

# Knowledge Atlas (2513 HK)

## Sailing toward the vast frontier of AGI

Knowledge Atlas (Zhipu or Z.ai) is a leading independent large language model (LLM) developer in China, with full-stack capabilities spanning foundation-model pre-training and fine-tuning, enterprise deployment, and application development. Per Frost & Sullivan (F&S), it ranked first among China's independent large-model developers and second across all developers in terms of revenue in 2024, with a market share of 6.6%. Our positive view is anchored on Z.ai's: 1) frontier model capability, underpinned by deep research that sustains rapid iteration; 2) strategic foresight in pinpointing monetization scenarios, positioning it to capture opportunities yet to unfold; and 3) a large institutional base (8k+ customers as of Jun 2025), served with tailor-made solutions. China's LLM industry revenue is projected to grow from RMB5.3bn in 2024 to RMB101.1bn by 2030 (63.5% CAGR, per F&S), and we regard Z.ai as one of the best-positioned independent names to capture such market opportunities. Initiate at BUY with target price of HK\$1,503.9.

- Strong capability in frontier-grade models, validated by pricing power.** Z.ai has demonstrated strong capability in developing frontier-grade models, in our view. Built on its proprietary General Language Model (GLM) framework, Z.ai operates a full-stack model portfolio. On aggregate capability, Artificial Analysis Intelligence Index data show that the company's GLM-5.1 achieved a score of 51 — the highest among Chinese models at launch on 8 Apr — and we believe Z.ai has the ability to sustain an industry-leading position in the upcoming model iteration. We read the cumulative 83% API price increase between December 2025 and March 2026 for its flagship models, absorbed without demand compression, as market-based validation of Z.ai's strong model capability.
- Coding-led monetization with runway beyond.** Z.ai identified coding early as a key monetization scenario and now ranks among the global leaders across multiple coding benchmarks. On a three-benchmark composite (SWE-Bench Pro, Terminal-Bench 2.0, and NL2Repo), GLM-5.1 scored 54.9, ranking first among Chinese models at launch and only behind GPT-5.4 and Claude Opus 4.6. Benchmarked against Anthropic, in our view, Z.ai has a credible path along a similar ARR trajectory, but at higher velocity, from a lower base, and at materially lower per-token cost. Our scenario analysis points to a medium-to-long-term global TAM of US\$90.5bn (base) / US\$348.5bn (bull) in developer-facing AI tools alone. In the longer term, we believe Z.ai can penetrate the full range of enterprise AI verticals, extending the runway well beyond.
- Initiate at BUY with a target price of HK\$1,503.9.** Our TP is based on 40.0x 2028E P/S, discounted to 2026E at a 15% WACC, which is justified in our view, by a 199% revenue CAGR over 2025-2028E on rapidly rising token usage. Key catalysts: 1) launch of next-generation frontier models; and 2) better-than-expected ARR growth.

### Earnings Summary

(YE 31 Dec)	FY24A	FY25A	FY26E	FY27E	FY28E
Revenue (RMB mn)	312	724	3,079	8,372	19,326
YoY growth (%)	150.9	131.9	325.1	171.9	130.8
Operating profit (RMB mn)	(2,538.4)	(3,786.8)	(4,847.5)	(4,934.6)	(3,022.1)
Net profit (RMB mn)	(2,956.5)	(4,698.2)	(4,802.4)	(4,883.4)	(2,968.6)
EPS (Adjusted) (RMB cents)	(1,542.18)	(1,990.28)	(918.84)	(904.46)	(465.22)
P/S (x)	1,292.0	557.3	131.1	48.2	20.9

Source: Company data, Bloomberg, CMBIGM estimates

**BUY (Initiate)**

**Target Price** HK\$1503.90  
**Up/Downside** 43.5%  
**Current Price** HK\$1048.00

### China Software & IT Services

**Saiyi HE, CFA**  
 (852) 3916 1739  
 hesaiyi@cmbi.com.hk

**Ye TAO, CFA**  
 (852) 3850 5226  
 franktao@cmbi.com.hk

**Shuyin GUO**  
 (852) 3916 3716  
 guoshuyin@cmbi.com.hk

**Wentao LU, CFA**  
 luwentao@cmbi.com.hk

### Stock Data

Mkt Cap (HK\$ mn)	467,243.6
Avg 3 mths t/o (HK\$ mn)	1,020.5
52w High/Low (HK\$)	1618.00/8.00
Total Issued Shares (mn)	445.8

Source: FactSet

### Shareholding Structure

Lhasa Junqi Enterprise Management	12.3%
Tianjin Sankuai Technology Co LTD	7.8%

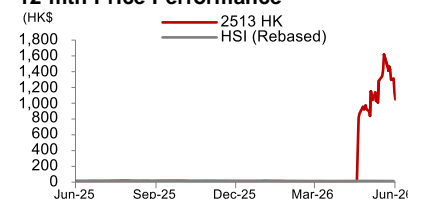
Source: HKEx

### Share Performance

	Absolute	Relative
1-mth	13.5%	22.8%
3-mth	10,862.3%	11,559.4%
6-mth	7,401.8%	7,750.0%

Source: FactSet

### 12-month Price Performance



Source: FactSet

## Contents

<b>Knowledge Atlas (2513 HK) - Sailing toward the vast frontier of AGI</b>	<b>1</b>
<b>Investment thesis</b>	<b>3</b>
China's LLM market is scaling rapidly, with Z.ai being the leading independent provider by revenue	3
Z.ai: frontier model capability with clear competitive edges	3
Sound judgment in picking monetization scenarios—with more to come	3
Benchmarking against global peers suggests considerable room for growth	3
<b>Z.ai — A leading independent large-model developer with full-stack capabilities and deep competitive moats</b>	<b>4</b>
Monetization through an integrated MaaS platform with two primary delivery formats	4
Full-stack model portfolio expands addressable demand and strengthens platform economics	6
Z.ai's model capability stands among China's top tier, with competitiveness extending to the global frontier	8
Open-source ecosystem and developer engagement drive adoption flywheel	10
Safety credentials and hardware compatibility deepen enterprise defensibility	11
Research depth and technical continuity support sustained model iteration	11
<b>China's LLM market: in initial but rapid development stage with strong commercialization prospects</b>	<b>12</b>
China's AI market is at an inflection point, with LLMs at the core	12
Enterprise demand is the primary monetization driver	13
Cloud MaaS to drive breadth of adoption	14
Value chain expanding beyond text into multimodal and agentic workflows	15
Z.ai ranked first among independent LLM providers in China in 2024	15
Developer-side market sizing: Z.ai has large potential to further penetrate	17
Benchmarking Against Anthropic: Sizable Runway to Drive ARR Growth	20
<b>Business segment analysis and financial forecast</b>	<b>23</b>
Open platform and API (64.9% of 2026E revenue)	24
On-premise deployment (35.1% of 2026E revenue)	24
Margin analysis	25
<b>Valuation: HK\$1,503.9 per share based on 40.0x 2028E PS</b>	<b>27</b>

