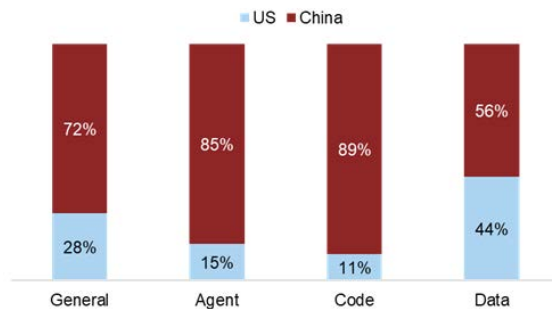
As of week starting Jul 1, 2026



资料来源: OpenRouter, Data compiled by Goldman Sachs Global Investment Research

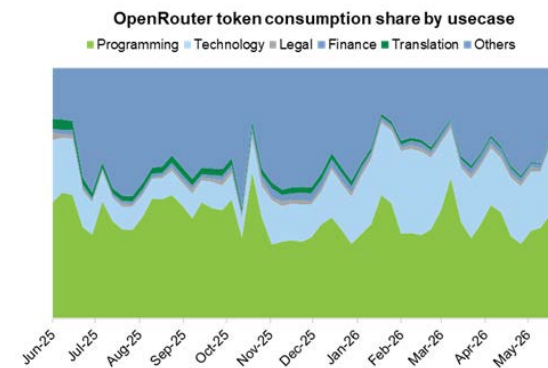
图表 16: While China AI models already capture majority of token usage on price competitiveness, esp. in long duration coding/agentic tasks

As of week starting Jul 1, 2026



资料来源: OpenRouter, Data compiled by Goldman Sachs Global Investment Research

图表 17: Token consumption is concentrating in complex, high value scenarios like programming/technology



资料来源: OpenRouter, Data compiled by Goldman Sachs Global Investment Research

Who are best positioned to be the long term winners? Introducing our Competitive Positioning framework

We expect players with the largest ARR scale with a gross margin advantage + financial strength to be the long-term winners.

ARR scale x gross margin advantage + financial strength

- Largest ARR scale (Token scale x pricing power)
- Gross margin advantage (Training & Inference efficiency, Technology)
- Financial strength (Balance sheet, Access to computing)

Accordingly, we introduce our Chinese AI model Competitive Positioning framework based on pricing power, cost advantage and finance strength, overlaying with token scale and market share progressions, and identify Knowledge Atlas (Zhipu initiation) and DeepSeek (private) as the most strongly positioned in foundation models, and Bytedance (private) in multi-modal capabilities.

Foundation models

We assess each key player's competitive positioning from its flagship foundation model, scored across 3 aspects: pricing power, cost advantage and financial strength (of the company), with each aspect further built upon granular quantitative metrics.

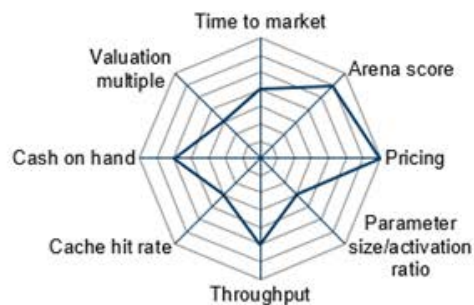
图表 18: Competitive positioning framework for AI model players

Key themes	Chinese models reaching critical stage of intelligence for wide adoption across agentic tasks/specific coding scenarios			Fragmentation between China AI models but strong to get stronger; pricing power to come from frontier performance/multi-modal			Agentic AI driving higher demand for value-for-money models/SLMs		Access to computing, where balance sheet and inference efficiency are key	
Drivers of corporate performance	Competitive positioning									
CP category	Pricing Power			Cost advantage				Financial strength		
CP metrics	Time to market	Arena score	Pricing	Token volume	Throughput /Cache hit rate	Parameter size/ activation ratio	Inference GPM	Cash on hand	Net cash as % of asset	Valuation multiples

资料来源：高盛全球投资研究部

图表 19: Overview of competitive analysis for key LLM labs' positioning

Alibaba Qwen: Full-stack AI player with multi-year AI/cloud upside, leading position of most areas



Tencent Hunyuan: Strong contender after AI strategic revamp; potential in Weixin AI agent



MiniMax M-series: Unique full-modal capabilities among independent players; cost-efficient computing architectures



Zhipu GLM: Leading model capabilities in coding/agent scenarios



DeepSeek: Extreme cost efficiency to sustain scalable token usage on competitive pricing



N/A for DeepSeek valuation multiple

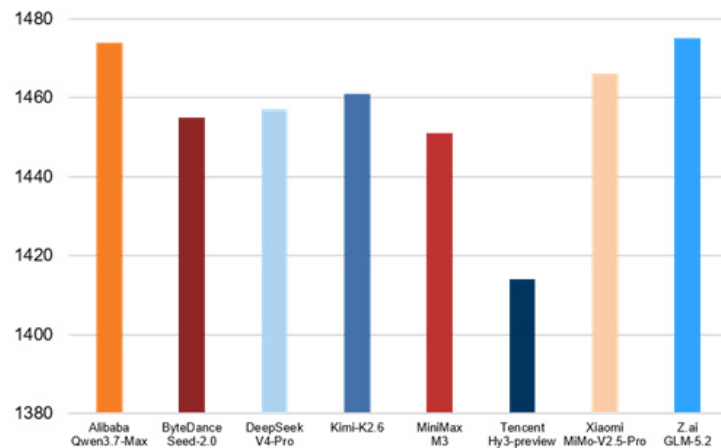
资料来源：高盛全球投资研究部

Pricing Power

- Time to market: We assess this in terms of how quickly and effectively a player delivers frontier competitive models. From comparing the launch date of the company's flagship models with its prior generation & other models at similar performance level, we aim to assess whether the release delivers a meaningful capability step-up over its past generation, and whether it narrows the gap to the current global SOTA models.

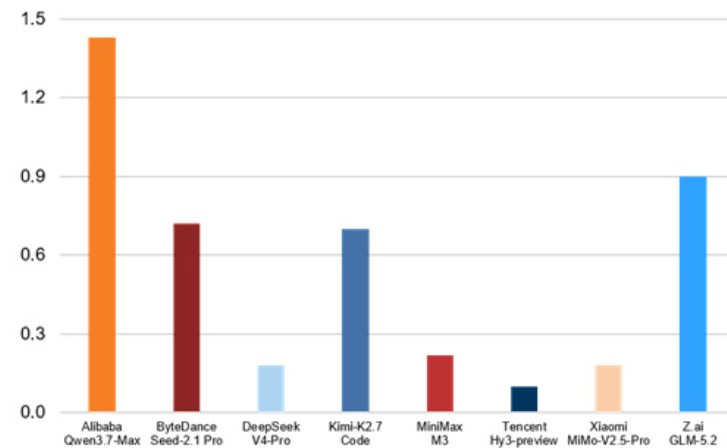
- **Arena score (overall text):** We view model intelligence as the primary determinant of pricing power, since models capable of handling higher-value tasks can deliver clearer ROI and therefore sustain their pricing premium. We use LMArena's score instead of any static benchmark because it reflects a large scale of blind user reviews, making it a more objective read on the real-world capability.
- **Pricing (blended, US\$ per 1M tokens):** We also refer to the realized headline price (across input/output/cached input) for flagship models, where sustained higher pricing with iteration signals stronger pricing power, whereas pricing cuts may indicate a more volume-prioritized strategy.

图表 20: Text Arena score of key domestic AI models



As of June 2026

资料来源: LMArena, Data compiled by Goldman Sachs Global Investment Research

图表 21: Weighted pricing of key player models
US\$ per 1M tokens

As of June 2026

资料来源: Artificial Analysis, 高盛全球投资研究部

Cost advantage: the structural cost-to-serve (i.e. inference costs) that sets the floor for price and margin

We refer to the below metrics as proxies for inference efficiency:

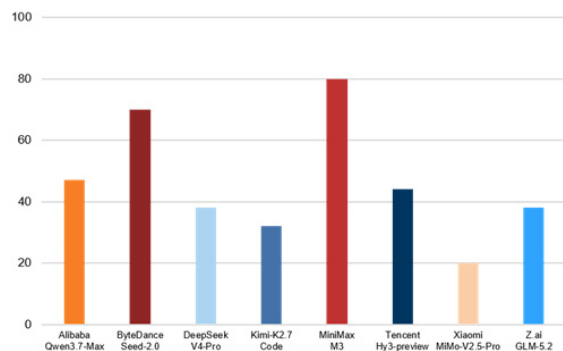
- **Throughput (tokens/second):** Measured by the number of tokens a single GPU generates per second. The more tokens a GPU outputs per second, the more fixed hourly compute cost is spread out. Therefore, throughput is a strong indicator of inference efficiency, which depends on model architecture (sparsity & attention design) and serving efficiency (batching & utilization).
- **Cache hit rate (%):** The share of input tokens served from cache rather than recomputed. In conversations and agentic workflows, much of the input repeats across calls, so models can read those tokens from cache memory

instead of recomputing from new inputs. A higher rate cuts compute, and since cached tokens are near-free to serve despite being billed at a discount (10X-100X cheaper than input cost), it is also margin accretive.

- Parameter size/activation ratio: The share of total parameters activated per token (available for open-source MoE models), where fewer active parameters mean fewer FLOPs per token and therefore lower inference cost, at any given level of performance. We also read the absolute parameter size since calculation scales with this degree of freedom.
- Inference GPM (where disclosed, or GS estimates based on total parameter size/activated parameters): The realized gross margin on model API, where ~90% of COGS is inference costs, as a direct indication on cost efficiency.

图表 22: Throughput of key player models (tokens per second)

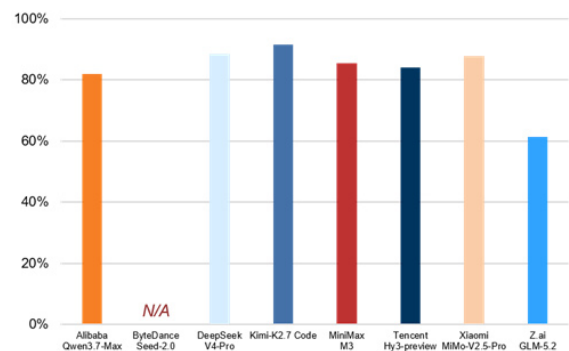
Processing capacity per unit time, representing how fast the model writes



As of June 2026

资料来源: OpenRouter, Data compiled by Goldman Sachs Global Investment Research

图表 23: Cache hit rate of key player models
% of request served from cache memory instead of new inputs, higher = better for cost/speed

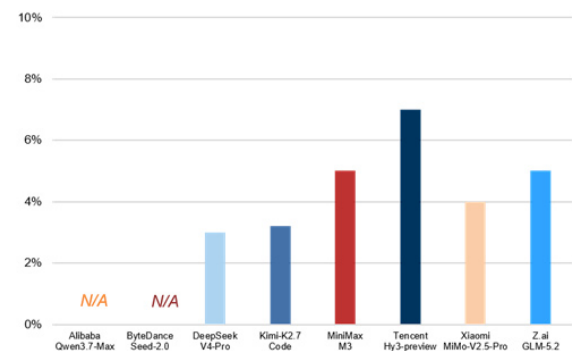


As of June 2026

资料来源: OpenRouter, Data compiled by Goldman Sachs Global Investment Research

图表 24: Active to total parameter ratio of key player models

Share of parameters activated per token; lower means cheaper inference costs given the same performance



Parameter data unavailable for closed-source models (Alibaba's Qwen and ByteDance's Seed); As of June 2026

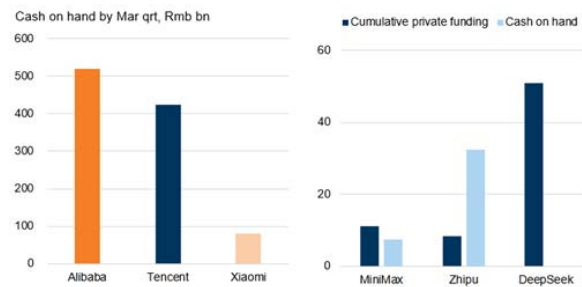
资料来源: 公司数据, Data compiled by Goldman Sachs Global Investment Research

Financial strength: capacity to keep funding frontier R&D and training before a profitability turnaround

Cash on hand and net cash/debt as % of assets: We use total cash on hand as a measure of balance sheet strength, with net cash as % of assets to normalize across players of different scale. We see mega-caps (compared with individual players) as having greater resources to accumulate compute, fund multi-modal exploration and push faster product distribution.