## Investment thesis

### China's LLM market is scaling rapidly, with Z.ai being the leading independent provider by revenue

Per F&S, China's LLM industry revenue is projected to grow from RMB5.3bn in 2024 to RMB101.1bn by 2030, a 63.5% CAGR. Three structural features define the opportunity, in our view: 1) enterprise-led monetization; 2) expanding usage, as applications evolve from text generation toward multimodal and agentic workflows, raising consumption per enterprise; and 3) dual deployment, with on-premise anchoring near-term revenue among large and regulated customers while cloud MaaS broadens adoption among SMEs, developers, and innovation teams. Z.ai sits at the front of this market: per F&S, it ranked first among independent providers and second overall in 2024, with 6.6% of market share in terms of revenue.

### Z.ai: frontier model capability with clear competitive edges

Built on its proprietary GLM framework, Z.ai has assembled a full-stack model portfolio organized around three human-cognition analogues: deep thinking, cognition, and tool use. On aggregate capability, GLM-5.1 scored 51 on the Artificial Analysis Intelligence Index (at launch on 8 Apr 2026)—the highest among Chinese models and only modestly below the U.S. frontier cluster at 52–57 (Fig. 6). We read pricing and demand as market-based validation: GLM-5.1 sits toward the high end of Chinese foundation models on a blended API basis, and Z.ai lifted API pricing by a cumulative 83% between Dec 2025 and Mar 2026, yet demand has held firm.

Z.ai's competitive position rests on four interrelated advantages, in our view: 1) a full-stack, increasingly diversified model portfolio, underpinned by deep research and technical continuity that sustain rapid iteration; 2) a commercialization model well-aligned with enterprise demand, serving a broad customer base; 3) a research-led ecosystem spanning open source, developer communities, and broad hardware compatibility; and 4) safety credentials that strengthen enterprise defensibility.

### Sound judgment in picking monetization scenarios—with more to come

Z.ai showed foresight in targeting coding as a key monetization scenario and has become a global leader in the field. On coding, GLM-5.1 has reached the global frontier: it ranked first worldwide on SWE-Bench Pro (Fig. 8) and third globally on a broader three-benchmark composite (SWE-Bench Pro, Terminal-Bench 2.0, and NL2Repo) as of 8 Apr 2026 (Fig. 9). Our scenario analysis points to a structural runway in China's developer-facing AI coding subscription market (Figs. 19 & 22), where we expect Z.ai to sustain leadership by leveraging these advantages. Beyond coding, continued gains in model intelligence position it to broaden monetization further, in our view.

### Benchmarking against global peers suggests considerable room for growth

We view Anthropic as the most analytically useful comparable for Z.ai, given the shared focus on coding, tool use, and agentic workloads. The comparison points to a credible path for Z.ai along a similar ARR trajectory—at higher velocity, from a lower base, and at materially lower per-token cost. On capability, we believe the residual gap to Anthropic is narrower than the pricing gap. We forecast Z.ai's total revenue to grow at a 199% CAGR over 2025-2028E, reaching RMB19.3bn in 2028E, driven by rapid token-usage growth and potential price increases as model intelligence improves further. We value Z.ai at RMB584.5bn (HK\$1,503.9/share), on 40.0x 2028E P/S discounted to 2026E at a 15% WACC. The 40.0x target sits at a premium to the 12.0x peer average, which in our view is justified by Z.ai's 199% revenue CAGR versus the peer average of 77% (Fig. 35).

## Z.ai — A leading independent large-model developer with full-stack capabilities and deep competitive moats

Z.ai is a full-stack AI company focusing on the development and commercialization of LLMs, with capabilities spanning the entire stack — from foundation model pre-training and fine-tuning to enterprise deployment and application development. Founded in 2019 as a spin-off from Tsinghua University's Knowledge Engineering Group (KEG), the company has become one of China's leading independent LLM developers, serving over 8,000 institutional customers and a developer community of more than 4mn as of 2025. Per F&S, Z.ai ranked first among China's independent large-model developers and second overall by revenue in 2024. Z.ai was the first Chinese AI company to train a model exceeding 100bn parameters (GLM-130B, 2022) and began commercialization in 2021, approximately two years ahead of the broader Chinese LLM industry. Its models were embedded in ~80mn end-user devices — spanning smartphones, PCs, and smart vehicles — as of 1H25, the highest penetration among independent providers.

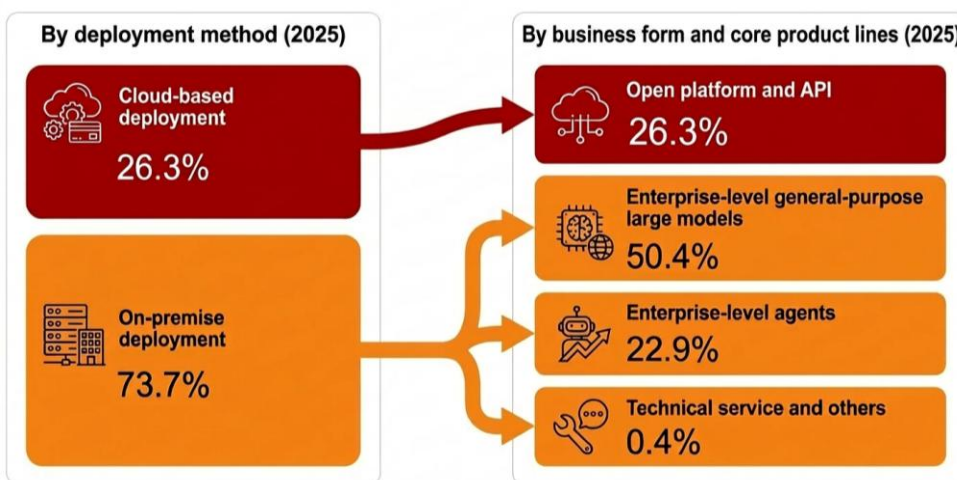
In our view, Z.ai's competitive positioning is supported by four interrelated advantages: 1) a full-stack and increasingly diversified model portfolio, underpinned by deep research capabilities and technical continuity that supports sustained model iteration; 2) a commercialization model well-aligned with enterprise demand; 3) a research-led ecosystem spanning open source, developer communities, and broad hardware compatibility; and 4) safety credentials and hardware compatibility that strengthen enterprise defensibility.