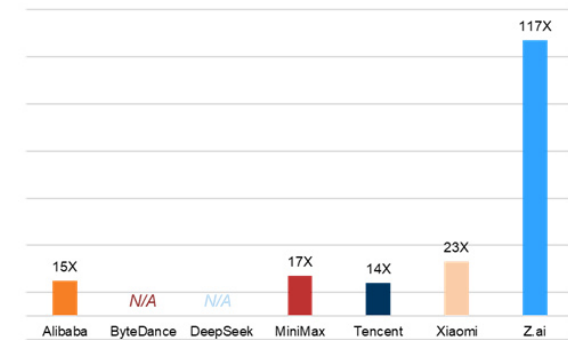
Valuation multiples: For independents, we use P/ARR 2026E multiples as they are not yet profitable and P/ARR is a more comparable measure of business scale and future monetization potential, and for mega-caps we use P/E 2026E.

图表 25: Cash on hand for key players
Rmb bn



资料来源：公司数据, The Information, Data compiled by Goldman Sachs Global Investment Research

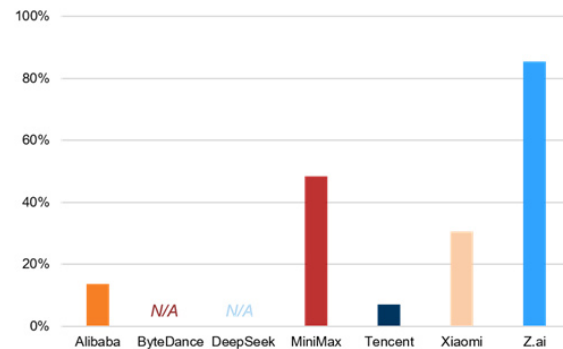
图表 28: Valuation multiples for key players
Mega-caps valuation on P/E, independent players valuation on P/ARR, 2026E



P/E for Alibaba, Tencent, and Xiaomi; P/ARR for MiniMax and Z.ai; Data as of Jun 26, 2026

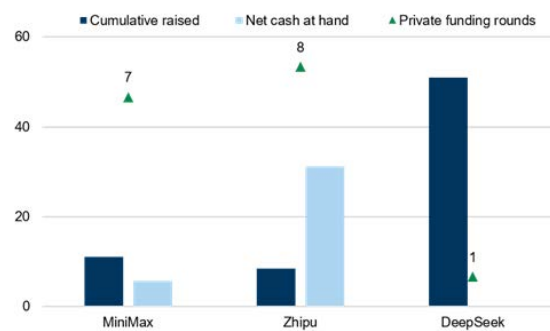
资料来源：公司数据, 高盛全球投资研究部

图表 26: Net cash as % of total asset for key players
As of July 2026



资料来源：公司数据, 高盛全球投资研究部

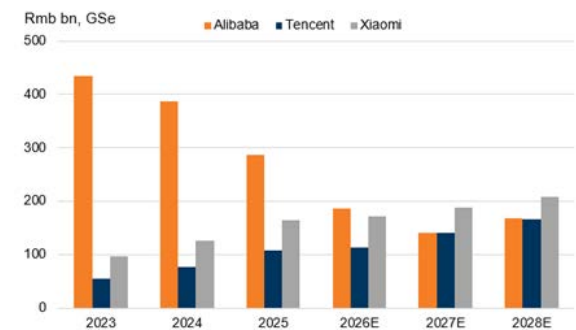
图表 29: Independent AI players' cumulative external private funding (actual) and GSe for net cash as of end-2026E
Rmb bn



N/A for DeepSeek's net cash at hand

资料来源：公司数据, Goldman Sachs Global Investment Research

图表 27: Net cash estimates for Alibaba, Tencent and Xiaomi
Rmb bn, GSe

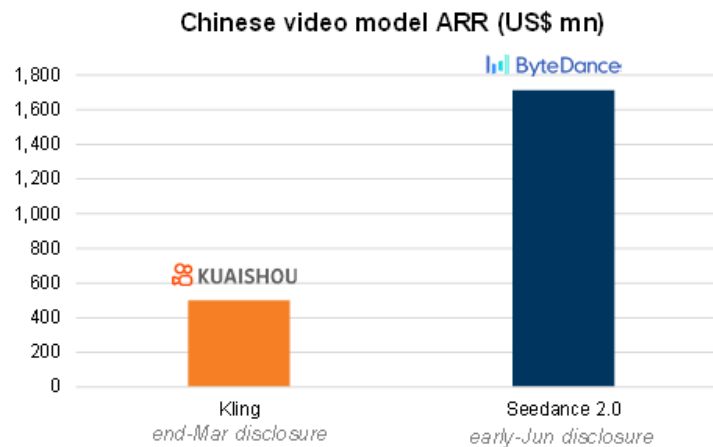


资料来源：公司数据, 高盛全球投资研究部

Multi-modal generation models

We assess the competitive positioning of video generation players (ByteDance, Alibaba, Kuaishou and MiniMax) under a similar framework across 3 dimensions: pricing power, cost advantage, and financial strength, with each built upon more detailed metrics.

图表 30: Latest disclosure of Chinese AI video model ARR (US\$ mn)



资料来源：公司数据, 36kr, Data compiled by Goldman Sachs Global Investment Research

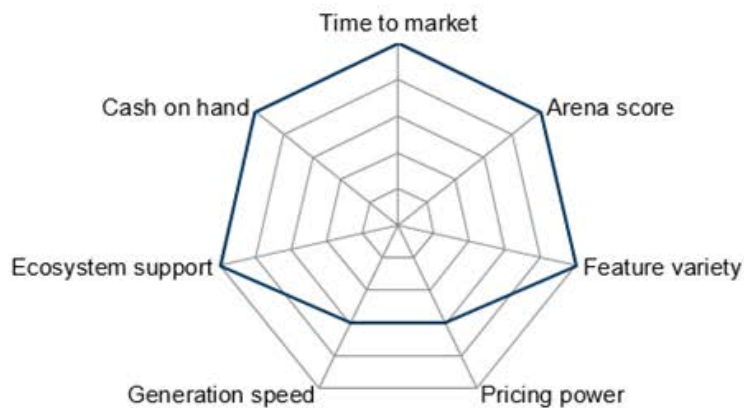
图表 31: Competitive positioning framework for multi-modal generation players

CP category	Pricing Power				Cost advantage	Financial strength		
CP metrics	Time to market	Arena score	Feature variety	Pricing	Generation speed	Cash on hand and net cash as % of asset	Ecosystem support	Valuation multiples (P/ARR)

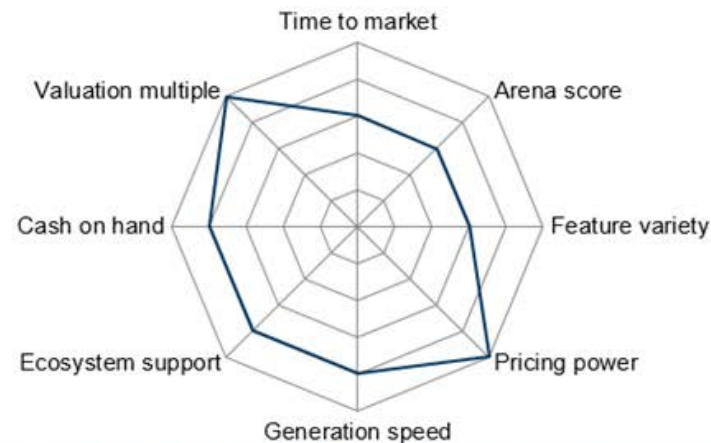
资料来源：高盛全球投资研究部

图表 32: Overview of competitive analysis for key multi-modal labs' positioning

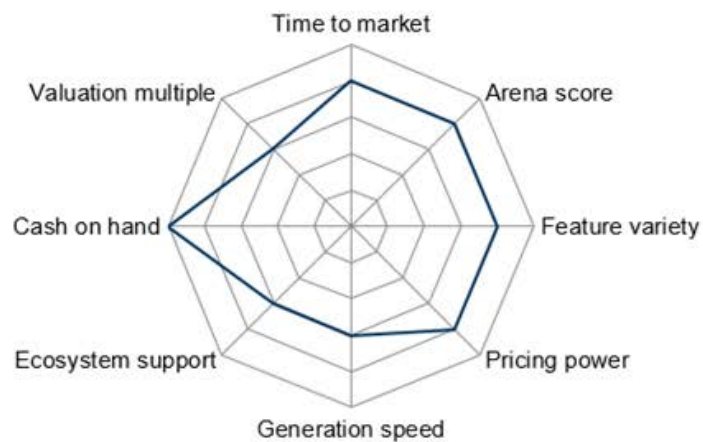
ByteDance: leading in model performance with strong financial/ecosystem back-up for quick adoption & commercialization



Kuaishou: prioritizing enterprise & professional users with strong pricing power and clearer monetization path, supporting a more manageable ROI



Alibaba: abundant resources to support rapid iteration for catching up in performance / feature offerings



MiniMax: native multi-modal training (text+video) to offer synergy on cost efficiency & capability, with anticipated H3 launch as key catalyst



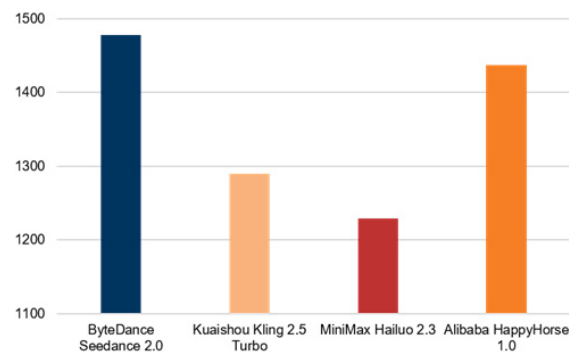
ByteDance's (Not Covered) relevant metrics are calculated by analyzing the industry and companies that we cover, and then extrapolating them to ByteDance. N/A for ByteDance's valuation multiple

资料来源: 高盛全球投资研究部

Pricing Power

- **Time to market:** Measured by how quickly and effectively a player delivers frontier competitive models. From comparing the launch date of the company's flagship models with its prior generation & other models at a similar performance level, we aim to assess whether the release delivers a meaningful capability step-up over its past generation, and whether it narrows the gap to the current global SOTA models.
- **Arena score (average from text-to-video generation and image-to-video generation):** We view model intelligence as the primary determinant of pricing power, since stronger generation models handle higher-value tasks/commercial work and deliver clearer ROI. We use the LMArena score instead of any static benchmark because it reflects a large scale of blind user reviews, making it a more objective read on the real-world capability.
- **Feature variety:** Reflects the generation quality and accessibility of the model, where we further look at 1) native audio (Y/N): whether the model directly outputs audio matched to the visuals, rather than requiring a separate voice/audio model layered on; 2) max duration: how long a single clip the model can generate; 3) max # of storyboards; how many distinct shots the model can generate within one clip, implying its ability to sustain narrative coherence; and 4) max # of multi-modal reference inputs: how many reference materials beyond text prompt (images or video clips) the user can feed in to guide generation.
- **Pricing (US\$ per min of 1080p video from text-to-video generation, without audio):** We also refer to the realized headline price for flagship models, where sustained higher pricing with iteration signals stronger pricing power.

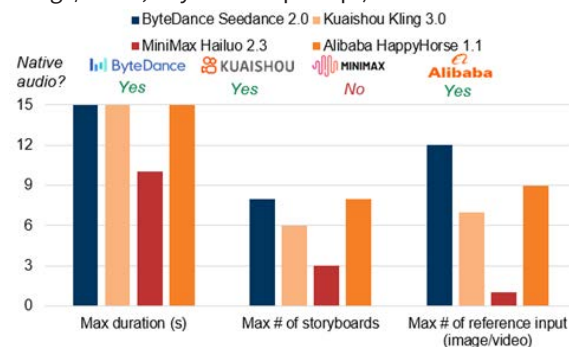
图表 33: Video Arena score of key players
Average score of text-to-video generation and image-to-video generation



As of Jun 2026

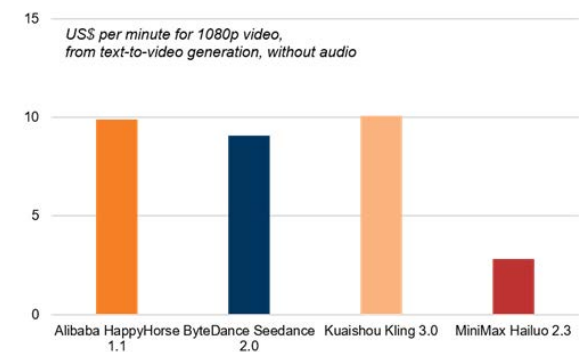
资料来源: LMArena, Data compiled by Goldman Sachs Global Investment Research

图表 34: Feature variety of key players
Native audio (Y/N), max duration (s), max # of storyboards and max # of reference inputs (as image/video, beyond text prompt)



资料来源: 公司数据, Data compiled by Goldman Sachs Global Investment Research

图表 35: API pricing of key players at the same settings
US\$ per minute



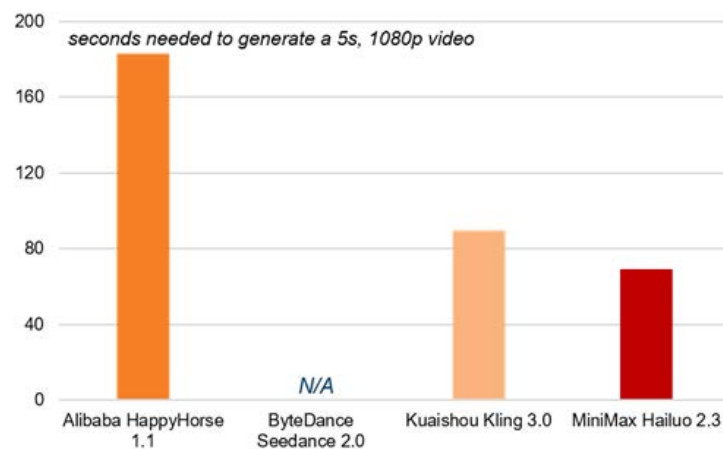
ByteDance Seedance 2.0 datapoint is for 720p video generation (as default setting), all else equal, as per Artificial Analysis

资料来源: Artificial Analysis, Data compiled by Goldman Sachs Global Investment Research

Cost advantage

- Generation speed: Seconds to generate a 5s 1080p clip, where faster speeds indicate higher unit compute utilization and lower inference cost per video.

图表 36: Generation speed of key players
seconds, lower = better



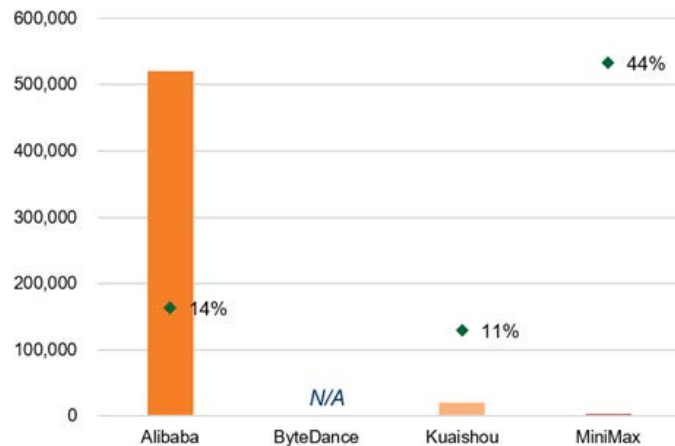
As of early-July, 2026

资料来源: Artificial Analysis, Data compiled by Goldman Sachs Global Investment Research

Financial strength

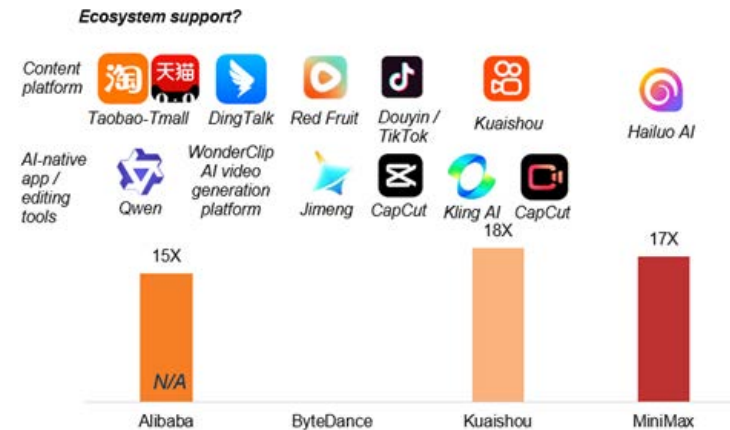
- Cash on hand and net cash/debt as % of assets: We use total cash on hand as a measure of balance sheet strength, with net cash as % of assets to normalize across players of different scale. We see mega-caps (compared with individual players) as having more resources to accumulate compute, fund multi-modal exploration and push faster product distribution.
- Ecosystem support: Customer acquisition and distribution channels within the player's ecosystem (including content platforms, editing tools and MaaS), where platform companies such as ByteDance and Kuaishou have an edge in strengthening enterprise commercial adoption and consumer traffic acquisition, thereby powering their data flywheels.
- Valuation multiples: For both MiniMax and Kuaishou (with Kling AI valuation at US\$18bn post-money implied valuation, see our team's [analysis](#)), we use P/ARR 2026E multiple.

图表 37: Cash on hand and net cash as % of total assets by 2025 for key players
Rmb mn



资料来源：公司数据, 高盛全球投资研究部

图表 38: Ecosystem support and valuation multiples of key players
P/ARR multiples, as of end-2026E



资料来源：LatePost, 公司数据, 高盛全球投资研究部

Key upside and downside risks

Upside risks:

- Stronger-than-expected model intelligence uplift from self-sustainable positive flywheel: Breakthroughs in coding and multi-modal performances, offering seamless integration and understanding across text, image, and audio, would unlock new applications, physical AI and AGI potentials.
- Faster-than-expected path to profit visibility: If Chinese AI model companies shift faster toward being open-weight (from pure open source) and demonstrate superior operational efficiency, such as optimizing model inference costs and training infrastructure, or diversify monetization channels to achieve higher revenue growth or better unit economics sooner, it would accelerate the timeline to positive cash flow and net income. Early profitability would boost investor confidence on the AI model industry, potentially reducing the need for future external funding.
- Additional non-inference take rate revenue streams: If Chinese AI model companies are able to train and achieve SOTA model performance, an open-weight approach charging take rates/revenue sharing on neoclouds/hyperscalers that house such models and sell to customers may transform AI model company business models to a high gross margin business (without needing to conduct inference for each token sold, but earning via revenue sharing with platforms that sell top performing Chinese models).

- Stronger-than-expected commercialization capability: Growth of new applications, Code and Co-Worker platforms, coupled with ‘digital worker’ agent models could provide multiple channels for revenue generation. If these channels achieve higher-than-expected adoption rates, conversion efficiencies, or ARPU, Chinese AI models could outperform current revenue forecasts.

Downside risks:

- Competition in global foundation model industry, vs. US/China mega-caps: Chinese AI models operate within the intensely competitive and rapidly evolving global foundation model industry, where the risks include the challenge of keeping pace with swift technological advancements, frequent new product introductions, and dynamic customer demand. Failure to adapt products and invest significantly in R&D in a cost-effective and timely way, could lead to diminished competitiveness. In addition, the early-stage nature of the market opportunities makes customer demand unpredictable and an inability to meet user requirements would risk losing existing users and adversely impact revenue.
- Limited near-term profit visibility due to high R&D expenses: Achieving and sustaining profitability of Chinese AI model companies in the near future could be a challenge, as success depends on continuous technological development, user experience improvement, and effective commercialization. Failure to manage escalating costs and generate adequate revenue could materially impact the company’s financial position and profitability.
- Risks around IP/content generation: Chinese AI models, especially multi-modal generation models, face risks from the generation of inappropriate content. Similar to challenges encountered by other AI applications, such as lawsuits against Character AI’s chatbot for its alleged role in teens’ safety, image/video-generation models could be exposed to reputational damage and potential lawsuits. Robust content moderation and ethical AI development are crucial to prevent adverse output and ensure a safe user experience. Chinese companies could also face risks from third party intellectual property/distillation method infringement claims, from both the data used to train their models and the content they generate.
- Cash burn/self-funding capability: The intense capital requirements for advanced AI model training, coupled with rapid technological advancements by competitors, could pose financing risks to the relatively fragmented Chinese AI model landscape. In addition, any models that fail to keep pace with leading peers could run into difficulties in generating sufficient revenue (ARR x inference gross margins) to cover training costs.
- Risk around geopolitics under an intensified tech race between the US and China: Any US restrictions on the use of Chinese AI models could limit market access and adoption, and Chinese AI model companies could also face increased costs if barred from training its models on US cloud service providers or be included into any list which limits procurement of US supplies from hardware to software. The evolving global chip availability dynamics also