### Monetization through an integrated MaaS platform with two primary delivery formats

Z.ai monetizes through an integrated MaaS platform with two primary delivery formats: on-premise deployment (software licensing and services) for customers with strict data security, compliance, or customization requirements, and cloud-based deployment (API, token-based billing) for lighter integration and greater scalability.

Starting from 2025, Z.ai updated its revenue disclosure from the legacy deployment-based framework to a more granular business-line structure. The "cloud-based deployment" category maps one-to-one to the new "Open Platform and API" segment, while the "on-premise deployment" pool has been disaggregated into three sub-segments: enterprise-level general-purpose large models, enterprise-level agents, and technical services and others.

**Figure 1: Mapping between deployment method and business form and core product lines**



Source: Company data, CMBIGM

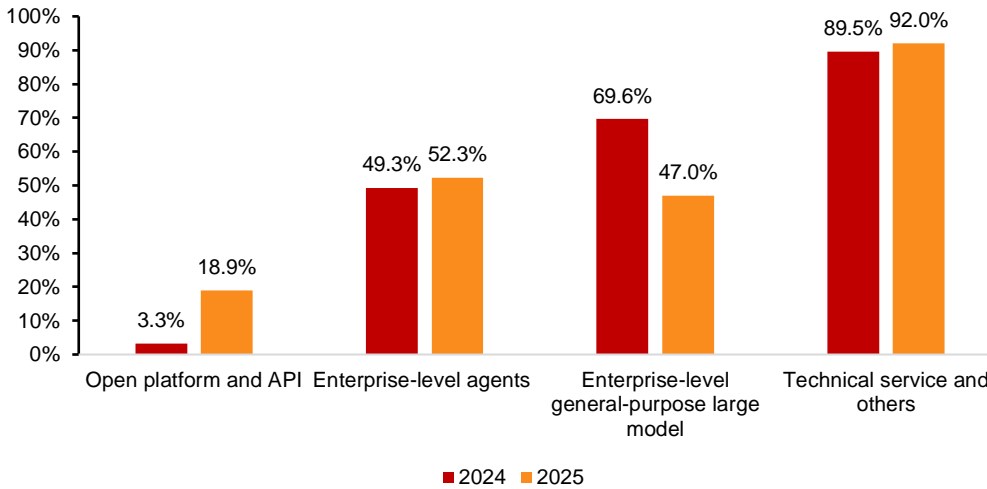
Note: figure is generated by Nano Banana

Z.ai's total revenue reached RMB724.3mn (+131.9% YoY) in 2025, with group-level gross margin compressing from 56.3% in 2024 to 41.0% in 2025 — a mix-driven decline that management attributed to the rising share of cloud deployment and a step-up in on-premise delivery costs.

1. **Open Platform and API (RMB190.4mn in 2025, 26.3% of revenue, +292.6% YoY; GPM 18.9%, +15.6ppts YoY):** Revenue growth was driven by explosive token consumption on Z.ai's BigModel platform, fueled by the rapid adoption of coding and agentic workflows, and we expect this to be further amplified by two successful subscription-based products in 2026: the Coding Plan (242,000+ paying developers as of March 2026, 15x token growth in six months) and the Claw Plan (400,000+ subscribers within 20 days of its launch on 16 March 2026).  
On profitability, the segment's gross profit margin improved sharply from 3.3% in 2024 to 18.9% in 2025, driven by improved model inference efficiency, economies of scale from expanding compute capacity, and API price increases without observable demand deterioration.
2. **Enterprise-level agents (RMB165.7mn in 2025, 22.9% of revenue, +248.8% YoY; GPM 52.3%, +3.0ppts YoY):** Revenue growth was driven by surging enterprise demand for agent-based solutions over basic model access, as businesses began automating complex workflows. Management views agent monetization as still in its early stages in terms of revenue contribution, but believes the category has reached an inflection point for volume growth, with long-horizon task execution expected to emerge as a more meaningful growth driver over time. Gross profit margin for the segment expanded modestly by 3.0ppts, from 49.3% in 2024 to 52.3% in 2025.
3. **Enterprise-level general-purpose large models (RMB365.7mn in 2025, 50.4% of revenue, +70.5% YoY; GPM 47.0%, -22.6ppts YoY):** Revenue growth was supported by continued strong traction in regulated sectors — including financial services, government, and public services — where data security, local control, regulatory compliance, and internal system integration remain key priorities.  
Gross profit margin for the segment contracted by 22.6ppts YoY, from 69.6% in 2024 to 47.0% in 2025. Management attributed this compression to a phased increase in delivery costs required to meet increasingly diverse and complex customer deployment demands.
4. **Technical services and others (RMB2.5mn in 2025, 0.4% of revenue, +31.6% YoY; GPM 92.0%, +2.5ppts YoY):** Revenue growth was primarily driven by increased customer demand. Gross profit margin expanded slightly by 2.5ppts, from 89.5% in 2024 to 92.0% in 2025, remaining elevated given the asset-light nature of the business.

Z.ai's revenue mix in 2025 reflects a business in transition: Open Platform and API is scaling rapidly with improving margins, and management has positioned it as the company's primary revenue growth driver over the next two to three years. Enterprise-level agents and general-purpose large models serve as strategic entry points to capture enterprise demand and funnel customers toward standardized, high-value API consumption.

**Figure 2: Z.ai: GPM profile by different business forms and core product lines**



Source: Company data, CMBIGM

**Full-stack model portfolio expands addressable demand and strengthens platform economics**

Based on its proprietary GLM framework, Z.ai has built a full-stack model portfolio organized around three human-cognition analogues. **Deep Thinking** is anchored by the foundation model GLM-5.1, complemented by reflection and rumination models (GLM-Z1 and GLM-Z1-Rumination). **Cognition** covers multimodal understanding and generation, including CogView (image), CogVideoX-3 (video), GLM-4.6V (visual reasoning), GLM-Realtime (real-time video), and GLM-4-Voice (end-to-end voice). **Tool Use** encompasses AI agents for autonomous task completion — AutoGLM, AutoGLM-Rumination, and CoCo (for enterprise use) — as well as the coding model CodeGeeX.

**Figure 3: Z.ai operates a three-layer, full-stack model portfolio**

Human faculties		Z.ai's models and agents		
Deep thinking	Foundation Models GLM-5.1	Reflection and rumination models	GLM-Z1	reflection
			GLM-Z1-Rumination	rumination
Cognition		Multimodal models	CogView	image generation
			GLM-4.6V	visual comprehension and reasoning
			CogVideoX-3	video generation
			GLM-Realtime	real-time video call
			GLM-4-Voice	end-to-end voice model
Tool use		AI agents	AutoGLM	"from chat to act" – agent for autonomous mission completion
			AutoGLM – Rumination	"thinking while working" – agent for autonomous mission completion with deep thinking capabilities
			CoCo	enterprise agent
	Coding models	CodeGeeX	coding	

Source: Company prospectus, CMBIGM

This breadth gives Z.ai a structural advantage over more narrowly positioned peers, in our view, allowing the company to serve a wider range of workloads through a unified model family and commercialization framework. Z.ai discloses nine application scenarios spanning consumer electronics, office software, recruitment, AIGC content creation, automotive intelligent cockpits, retail and consumer goods, financial services and diversified conglomerates, academic research, and consumer applications — illustrating both the breadth of vertical coverage and the variety of capability combinations being deployed in production.

**Figure 4: Z.ai's full-stack model portfolio translates into cross-industry deployments**

Industry vertical	Client/Partner	Deployment summary
Consumer electronics	A global leading consumer electronics manufacturer	AutoGLM agent embedded in a flagship smartphone series, enabling system-level function calls, cross-app operations, and multimodal social-media content generation from photo-album images via natural-language commands.
Office software	Kingsoft Office (WPS AI)	GLM models integrated across the WPS suite to upgrade content quality (autonomous outline-to-slide-to-speech generation from a single theme) and replace scenario-specific small models, improving scalability and lowering R&D and maintenance costs.
Recruitment services	Zhaopin.com	End-to-end AI recruiting assistant: conversational requirement capture, resume screening, candidate-job matching, and post-interview evaluation for employers; AI-generated and polished resumes for job seekers.
Content Creation / AIGC	Nieta	Intelligent short-video generation agent (launched April 2025) that converts a natural-language idea into a complete short video with visuals, narration, and background music — lowering technical barriers for everyday creators.
Automotive	A high-end intelligent EV manufacturer	Intelligent cabin system upgraded from rigid command execution to multi-turn guided dialogue with emotional intelligence, persona switching, and dynamic content generation (jokes, stories, voice-controlled trivia games).
Retail / Consumer goods	Mengniu Dairy	Consumer-facing AI nutritionist "Mengmeng" delivering personalized nutrition and health consultation, with an embedded AI Health Planner offering adaptive goal-setting, real-time progress tracking, and proactive coaching — strengthening brand differentiation in health and wellness.
Financial & diversified conglomerate	A China-based financial-services-led conglomerate	Three-pronged deployment: 1) financial capital-management LLM achieving >50% AI automation in regulatory Q&A for the commercial banking arm; 2) transportation and logistics LLM covering order management, route planning, customer service, and equipment O&M; 3) "Document + AI" employee knowledge base built on the CoCo enterprise agent, integrating "knowledge base + workflow + agent."
Public Service / Academic research	AMiner	GLM-powered scientific intelligence platform offering natural-language academic search, scholar profiling, AI-assisted literature analysis, automatic summarization, interactive Q&A, and writing/translation tools across millions of academic papers.
Consumer application	Zhipu QingYan	Z.ai's flagship consumer chatbot powered by the GLM series, supporting general-purpose Q&A (with web search, image interpretation, document reading, and GLM-Z1-Rumination-driven deep reasoning), multimodal video/image generation, and a no-code custom-agent platform hosting 800,000+ active user-created agents.

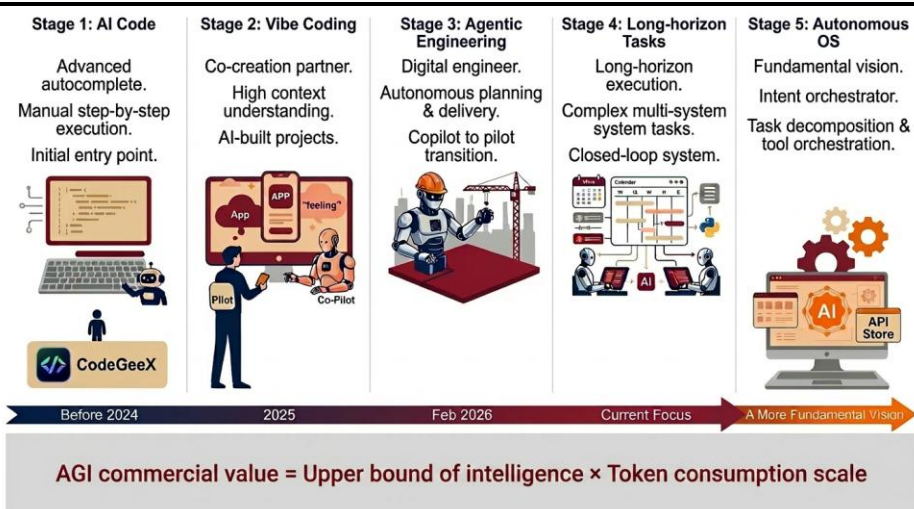
Source: Company data, CMBIGM

Z.ai sees that AI Coding domain is where large model capabilities are evolving the fastest, and the GLM family has tracked each major paradigm shift in AI-assisted software development and enterprise AI more broadly, evolving through four stages to date:

- **Stage 1 - AI Coding:** Autocomplete-style assistance, anchored by the early CodeGeeX release.
- **Stage 2 - Vibe Coding:** Intent-driven development, powered by GLM-4.5/4.6/4.7 across 2025 — a cycle during which API ARR scaled approximately 10x.
- **Stage 3 - Agentic Engineering:** Marked by the Feb 2026 launch of GLM-5.
- **Stage 4 - Long-Horizon Tasks:** The most recent development, anchored by GLM-5.1 in Apr 2026, which extends goal coherence across multi-day, multi-system workflows.

Looking ahead, Z.ai's management has identified Autonomous OS — where the model acts as a scheduler of user intent across applications and platforms — as the next technical frontier, central to Z.ai's positioning within the next-generation enterprise AI computing stack.