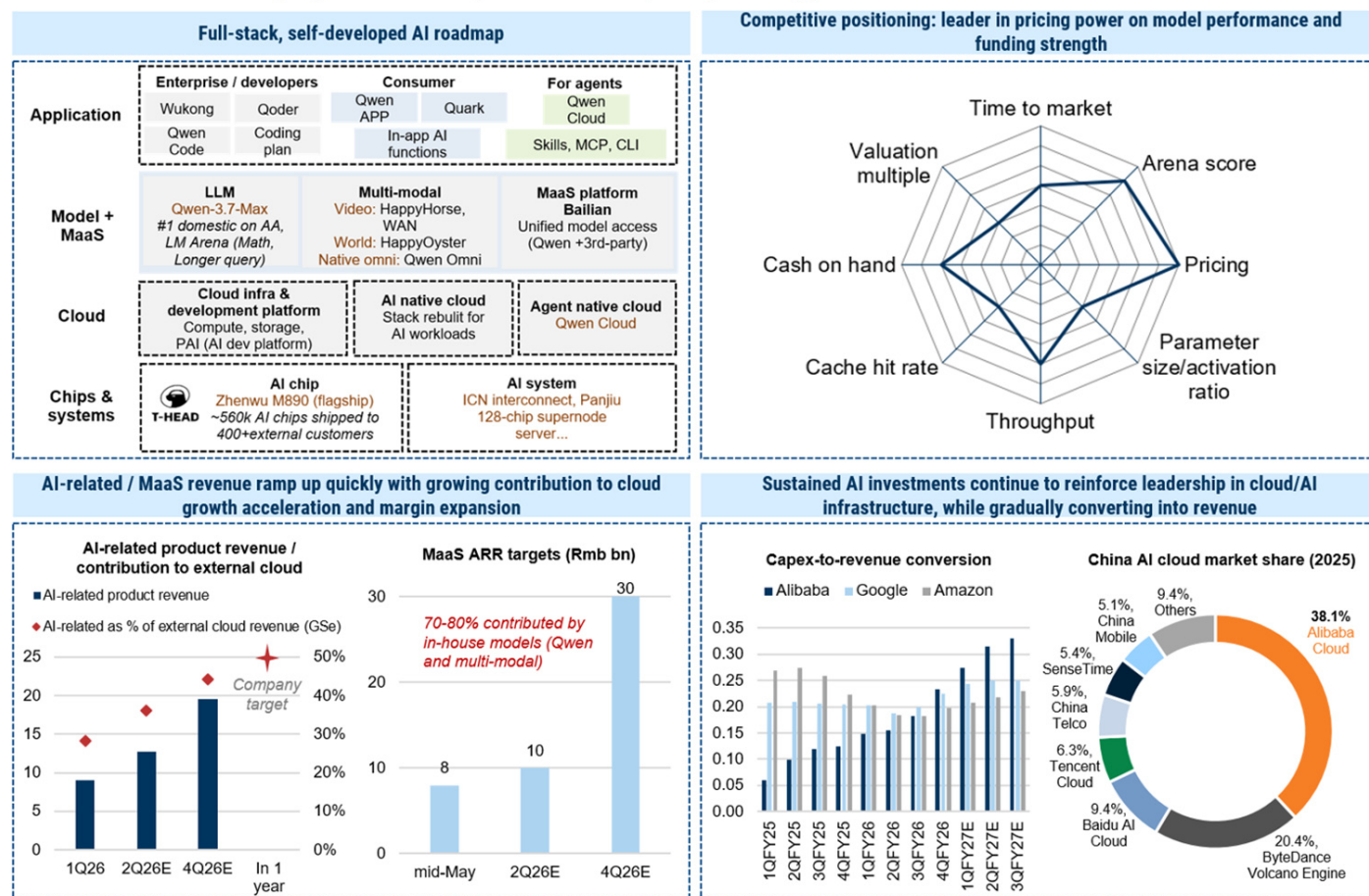
present a critical supply chain risk.

China's key AI players at a glance

Mega-caps

图表 39: Alibaba

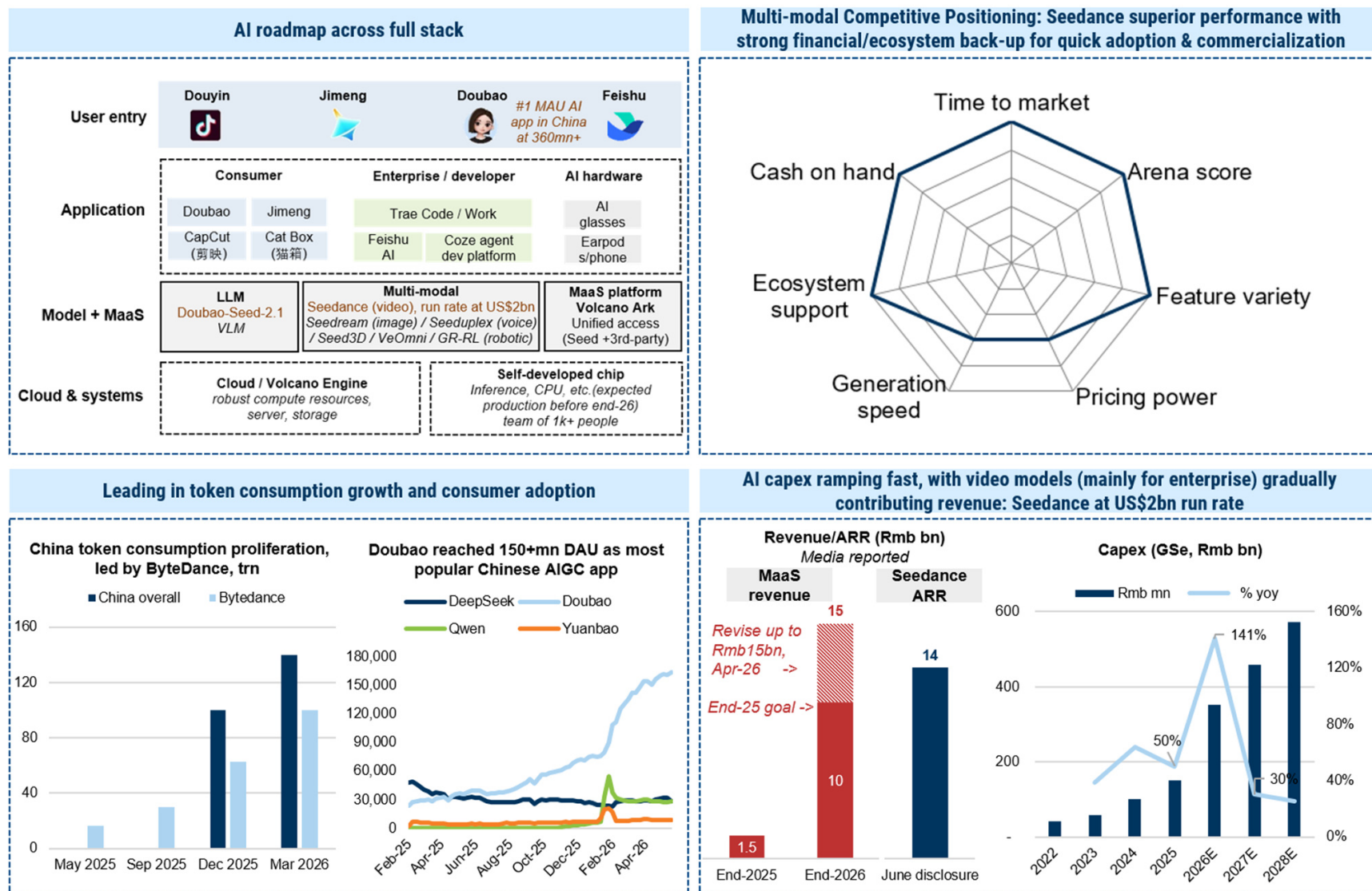
Alibaba: Full-stack AI player with multi-year AI/cloud upside, leading position of most areas



资料来源：公司数据, 高盛全球投资研究部, Omdia

图表 40: ByteDance

ByteDance: Deep resources for broad AI exploration with strong consumer base and product portfolio

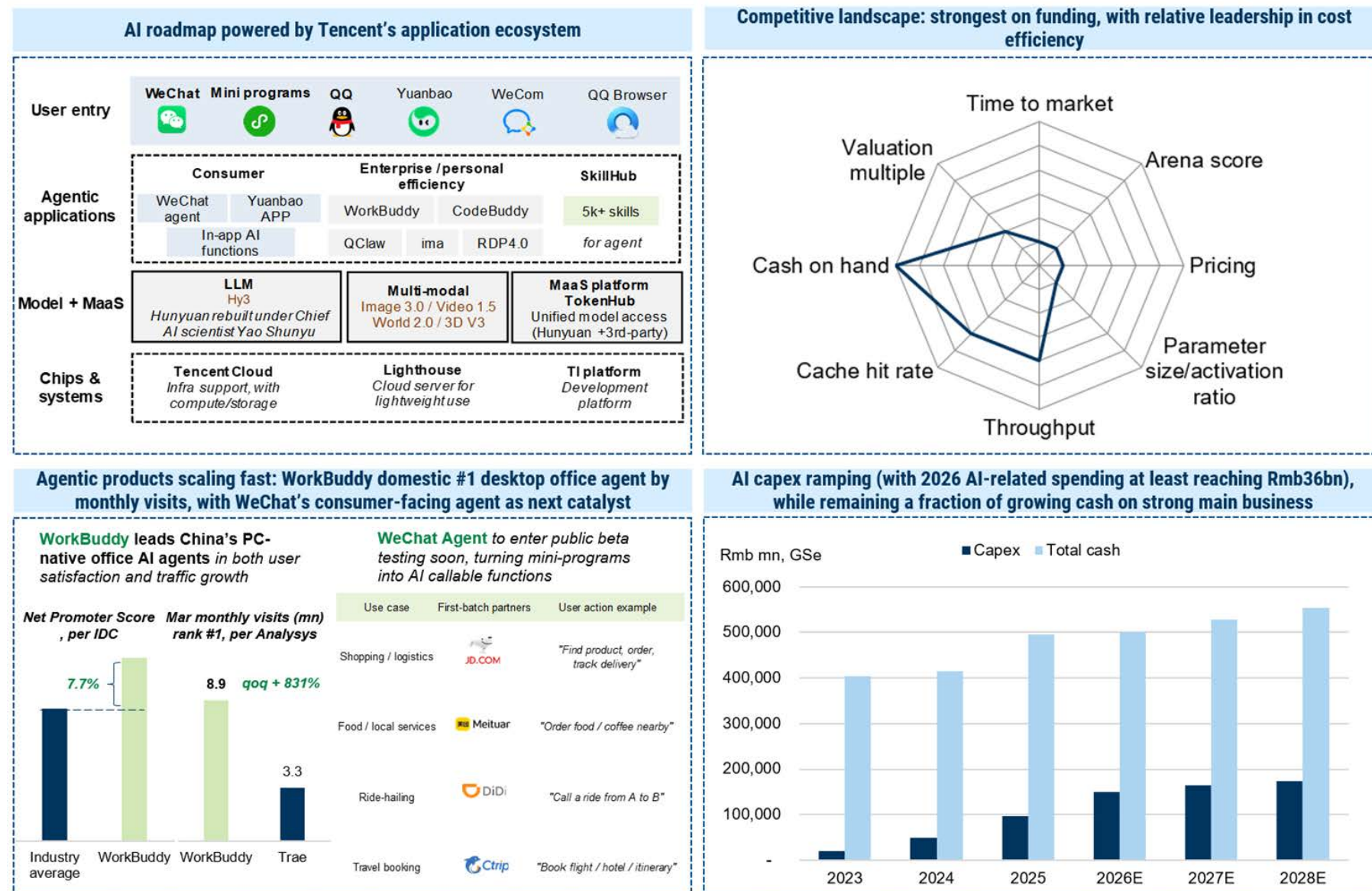


ByteDance's (Not Covered) competitive positioning metrics are calculated by analyzing the industry and companies that we cover, and then extrapolating them to ByteDance. N/A for the valuation multiple in the competitive positioning chart

资料来源: 公司数据, 高盛全球投资研究部, LatePost, QuestMobile

图表 41: Tencent

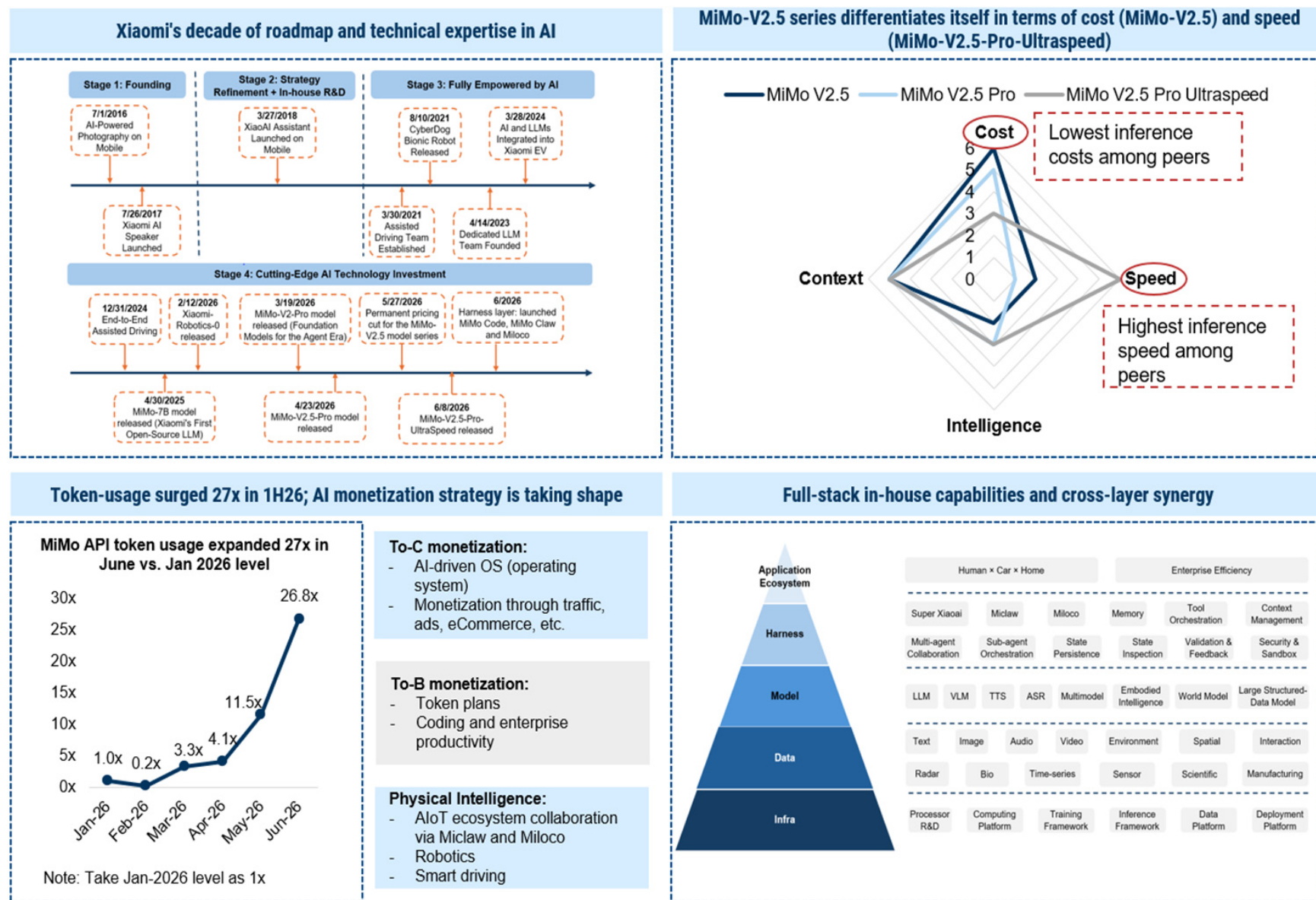
Tencent: Strong contender after AI strategic revamp; potential in Weixin AI agent



资料来源：公司数据，高盛全球投资研究部，Analysys, IDC

图表 42: Xiaomi

Xiaomi: Pushes efficiency frontiers of LLMs; Rich applications scenarios in to-C/to-B and physical intelligence

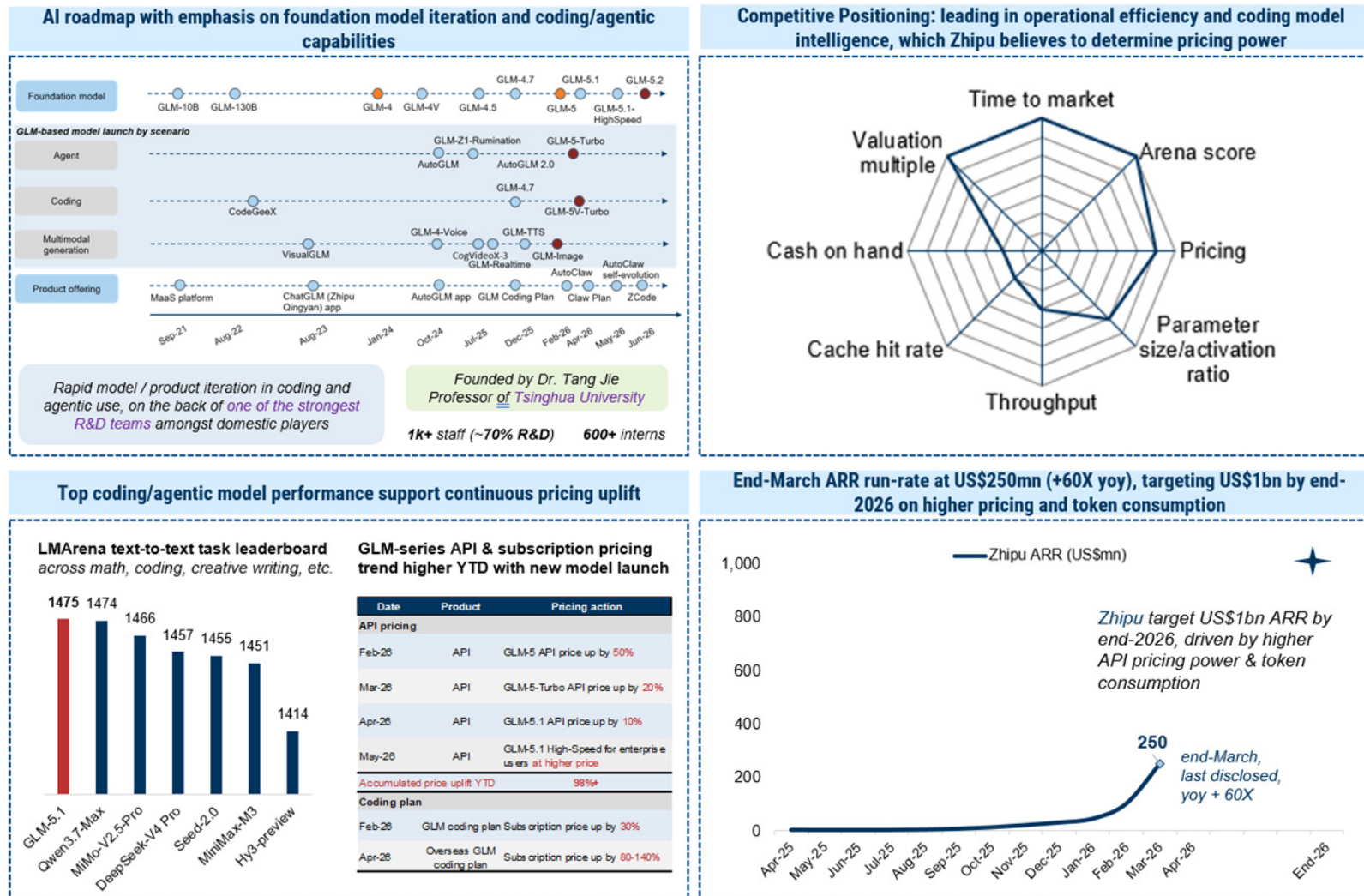


资料来源：公司数据，高盛全球投资研究部

Key independent AI players

图表 43: Zhipu

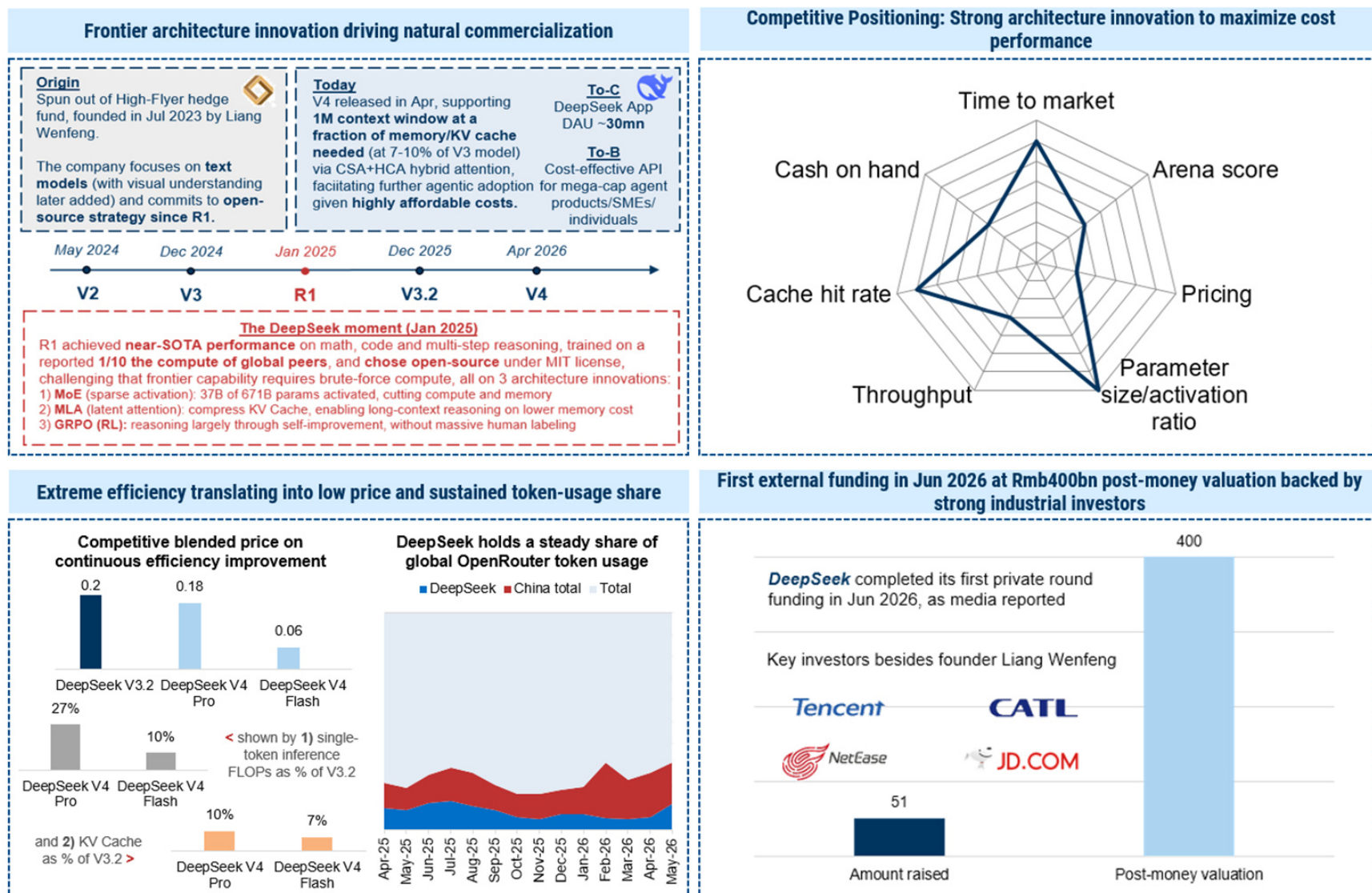
Zhipu: Leading model capabilities in coding/agent scenarios



资料来源：公司数据，高盛全球投资研究部，LMarena

图表 44: DeepSeek

DeepSeek: Extreme cost efficiency to sustain scalable token usage on competitive pricing