**Figure 5: Z.ai identifies itself at Stage 4 of the five-stage paradigm shift in enterprise AI domain**



Source: Company data, CMBIGM  
Note: figure generated by Nano Banana

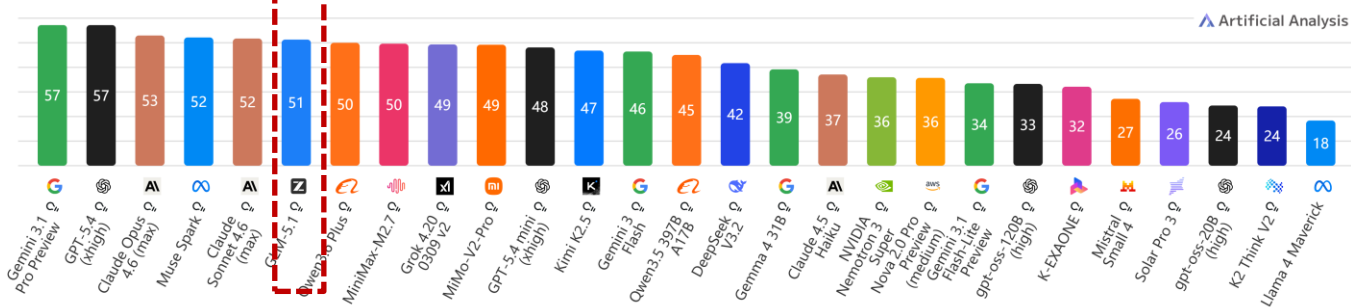
**Z.ai's model capability stands among China's top tier, with competitiveness extending to the global frontier**

On aggregate model capability, Artificial Analysis Intelligence Index data (as of April 8, 2026, the date GLM-5.1 was launched) showed GLM-5.1 scoring 51 — the highest among Chinese models and modestly below the U.S. frontier cluster at 52–57 (Gemini 3.1 Pro Preview, GPT-5.4, and Claude Opus 4.6). The gap narrows on certain task-specific benchmarks: on SWE-Bench Pro, GLM-5.1 scored 58.4, ahead of GPT-5.4, Claude Opus 4.6, and Gemini 3.1 Pro. Across 12 benchmarks spanning reasoning, coding, agentic behavior, tool use, and browsing, the model also demonstrates a relatively balanced capability profile with no pronounced weak spots, according to Z.ai.

**Figure 6: GLM-5.1 scoring 51 at launch, the highest among Chinese models in Artificial Analysis Intelligence Index.....**

**Artificial Analysis Intelligence Index**

Artificial Analysis Intelligence Index v4.0 incorporates 10 evaluations: GDPval-AA, r<sup>2</sup>-Bench Telecom, Terminal-Bench Hard, SciCode, AA-LCR, AA-Omniscience, IFBench, Humanity's Last Exam, GPQA Diamond, CritPt



Source: Artificial Analysis, CMBIGM  
Note: data as of 8 Apr 2026.

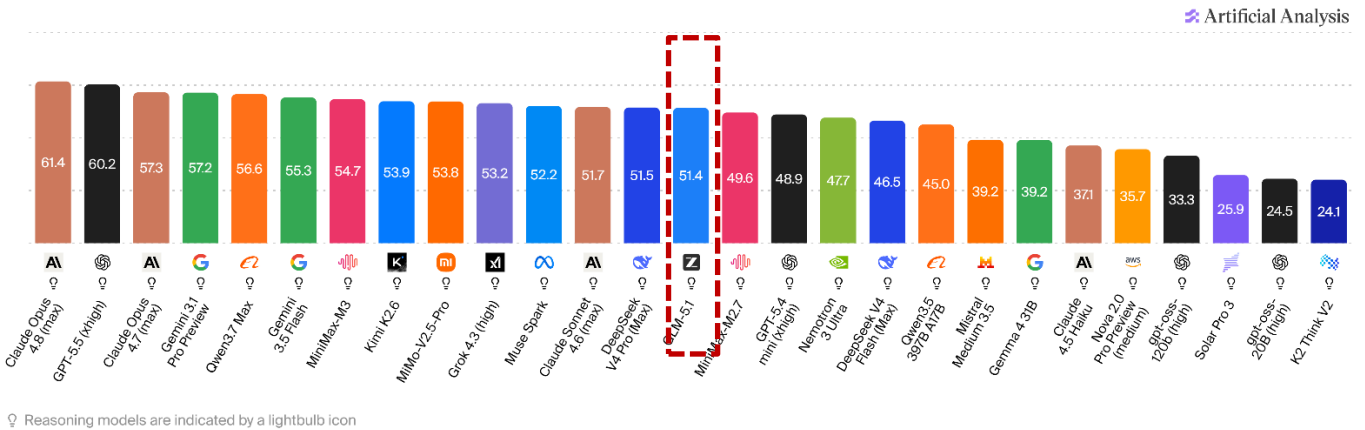
On the latest reading (as of June 9, 2026), GLM-5.1 scored 51.4 and has been overtaken by several newer Chinese releases, including Qwen3.7 Max (56.6), MiniMax-M3 (54.7), Kimi K2.6 (53.9), MiMo-V2.5-Pro (53.8), and DeepSeek V4 Pro (51.5), underscoring the fast shift in the pace of releases. While model rankings have become dynamic amid rapid industry-wide iteration, GLM-5.1's leading position among Chinese models at launch demonstrates Z.ai's ability to push its model capability to China's top tier and the global

frontier. We believe Z.ai has the ability to sustain an industry-leading position in the upcoming model iteration, supported by its deep market research capability.

**Figure 7: GLM-5.1's relative standing shifting since launch amid rapid model iteration in Artificial Analysis Intelligence Index**

Artificial Analysis Intelligence Index

Artificial Analysis Intelligence Index v4.0 incorporates 10 evaluations: GDPval-AA, r<sup>2</sup>-Bench Telecom, Terminal-Bench Hard, SciCode, AA-LCR, AA-Omniscience, IFBench, Humanity's Last Exam, GPQA Diamond, CritPt

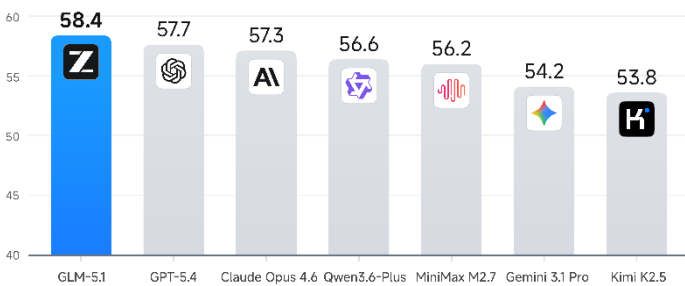


Source: Artificial Analysis, CMBIGM  
Note: data as of 9 Jun 2026.