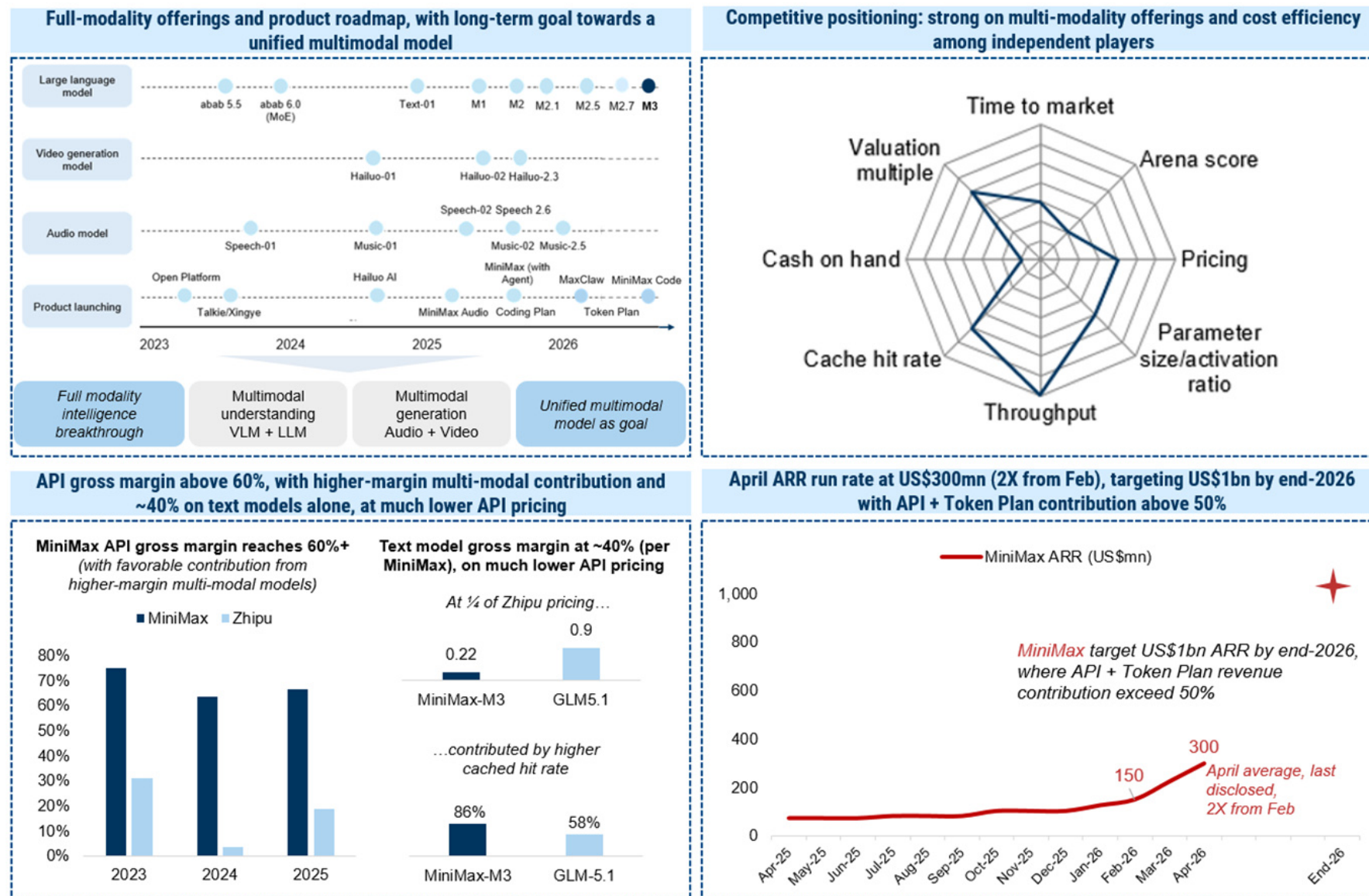


DeepSeek's (Not Covered) competitive positioning metrics are calculated by analyzing the industry and companies that we cover, and then extrapolating them to DeepSeek. DeepSeek's latest funding round based on Reuters.

资料来源：公司数据, 高盛全球投资研究部, OpenRouter, 路透

图表 45: MiniMax

MiniMax: Unique full-modal capabilities among independent players; cost-efficient computing architectures



资料来源：公司数据，高盛全球投资研究部

Price Target Methodology & Risks

Price Target Risks and Methodology - Alibaba Group

Valuation methodology: We are Buy rated on BABA/9988.HK, with SOTP-based 12-month target prices of US\$186/HK\$180 per ADS/share (FY27E-based).

Key risks: (1) Lower-than-expected GMV growth due to macro/competition; (2) Slower-than-expected monetization in China retail; (3) Weaker-than-expected execution in key strategic investments; and (4) Cloud revenue growth deceleration.

Price Target Risks and Methodology - Tencent Holdings

Valuation methodology: We are Buy rated on Tencent (0700.HK), with a 12-m SOTP-based TP of HK\$700.

Key risks: More intense industry competition in performance-based advertising, unexpected delays in game launches/Banhao approvals, moderated growth in FinTech and Cloud businesses, slower-than-expected AI development progress, reinvestment risk.

Price Target Risks and Methodology - Xiaomi Corp.

We are Buy rated on Xiaomi with a 12-month target price of HK\$40, based on a SOTP valuation methodology: 1) 16x target 12m-fwd EV/NOPAT on Xiaomi core; 2) DCF-based valuation for Xiaomi EV of US\$42bn (12% WACC and 3% TGR); 3) a 10% holdco discount. Key downside risks: 1) More intense competition and weaker market share gains within the global smartphone industry. 2) Higher GPM pressure on the smartphone/EV business. 3) Below-expected execution on Xiaomi brand premiumization and the EV business. 4) Intensifying geopolitical risks and regulatory uncertainties. 5) Softer macro environment and weaker smartphone/IoT demand. 6) FX fluctuation.

Price Target Risks and Methodology - MiniMax Group

We are Buy rated on MiniMax with a 12-month target price of HK\$860. Our valuation is based on DCF valuation, with 12% WACC and 2% terminal growth rate.

Key risks: Weaker-than-expected model performance amid the competition in the global foundation model industry; Slower-than-expected path to profit visibility; Weaker-than-expected commercialization capability; risks around IP/content generation; cash burn/self-funding capability; risks around geopolitics under an intensified tech race between the US and China.

Price Target Risks and Methodology - Kuaishou Technology

Valuation methodology: We are Buy rated on Kuaishou with a 12m target price of HK\$68, based on P/E valuation (target P/E multiple of 15X applied to our avg. 2026E-27E EPS).

Key downside risks: 1) Slower-than-expected ad budget recovery; 2) Weaker-than-expected Kling monetization; 3) Slower-than-expected growth of its aggregate user engagement base (DAU x time spent per DAU); 4) Lower-than-expected profitability; 5) Weaker-than-expected AI progress.

信息披露附录

申明

我们，姜朗霆, CFA、Eric Sheridan、张博凯、孔令侃, CFA、Timothy Zhao、谢元滔、Iris Xiao、邱品儒、周璐晴、刘昱岑，在此申明，本报告所表述的所有观点准确反映了我们对上述公司或其证券的个人看法。此外，我们的薪金的任何部分不曾与，不与，也将不会与本报告中的具体推荐意见或观点直接或间接相关。

本报告首页所列作者为高盛全球投资研究部分分析师，除非另有说明。

贡献作者: 姜朗霆, CFA Goldman Sachs (Asia) L.L.C.、Eric Sheridan Goldman Sachs & Co. LLC、张博凯 Goldman Sachs (Asia) L.L.C.、孔令侃, CFA Goldman Sachs (Asia) L.L.C.、Timothy Zhao Goldman Sachs (Asia) L.L.C.、谢元滔 Goldman Sachs (Asia) L.L.C.、Iris Xiao Goldman Sachs (Asia) L.L.C.、邱品儒 Goldman Sachs (Asia) L.L.C.、周璐晴 Goldman Sachs (Asia) L.L.C.、刘昱岑 Goldman Sachs (Asia) L.L.C.。

本报告披露的“贡献作者”为高盛全球投资研究部分分析师，除非另有说明。

高盛要素概要

高盛要素概要部分通过将一只股票的主要指标与市场（即我们给予评级的股票范围）和可比同业相比较来评价该股的投资背景。四个主要指标是增长、财务回报、估值倍数（估值）和综合状况（增长、财务回报、估值倍数的综合情况）。增长、财务回报和估值倍数是运用每只股票具体指标的标准化排名计算。随后取这些指标标准化排名的均值并转化为相关指标的百分位。每项指标的具体计算方式可能随着财务年度、行业和所属地区的不同而有所变化，但标准方法如下：

增长指标是基于一只股票的预期销售增速、EBITDA增速和每股盈利增速（金融股仅采用每股盈利和销售增速），较高的百分位表示公司增长较快。财务回报是基于一只股票的预期净资产回报率、ROCE和CROCI（金融股仅采用净资产回报率），较高的百分位表示公司的财务回报较高。估值倍数基于一只股票的预期市盈率、市净率、股价/股息、EV/EBITDA、EV/FCF、EV/DACF（EV/经债务调整的现金流）（金融股仅采用市盈率、市净率和股价/股息），较高的百分位表示公司的估值倍数较高。综合状况百分位为增长百分位、财务回报百分位和（100% - 估值倍数百分位）的平均值。

财务回报和估值倍数使用高盛分析师在财政年度末对未来至少3个季度的预测。增长使用未来至少7个季度的财政年度预测与未来至少3个季度的财政年度预测的比较（所有指标均使用每股数据）。

如需了解高盛要素概要更具体的计算，请联络您的高盛代表。

并购评分

在我们的全球覆盖范围中，我们使用并购框架来分析股票，综合考虑定性和定量因素（各行业和地区可能会有所不同）以计入某些公司被收购的可能性。然后我们按照从1到3对公司进行并购评分，其中1分代表公司成为并购标的的概率较高(30%-50%)，2分代表概率为中等(15%-30%)，3分代表概率较低(0%-15%)。对于评分为1或2的公司，我们按照研究部统一标准将并购因素体现在我们的目标价格当中。并购评分为3被认为意义不大，因此不予体现在我们的目标价格当中，分析师在研究报告中可以予以讨论或不予讨论。

Quantum

Quantum是提供具体财务报表数据历史、预测和比率的高盛专有数据库，它可以用于对单一公司的深入分析，或在不同行业和市场的公司之间进行比较。

信息披露

Logo disclosure

Third party brands used in this presentation are the property of their respective owners, and are used here for informational purposes only. The use of such brands should not be viewed as an endorsement, affiliation or sponsorship by or for Goldman Sachs or any of its products/services.

Financial Advisory Disclosure

Goldman Sachs and/or one of its affiliates is acting as a financial advisor in connection with an announced strategic matter involving the following company or one of its affiliates: Baidu, Inc.

Rating and pricing information

Kuaishou Technology (Buy, HK\$40.46) and MiniMax Group (Buy, HK\$323.80)

Alibaba Group (ADR), Alibaba Group (H), Knowledge Atlas Technology, MiniMax Group 和 Tencent Holdings 的股票评级是相对于其所属研究范围内的其他公司的相对评级：Alibaba Group (ADR), Alibaba Group (H), DiDi Global Inc., Full Truck Alliance Co., J&T Global Express Ltd., JD Logistics, JD.com Inc. (ADR), JD.com Inc. (H), Jingdong Industrials Inc., Kerry Logistics Network Ltd., Knowledge Atlas Technology, Meituan, MiniMax Group, PDD Holdings, S.F. Holding (A), S.F. Holding (H), STO Express, Sinotrans Ltd. (A), Sinotrans Ltd. (H), Tencent Holdings, Vipshop Holdings, YTO Express Group, Yunda Holding, ZTO Express (Cayman) Inc. (ADR), ZTO Express (Cayman) Inc. (H)

Kuaishou Technology 的股票评级是相对于其所属研究范围内的其他公司的相对评级：37 Interactive Entertainment, Alibaba Health, Bilibili Inc. (ADR), Bilibili Inc. (H), Damai Entertainment Holdings, Focus Media Information Tech, G-Bits, JD Health International, Kuaishou Technology, Mango Excellent Media, NetEase Inc. (ADR), NetEase Inc. (H), Perfect World, Tencent Music Entertainment Group (ADR), Tencent Music Entertainment Group (H), XD Inc., Zhihu Inc. (ADR), Zhihu Inc. (H), iQIYI Inc.

Xiaomi Corp. 的股票评级是相对于其所属研究范围内的其他公司的相对评级：Autohome Inc. (ADR), Autohome Inc. (H), Beijing Sinnet Technology Co Ltd., East Buy, GDS Holdings (ADR), GDS Holdings (H), KE Holdings (ADR), KE Holdings (H), Kanzhun Ltd. (ADR), Kanzhun Ltd. (H), Kingsoft Cloud, Ming Yuan Cloud, New Oriental Education & Technology (ADR), New Oriental Education & Technology (H), Offcn Education Technology, Range Intelligent, SUNeVision Holdings, Shanghai Athub, TAL Education Group, Tuhu Car Inc., Tuya, VNET Group, Weibo Corp. (ADR), Weibo Corp. (H), Weimob, Xiaomi Corp.

与公司有关的法定披露

信息披露摘要：请参阅<https://www.gs.com/research/hedge.html>的披露信息。本摘要中所涉及公司适用的披露信息也可参见最近出版的相关研究报告。

评级分布/投资银行关系

高盛投资研究部的全球研究覆盖范围

	评级分布			投资银行关系		
	买入	持有	卖出	买入	持有	卖出
全球	50%	34%	16%	65%	60%	45%

截至2026年4月1日，高盛全球投资研究部对3,074种股票评定了投资评级。高盛给予股票在各种地区投资名单中的买入和卖出评级；未给予这些评级的股票被视为中性评级，根据FINRA的披露要求，这些评级分别对应买入，持有及卖出。详情见以下“公司评级，研究范围和相关定义”部分。投资银行关系表反映了高盛在过去12个月已提供投资银行服务的公司在各评级类别中所占的比例。

目标价格和评级历史图

信息披露摘要：请参阅<https://www.gs.com/research/hedge.html>的披露信息。本摘要中所涉及公司适用的披露信息也可参见最近出版的相关研究报告。

目标价格历史列表

Alibaba Group (ADR) (BABA)

Date of report	Target price (\$)	Closing price (\$)
27-Feb-26	186.00	144.11
25-Nov-25	197.00	157.01
12-Oct-25	205.00	159.01
16-Sep-25	179.00	162.21
31-Aug-25	163.00	135.00
15-Jul-25	147.00	116.97
02-Jul-25	150.00	110.71
21-Apr-25	159.00	110.15
20-Feb-25	160.00	135.97
20-Jan-25	117.00	85.12
29-Sep-24	134.00	107.33
15-Aug-24	108.00	79.54
16-Jul-24	107.00	78.38
15-May-24	106.00	80.99
11-Apr-24	102.00	74.85
07-Feb-24	105.00	73.64
15-Jan-24	121.00	71.84
09-Oct-23	134.00	84.85
10-Aug-23	138.00	99.21
16-Jul-23	133.00	94.56

Tencent Holdings (0700.HK)

Date of report	Target price (HK\$)	Closing price (HK\$)
18-Mar-26	700.00	550.50
18-Jan-26	752.00	617.50
12-Oct-25	770.00	651.50
13-Aug-25	701.00	586.00
05-Aug-25	658.00	559.00
14-May-25	595.00	521.00
19-Mar-25	590.00	540.00
20-Jan-25	534.00	389.80
13-Nov-24	542.00	403.80
29-Sep-24	521.00	437.80
14-Aug-24	464.00	373.80
12-Jun-24	477.00	370.80
15-May-24	466.00	381.80
11-Apr-24	408.00	315.00
20-Mar-24	413.00	288.80
15-Jan-24	426.00	289.40
15-Nov-23	435.00	322.60
11-Oct-23	423.00	311.00
16-Aug-23	431.00	328.80
07-Aug-23	428.00	341.20

Xiaomi Corp. (1810.HK)

Date of report	Target price (HK\$)	Closing price (HK\$)
26-May-26	40.00	29.76
09-Mar-26	41.00	33.68
23-Jan-26	47.50	36.24
18-Nov-25	53.50	40.78
01-Nov-25	56.50	43.20

Kuaishou Technology (1024.HK)

Date of report	Target price (HK\$)	Closing price (HK\$)
29-Jun-26	68.00	41.00
27-May-26	70.00	45.28
14-May-26	74.00	50.85
25-Mar-26	75.00	53.05
18-Jan-26	87.00	78.35

Xiaomi Corp. (1810.HK)

Date of report	Target price (HK\$)	Closing price (HK\$)
25-Sep-25	66.00	59.45
19-Aug-25	65.00	52.40
26-Jun-25	69.00	56.90
27-May-25	65.00	51.55
18-May-25	62.00	51.00
16-Apr-25	59.00	41.25
18-Mar-25	63.00	57.65
20-Feb-25	58.00	49.15
14-Jan-25	38.00	33.75
18-Nov-24	33.30	28.80
29-Oct-24	30.70	25.85
15-Oct-24	27.80	23.00
03-Oct-24	27.50	24.05
21-Aug-24	24.70	17.52
21-Jul-24	23.20	16.52
24-May-24	23.00	18.30
12-May-24	22.60	19.40
01-Apr-24	20.00	14.94
19-Mar-24	18.90	14.86
06-Feb-24	18.10	12.90
20-Nov-23	18.70	16.18
18-Oct-23	16.30	13.18
31-Aug-23	14.90	12.36
09-Aug-23	14.70	12.14

Knowledge Atlas Technology (2513.HK)

Date of report	Target price (HK\$)	Closing price (HK\$)
09-Jul-26	1,880.00	2,032.00

Kuaishou Technology (1024.HK)

Date of report	Target price (HK\$)	Closing price (HK\$)
31-Aug-25	83.00	75.35
21-Aug-25	77.00	71.75
17-Jun-25	68.00	59.40
27-May-25	63.00	48.75
25-Mar-25	62.00	56.80
19-Jan-25	55.00	40.75
20-Nov-24	63.00	52.55
20-Aug-24	67.00	44.40
08-Aug-24	70.00	43.85
22-May-24	80.00	58.15
20-Mar-24	75.00	52.25
21-Jan-24	82.00	41.55
21-Nov-23	88.00	58.50
19-Oct-23	87.00	56.10
22-Aug-23	91.00	62.50

Alibaba Group (H) (9988.HK)

Date of report	Target price (HK\$)	Closing price (HK\$)
27-Feb-26	180.00	142.90
25-Nov-25	192.00	157.80
12-Oct-25	199.00	165.40
16-Sep-25	174.00	153.50
31-Aug-25	158.00	115.70
15-Jul-25	142.00	113.50
02-Jul-25	146.00	109.40
21-Apr-25	154.00	108.70
20-Feb-25	156.00	120.90
20-Jan-25	114.00	84.55
29-Sep-24	131.00	102.50
15-Aug-24	105.00	76.40
16-Jul-24	104.00	75.50
15-May-24	103.00	82.65
11-Apr-24	99.00	74.20

Knowledge Atlas Technology (2513.HK)

Date of report Target price (HK\$) Closing price (HK\$)

07-Feb-24	102.00	74.90
15-Jan-24	118.00	70.00
09-Oct-23	131.00	82.10
10-Aug-23	134.00	94.30
16-Jul-23	130.00	92.95

Alibaba Group (H) (9988.HK)

Date of report Target price (HK\$) Closing price (HK\$)

07-Feb-24	102.00	74.90
15-Jan-24	118.00	70.00
09-Oct-23	131.00	82.10
10-Aug-23	134.00	94.30
16-Jul-23	130.00	92.95

MiniMax Group (0100.HK)

Date of report Target price (HK\$) Closing price (HK\$)

16-Jun-26	860.00	416.60
03-Mar-26	1,000.00	821.00
22-Feb-26	1,018.00	970.00

法定披露
美国法定披露

任何本报告中研究企业所需的特定公司法定披露见上文：包括即将进行交易的承销商或副承销商，1%或其他股权，特定服务的补偿，客户关系种类，之前担任承销商或副承销商的公开发行人，担任董事，担任股票做市及／或专家的角色。高盛担任或可能担任本报告中所涉及发行方的债券（或相关衍生品）的交易对手。

以下为额外要求的披露：股权及重大利益冲突：高盛的政策为禁止其分析师、分析师属下专业人员及其家庭成员持有分析师负责研究的任何公司的证券。分析师薪酬：分析师薪酬部分取决于高盛的盈利，其中包括投资银行的收入。分析师担任高级职员或董事：高盛的政策通常禁止其分析师、分析师属下人员及其家庭成员担任分析师负责研究的任何公司的高级职员、董事或顾问。非美国分析师：非美国分析师可能与高盛无关，因此可以不受FINRA 2241条FINRA 2242条对于与所研究公司的交流、公开露面及交易分析师所研究证券的限制。

美国以外司法管辖区规定的额外披露

以下为除了根据美国法律法规规定作出的上述信息披露之外其他司法管辖区法律所要求的披露。澳大利亚: Goldman Sachs Australia Pty Ltd及其相关机构不是澳大利亚经授权的存款机构（1959年《银行法》所定义），因此不在澳大利亚境内提供银行服务，也不经营银行业务。本研究报告或本报告的其他形式内容只可分发给根据澳大利亚公司法定义的“批发客户”，在事先获得高盛许可的情况下可以有例外。在撰写研究报告期间，Goldman Sachs Australia全球投资研究部的职员可能参与本研究报告中所讨论证券的发行公司或其他实体组织的现场调研或会议。在某些情况下，如果视具体情形Goldman Sachs Australia认为恰当或合理，此类调研或会议的成本可能部分或全部由该证券发行人承担。如本报告内容包含任何金融产品建议，则该建议仅为一般建议，且高盛提出该建议时并未考虑客户的目标、财务状况或需求。客户在就此类建议采取行动之前，应结合其自身目标、财务状况和需求来考虑该建议的适当性。高盛澳大利亚和新西兰的利益披露，以及高盛澳大利亚卖方研究独立性制度声明请参见<https://www.goldmansachs.com/disclosures/australia-new-zealand/index.html>。巴西: 与CVM Resolution n. 20相关的信息披露请参阅<https://www.gs.com/worldwide/brazil/area/gir/index.html>。根据CVM Resolution n. 20第20条，在适用的情况下，对本研究报告内容负主要责任的巴西注册分析师为本报告开头部分标明的第一作者，除非报告末另有说明。加拿大: 这些信息仅供您参考，在任何情况下都不应被理解为Goldman Sachs & Co. LLC对加拿大证券购买者进行有关任何加拿大证券交易的广告、要约或征求行为。Goldman Sachs & Co. LLC未在适用的加拿大证券法规下注册为任何加拿大司法管辖区内的交易商，通常不被允许交易加拿大证券，并且可能被禁止在加拿大某些司法管辖区内销售某些证券和产品。若您想在加拿大交易任何加拿大证券或其他产品，请联系 Goldman Sachs Canada Inc. (高盛集团的关联机构) 或其他已注册的加拿大交易商。香港: 可从高盛（亚洲）有限责任公司获取有关本报告中所研究公司的证券的额外资料。印度: 可从高盛（印度）证券私人有限公司（分析师 - 印度证券交易委员会(SEBI) 编号INH000001493，地址10th Floor, Ascent-Worli, Sudam Kalu Ahire Marg, Worli, Mumbai-400 025, India, 公司编号 U74140MH2006FTC160634, 电话 +91 22 6616 9000, 传真 +91 22 6616 9001）获取有关本报告中研究对象或所提及公司的额外资料。高盛可能持有本报告中研究对象或所提及公司的证券（1956年印度《证券合同(管理)法》条款2(h)之定义）的1%或更高比例。在SEBI注册并获得NISM认证并非对该中介机构表现的担保，亦不能对投资者回报做出保障。高盛（印度）证券私人有限公司合规负责人和投资者支持部门联系方式、年度合规审计报告副本以及其他相关信息和披露请点击此链接<https://www.goldmansachs.com/worldwide/india/research-analyst>。日本: 见下文。韩国: 除非高盛另行同意，本报告无论以何种方式取得，仅供《金融服务与资本市场法》定义的“专业投资者”使用。可从高盛（亚洲）有限责任公司首尔分公司获取有关本报告所研究公司的额外资料。新西兰: Goldman Sachs New Zealand Limited及其关联机构并非1989年新西兰储备银行法定义的“注册银行”或“存款机构”。本研究报告以及本报告的其他形式内容只可分发给2008年财务顾问法案定义的“批发客户”，在事先获得高盛许可的情况下可以有例外。高盛澳大利亚和新西兰的利益披露请参见<https://www.goldmansachs.com/disclosures/australia-new-zealand/index.html>。俄罗斯: 在俄罗斯联邦分发的研究报告并非俄罗斯法律所定义的广告，而是不以产品推广为主要目的的信息和分析，也不属于俄罗斯法律所界定的评估行为。研究报告不构成俄罗斯法律规定的个性化投资建议，并非针对某个具体客户，在报告准备阶段也未分析客户的财务状况、投资特征或风险特征。高盛不对某个客户或任何其他人士基于本报告可能做出的任何投资决策承担责任。新加坡: 高盛（新加坡）私人公司（公司编号: 198602165W）（受新加坡金融管理局监管）为本研究报告承担法律责任，若有由本研究报告所引发或与他人基于本报告相关的任何事宜，请联系高盛（新加坡）私人公司。台湾: 本信息仅供参考，未经允许不得翻印。投资者应当谨慎考虑他们自身的投资风险，投资结果由投资者自行负责。英国: 在英国根据金融市场行为监管局的定义可被分类为私人客户的人士参阅本报告的同时应当参阅高盛以往对本报告研究企业的研究报告，并应当参考高盛国际已经发给这些客户的风险警告资料。该风险警告资料复本，以及本报告中采用部分金融辞汇的解释可向高盛国际索取。