Regarding coding performance, GLM-5.1 has reached the global frontier at launch: it ranks first worldwide at launch on SWE-Bench Pro (58.4), surpassing GPT-5.4 (57.7) and Claude Opus 4.6 (57.3). On a broader three-benchmark composite (SWE-Bench Pro, Terminal-Bench 2.0, and NL2Repo), GLM-5.1 scored 54.9, ranking first among Chinese models and trailing only GPT-5.4 (58.0) and Claude Opus 4.6 (57.5) by less than 3 points.

**Figure 8: GLM-5.1 ranked first at launch in agentic coding SWE-Bench Pro**

Agentic Coding: SWE-Bench Pro

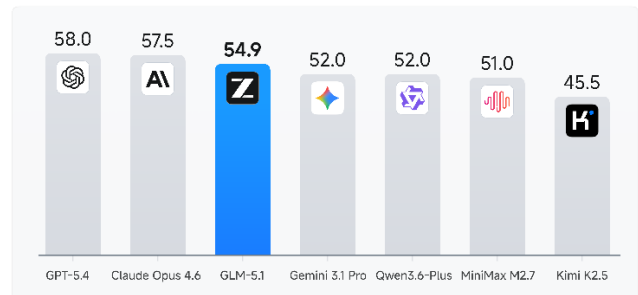


Source: Company data, CMBIGM  
Note: data as of 8 Apr 2026

**Figure 9: GLM-5.1 ranked third globally at launch in coding performance evaluation**

Coding Performance Evaluation

3 Benchmarks: SWE-Bench Pro, Terminal-Bench 2.0, NL2Repo



Source: Company data, CMBIGM  
Note: data as of 8 Apr 2026

A more differentiated capability is GLM-5.1's progress in long-horizon task execution. The model is designed to work autonomously for up to eight hours on a single task, covering the full loop of planning, execution, testing, debugging, and delivery. This places GLM-5.1 among a small group of models globally — and the first from China — to reach this level of sustained autonomous operation. According to the company, GLM-5.1 built a complete Linux desktop environment from scratch within eight hours and delivered a 3.6x geometric-mean speedup on the KernelBench Level 3 optimization benchmark, compared with 1.49x for torch.compile max-autotune.

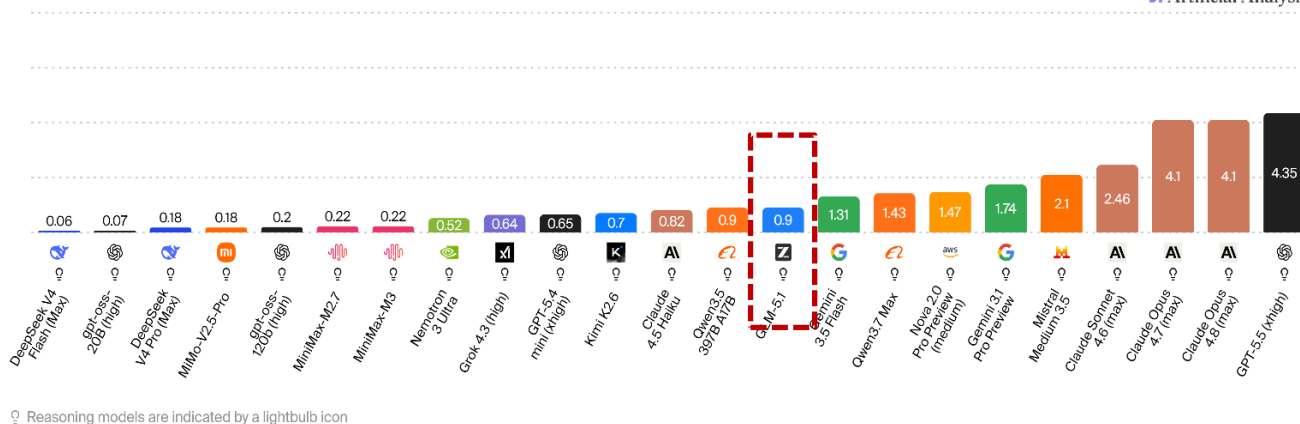
We view pricing and demand dynamics as market-based validation of Z.ai's model capability. GLM-5.1 is priced toward the high end among Chinese foundation models on a blended API basis, and Z.ai raised API pricing by a cumulative 83% between Dec 2025 and Mar 2026, yet demand has remained strong — with call volumes continuing to scale and periodic token stock-outs reported as capacity was absorbed following release. Management indicated on 2025 earnings conference call that the actual demand is running at roughly 1–2x of current daily serving capacity, driven by the capability uplift in the GLM-5 series and broader adoption in higher-value enterprise use cases. At the same time, GLM-5.1 remains meaningfully cheaper than leading global frontier and reasoning models, suggesting potential room for further monetization if its capability gap continues to narrow.

**Figure 10: GLM-5.1 commands premium pricing among domestic peers, while remaining discounted vs global frontier models**

Pricing: Blended Price

Blended at 7:2:1 (cache-input-output) · USD per 1M tokens (blended) · Lower is better

Artificial Analysis



Source: Artificial Analysis, CMBIGM  
 Note: data as of 9 June 2026

### Open-source ecosystem and developer engagement drive adoption flywheel

Open source serves as a key catalyst for Z.ai's ecosystem vitality. Z.ai was among the earliest Chinese AI companies to open-source proprietary large models, and as of 30 June 2025, Z.ai open-sourced more than 50 models with cumulative downloads exceeding 45mn globally. Stanford University research ranked Z.ai as the fifth most popular open-source large-model provider globally in 2023. Open-source usage allows developers to experiment at low cost, generates community feedback, and lowers adoption friction — functioning as the entry point of a funnel that ultimately converts into closed-source API consumption, paid subscriptions, and enterprise deployments.