欧盟和英国: 与欧盟委员会实施条例(EU) (2016/958)（欧盟议会和欧盟理事会条例(EU) No 596/2014的补充条款,规定了有关投资建议或其他投资策略的推荐或建议之信息的客观陈述,以及对特定利益或利益冲突进行披露的技术安排应达到的监管技术标准；英国脱离欧盟和欧洲经济区之后该实施条例被纳入英国国内法律法规）第6(2)条相关的披露信息可在<https://www.gs.com/disclosures/europeanpolicy.html>上获取，该网址介绍在处理和投资研究有关的利益冲突时应参照的欧洲政策。

日本: 高盛证券株式会社是在关东财务局注册（注册号：No. 69）的金融工具交易商，同时也是日本证券业协会日本金融期货业协会、第二类金融工具公司协会、日本投资信托协会以及日本投资顾问协会的成员。股票买卖需要缴纳与客户事先约定的佣金及消费税。关于日本证券交易所、日本证券交易商协会或日本证券金融公司所要求的适用的信息披露，请参见与公司有关的法定披露部分。

公司评级、研究范围和相关定义

买入、中性、卖出：分析师建议将评为买入或卖出的股票纳入地区投资名单。一只股票在投资名单中评为买入或卖出由其相对于所属研究范围的总体潜在回报决定。任何未获得买入或卖出评级且拥有活跃评级（即不属于暂停评级、暂无评级、处于初期阶段的生物科技公司、暂停研究或没有研究的股票）的股票均被视为中性评级。每个地区管理着地区强力买入名单，该名单选自各地区投资名单上评级为买入的股票，以总体潜在回报规模和/或实现回报的可能性为主要依据确立各自研究范围内的投资建议。将股票加入或移出此类强力买入名单，由各地区的投资评估委员会或其他指定委员会进行管理，并不意味着分析师对这些股票的投资评级发生了改变。

总体潜在回报：代表当前股价低于或高于一定时间范围内预测目标价格的幅度，包括所有已付或预期股息。分析师被要求对研究范围内的所有股票给出目标价格。总体潜在回报、目标价格及相关时间范围在每份加入投资名单或重申维持在投资名单的研究报告中都有注明。

研究范围：每个研究范围的所有股票名单可登陆<https://www.gs.com/research/hedge.html>通过主要分析师、股票和研究范围进行查询。

暂无评级(NR)：在高盛于涉及该公司的一项合并交易或战略性交易中担任咨询顾问时，或由于高盛参加一项交易而存在法律、监管或政策的限制时，并在某些其他情况下，投资评级、目标价格和盈利预测（如相关）根据高盛的政策被移除。处于初期阶段的生物科技公司：在该公司尚未拥有已通过II期临床试验的药品、治疗方案或医疗设备，或者尚未获得经销II期临床试验完成后药品、治疗方案或医疗设备的执照的情况下，根据高盛政策，不给予投资评级和目标价格。暂停评级(RS)：由于缺乏足够的基础去确定投资评级或价格目标，我们已经暂停对这种股票给予投资评级和目标价格。此前对这种股票作出的投资评级和目标价格(如有的话)将不再有效，因此投资者不应依赖该等资料。暂停研究(CS)：我们已经暂停对该公司的研究。没有研究(NC)：我们没有对该公司进行研究。

全球产品；分发机构

高盛全球投资研究部在全球范围内为高盛的客户制作并分发研究产品。高盛分布在其全球各办事处的分析师提供行业和研究，以及宏观经济、货币、商品及投资组合策略的研究。本研究报告在澳大利亚由Goldman Sachs Australia Pty Ltd (ABN 21 006 797 897) 分发；在巴西由Goldman Sachs do Brasil Corretora de Títulos e Valores Mobiliários S.A.分发；Public Communication Channel Goldman Sachs Brazil: 0800 727 5764和/或contatogoldmanbrasil@gs.com。工作日（假期除外）上午9点至下午6点。Canal de Comunicação com o Público Goldman Sachs Brasil: 0800 727 5764 e/ou contatogoldmanbrasil@gs.com. Horário de funcionamento: segunda-feira à sexta-feira (exceto feriados), das 9h às 18h; 在加拿大由Goldman Sachs & Co. LLC 分发；在香港由高盛（亚洲）有限责任公司分发；在印度由高盛（印度）证券私人有限公司分发；在日本由高盛证券株式会社分发；在韩国由高盛（亚洲）有限责任公司首尔分公司分发；在新西兰由Goldman Sachs New Zealand Limited 分发；在俄罗斯由高盛OOO 分发；在新加坡由高盛（新加坡）私人公司（公司号：198602165W）分发；在美国由高盛集团分发。高盛国际已批准本研究报告在英国分发。

高盛国际（由审慎监管局授权并接受金融市场行为监管局和审慎监管局的监管）已批准本研究报告在英国分发。

欧洲经济区：Goldman Sachs Bank Europe SE是一家在德国注册成立的信贷机构，在单一监管机制下接受欧洲央行的直接审慎监督，在其他方面接受德国联邦金融监管局(Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin)和德国联邦银行的监督，由该机构向欧洲经济区内分发研究报告。

一般性披露

本研究报告仅供我们的客户使用。除了与高盛相关的披露，本研究报告是基于我们认为可靠的目前已公开的信息，但我们不保证该信息的准确性和完整性，客户也不应该依赖该信息是准确和完整的。报告中的信息、观点、估算和预测均截至报告的发表日，且可能在不事先通知的情况下进行调整。我们会适时地更新我们的研究，但各种规定可能会阻止我们这样做。除了一些定期出版的行业报告之外，绝大多数报告是在分析师认为适当的时候不定期地出版。

高盛是一家集投资银行、投资管理和证券经纪业务于一身的全球性综合服务公司。全球投资研究部所研究的大部分公司与我们保持着投资银行业务和其它业务关系。美国证券经纪交易商高盛是SIPC(<https://www.sipc.org>)的成员。

我们的销售人员、交易员和其它专业人员可能会向我们的客户及自营交易部提供与本研究报告中的观点截然相反的口头或书面市场评论或交易策略。我们的资产管理部门、自营交易部和投资业务部可能会做出与本报告的提议或表达的意见不一致的投资决策。

本报告中署名的分析师可能已经与包括高盛销售人员和交易员在内的我们的客户讨论，或在本报告中讨论交易策略，其中提及可能会对本报告讨论的证券市场价格产生短期影响的推动因素或事件，该影响在方向上可能与分析师发布的股票目标价格相反。任何此类交易策略都区别于且不影响分析师对于该股的基本评级，此类评级反映了某只股票相对于报告中描述的研究范围内股票的回报潜力。

我公司及其关联机构、高级职员、董事和雇员，除法规及高盛的制度所禁止的情况外，将不时地对本研究报告所涉及的证券或衍生工具持有多头或空头头寸，担任上述证券或衍生工具的交易对手，或买卖上述证券或衍生工具。

在高盛组织的会议上的第三方演讲嘉宾（包括高盛其它部门人员）的观点不一定反映全球投资研究部的观点，也并非高盛的正式观点。

在此提到的任何第三方，包括销售人员、交易员和其它专业人士或其家庭成员，可能持有本报告提及的且与本报告分析师所表达的观点不一致的产品头寸。

在任何要约出售股票或征购买股票要约的行为为非法的司法管辖区内，本报告不构成该等出售要约或征购买要约。本报告不构成个人投资建议，也没有考虑到个别客户特殊的投资目标、财务状况或需求。客户应考虑本报告中的任何意见或建议是否符合其特定状况，以及(若有必要)寻求专家的意见，包括税务意见。本报告中提及的投资价格和价值以及这些投资带来的收入可能会波动。过去的表现并不代表未来的表现，未来的回报也无法保证，投资者可能会损失本金。外汇汇率波动有可能对某些投资的价值或价格或来自这一投资的收入产生不良影响。

某些交易，包括牵涉期货、期权和其它衍生工具的交易，有很大的风险，因此并不适合所有投资者。投资者可以向高盛销售代表取得或通过<https://www.theocc.com/about/publications/character-risks.jsp>和https://www.goldmansachs.com/disclosures/cftc_fcm_disclosures取得当前期权和期货的披露文件。对于包含多重期权买卖的期权策略结构产品，例如，期权差价结构产品，其交易成本可能较高。与交易相关的文件将根据要求提供。

全球投资研究部提供的不同服务层级：根据您接收沟通信息的频率和方式的个人偏好、您的风险承受能力、投资重心和视角（例如整体市场、具体行业、长线、短线）、您与高盛的整体客户关系的规模和范围、以及法律法规限制等各种因素，高盛全球投资研究部向您提供的服务层级和类型可能与高盛提供给内部和其他外部客户的服务层级和类型有所不同。例如，某些客户可能要求在关于某个证券的研究报告发表之时收到通知，某些客户可能要求我们将内部客户网上提供的分析师基本面分析背后的某些具体数据通过数据流或其它途径以电子方式发送给他们。分析师基本面研究观点（如股票评级、目标价格或盈利预测大幅调整）的改变，在被纳入研究报告、并通过电子形式发表在内部客户网上或通过其它必要方式向有权接收此类研究报告的所有客户大范围发布之前，不得向任何客户透露。

所有研究报告均以电子出版物的形式刊登在我们的内部客户网上并向所有客户同步提供。并非所有研究内容都转发给我们的客户或者向第三方整合者提供，高盛也并不对由第三方整合者转发的我们研究报告承担任何责任。如需了解可向您提供的有关一个或多个证券、市场或资产类别的研究报告、模型或其它数据（包括相关服务），请联络您的高盛销售代表或登陆<https://research.gs.com>。

披露信息可以查阅<https://www.gs.com/research/hedge.html>或向研究合规部索取，地址是200 West Street, New York, NY 10282。

高盛版权所有 © 2026年

对于通过这些服务获取的信息，您仅可出于自行使用的目的进行保存、展示、分析、修改、重新排版和打印。未经高盛明确书面同意，您不得将这些信息转售或进行逆向处理从而用于计算或开发任何用于披露和/或营销的指数、或生成任何其他衍生产品或商业产品、数据或其他产品或服务。未经高盛明确书面同意，您不得以任何形式将这些信息的整体或部分向任何第三方发布、传输或复制。上述限制性要求的适用范围包括但不限于：使用、提取、下载或获取这些信息的整体或部分从而用于对机器学习或人工智能系统的训练或微调、或将这些信息的整体或部分提供或复制给任何此类系统作为提示词或输入信息。