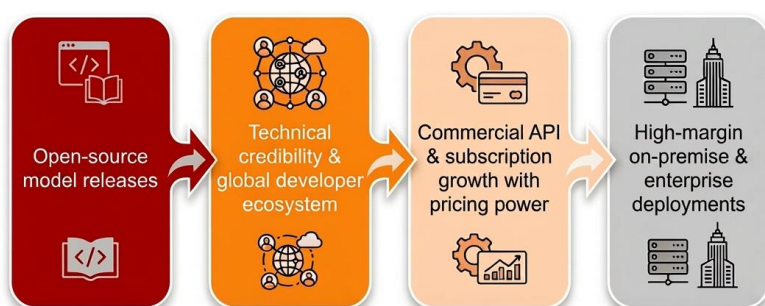
The ecosystem effect is now visible at scale. As of March 2026, Z.ai's open platform served over 4mn of developers and SME customers across 218 countries and regions, with more than 1,000 open-source projects built on its models. Nine of China's top ten internet companies have deeply integrated GLM, and GLM-5 was adopted by ByteDance, Alibaba, Tencent, Meituan, Kuaishou, and Kingsoft's WPS within 24 hours of release — suggesting that the ecosystem has progressed beyond developer experimentation into broad-based commercial adoption.

Behind these numbers is a closed feedback loop that Z.ai labels "MaaS in the Loop". Open-source releases drive developer adoption; developer projects generate usage data and surface edge cases; enterprise deployments provide high-quality demand signals; and each cycle's data feeds back into model refinement, which in turn attracts further developers and enterprise customers.

Monetization runs through three related channels built on top of this ecosystem. Developers who adopt open-source models can upgrade to closed-source commercial APIs — GLM-5, GLM-5-Turbo, and GLM-5.1 — for production-grade access; cloud API ARR reached approximately US\$250mn as of March 2026, up roughly 60x YoY. Layered above the API are subscription products, including the Coding Plan (over 242,000 paying developers across 196 countries as of March 2026) and the higher-priced Claw Plan (over 400,000 subscribers within 20 days of its launch on 16 March 2026).

Two additional initiatives aim to extend this model internationally. From March 2026, Z.ai began hosting closed-source models — including GLM-5-Turbo — on overseas third-party compute platforms under a revenue-sharing arrangement, and has established a national-level MaaS platform and Z-UM AI Lab in Malaysia.

**Figure 11: Z.ai's open-source-to-monetization roadmap**



Source: CMBIGM  
 Note: figure generated by Nano Banana

### Safety credentials and hardware compatibility deepen enterprise defensibility

Safety credentials reinforce enterprise trust, particularly in regulated and cross-border deployments. GLM-4.5 achieved the world's second-lowest and China's lowest hallucination rate on the LLM Hallucination Leaderboard for Retrieval-Augmented Generation (September 2025); GLM-4-9B recorded a 1.3% hallucination rate on HHEM-2.1-Open, among the lowest globally per Stanford's 2025 AI Index Report. Z.ai was the only Chinese AI company to sign the Frontier AI Safety Commitments at the 2024 AI Seoul Summit, and the only Chinese AI company acknowledged in the Future of Life Institute's AI Safety Index 2024 — credentials particularly relevant for regulated-sector buyers.

Hardware compatibility adds a second layer of defensibility. As of June 2025, Z.ai's models were compatible with more than 40 major chip platforms, spanning international and domestic Chinese architectures. Management noted in 2025 annual report that GLM-5's domestic chip adaptation has progressed beyond operator migration into the co-design phase, with customized fusion kernels (Lightning Indexer and FlashComm) enabling inference efficiency on domestic chips comparable to top-tier international hardware.

### Research depth and technical continuity support sustained model iteration

Z.ai originated from Tsinghua University's KEG lab and maintains deep academic ties. As of 30 June 2025, its elite research and advisory teams had published approximately 500 papers with over 58,000 citations, and its ChatGLM work was featured in Nature in May

2024. Z.ai's research credibility is corroborated by independent international benchmarks: GLM-130B (ICLR 2023) was the only Asian model selected in Stanford HELM's foundation model evaluation, with accuracy and toxicity indicators on par with GPT-3 Davinci. More than 74% of employees were engaged in R&D in 1H25, and total R&D expenditure reached RMB3.18bn in 2025, up 44.9% YoY — underscoring a research-heavy organizational structure and sustained commitment to model leadership.

The KEG lineage is reinforced institutionally: Z.ai is a co-founder of the Beijing Key Laboratory of Foundation Models (alongside Tsinghua and Peking University) and a founding partner of the Tsinghua University Foundation Model Research Center, supporting continuous access to frontier research and systematic recruitment of top AI talent. Chief Scientist Zhang Bo is a Chinese Academy of Sciences academician; the core technical leadership — Chairman Liu Debing, CEO Zhang Peng, CAIO Tang Jie, and Co-founder Li Juanzi — averages over 12 years of domain experience. Tang Jie is internationally recognized for his work on ChatGLM (featured in Nature in May 2024) and as the first Chinese recipient of the ACM SIGKDD 2020 Test of Time Applied Science Award. Core technical leaders have published extensively across five major AI conferences across top-tier AI conferences including ICLR, NeurIPS, ICML, KDD, and ACL, providing peer-reviewed credibility and structural visibility in frontier research that few Chinese large-model peers can match.

Beyond publications, Z.ai has built academic infrastructure that functions as a durable moat. Its AMiner platform — a scientific knowledge graph covering ~320mn papers, ~133mn scholars, and ~160mn patents as of 2025 — serves ~10mn users across 220+ countries and publishes the AI 2000 Global Scholar Influence Ranking. The company also co-sponsors the CCF-Zhipu Large Model Fund (inaugurated 2022, funding 26 projects from 20 universities in its first batch) and the SMP-Zhipu Interdisciplinary Fund — channels that seed GLM adoption among future enterprise technical decision-makers.

The technical advances underlying GLM-5 are substantive. According to the company, three innovations underpinned the release: 1) architectural improvements through MuonSplit optimization and MLA-256, achieving stable training while significantly reducing KV-Cache memory usage; 2) the Slime asynchronous reinforcement learning framework, which decouples generation from training and addresses the GPU idle-time bottleneck of traditional synchronous RL; and 3) more than 10,000 verifiable real-world engineering environments for authentic code-challenge learning. These innovations enabled GLM-5-Turbo's OpenClaw targeting and enterprise-grade long-chain task execution.

## **China's LLM market: in initial but rapid development stage with strong commercialization prospects**

China's LLM market is transitioning from early experimentation into scaled enterprise adoption. According to F&S, China's LLM industry revenue is projected to grow from RMB5.3bn in 2024 to RMB101.1bn by 2030, implying a 63.5% CAGR. Three structural features define the opportunity: 1) enterprise-led monetization, as enterprise customers have clearer use cases and higher willingness to pay than consumers; 2) dual deployment demand, with on-premise deployment anchoring near-term revenue among large and regulated customers, while cloud MaaS broadens adoption among SMEs, developers, and innovation teams; 3) expanding model usage, as LLM applications evolve from text generation to multimodal and agentic workflows, increasing potential model consumption per enterprise.

However, the market remains fragmented, with the top five players accounting for only 33% of revenue, though barriers to scaled participation are rising. Z.ai ranks second overall and first among independent providers, with a 6.6% market share in 2024, according to F&S.

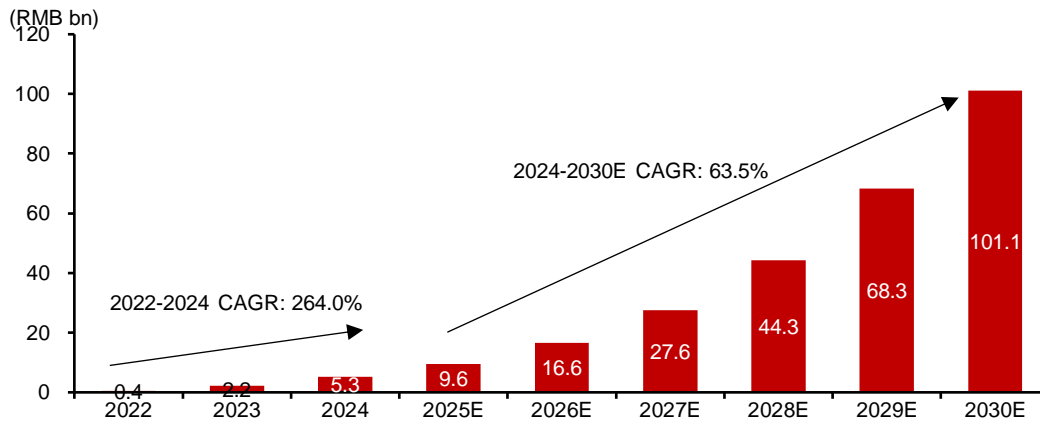
## **China's AI market is at an inflection point, with LLMs at the core**

According to F&S, China's AI market grew from RMB93.7bn in 2022 to RMB160.7bn in 2024, representing a 31.0% CAGR, and is projected to reach RMB993.0bn by 2030 at a 35.5% CAGR — driven by the convergence of compute, algorithms, and data.

Within AI, the industry is undergoing a structural shift from discriminative models — limited to recognition and classification tasks — to generative LLMs capable of multimodal generation, reasoning, and autonomous task execution. LLMs overcome the limitations of legacy discriminative AI in semantic understanding, generalization, and autonomous decision-making, and are widely regarded as the key vehicle for the transition from Artificial Narrow Intelligence toward Artificial General Intelligence.

Against this backdrop, F&S estimates China's LLM market reached RMB5.3bn in 2024, up from RMB0.5bn in 2022, and forecasts growth to RMB101.1bn by 2030 — implying a 63.5% CAGR.

**Figure 12: China's LLM market size & forecast**



Source: F&S, CAICT, CMBIGM

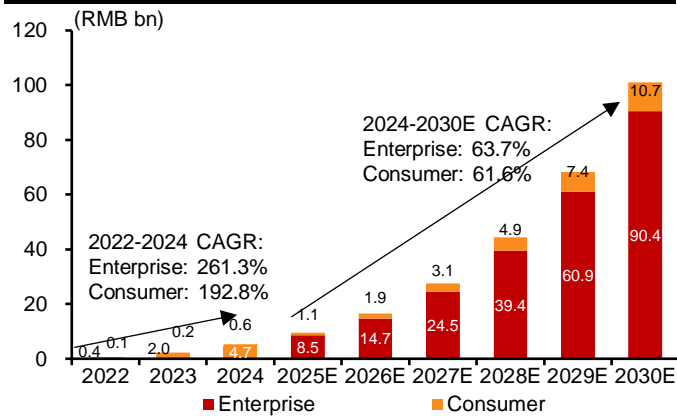
Note: The size of AI market in China represents the market players' revenue derived from the provision of AI solutions.

### Enterprise demand is the primary monetization driver

According to F&S, enterprise customers contributed RMB4.7bn (~89%) of China's 2024 LLM revenue and are expected to reach RMB90.4bn by 2030 at a 63.7% CAGR — outpacing the consumer segment, which is forecast to grow from RMB0.6bn to RMB10.7bn at a 61.6% CAGR. The gap reflects fundamental differences in demand: consumer willingness to pay for LLM-powered applications remains limited, while enterprises have well-defined, high-ROI use cases spanning customer service automation, marketing content generation, productivity and workflow automation, code assistance, and industry-specific knowledge management — all scenarios where LLMs automate text-heavy processes, enable data-driven decisions, and generate contextually relevant content at scale.

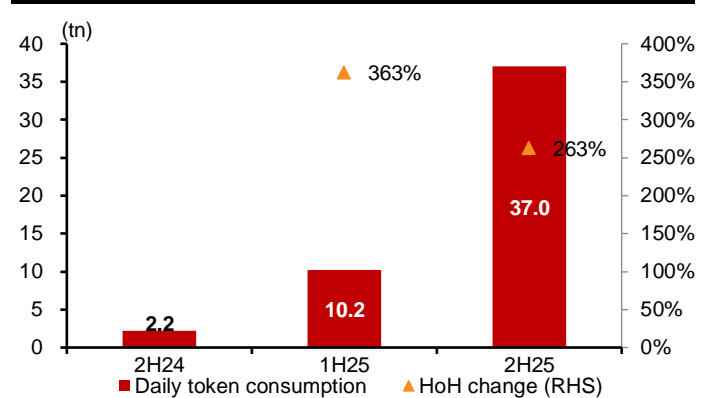
Demand intensity is also accelerating sharply. Based on F&S data, daily enterprise LLM token consumption in China rose from ~2.2tn in 2H24 to 10.2tn in 1H25 and 37.0tn in 2H25. The acceleration in 2H25 was driven primarily by the adoption of agent-based architectures that replace single-turn Q&A with continuous, multi-step task execution — planning → tool calling → execution → verification. The structural implication is that per-enterprise token consumption is scaling independently of new customer additions, creating a compounding revenue pool as existing adopters deepen usage. Financial services, government, and healthcare have been the earliest movers, reflecting their high density of unstructured data and knowledge-intensive workflows.

**Figure 13: China enterprise vs consumer LLM revenue breakdown**



Source: F&S, CAICT, CMBIGM

**Figure 14: China enterprise daily LLM token consumption**



Source: F&S, CMBIGM

Note: Survey of 700 Chinese enterprises across finance, manufacturing, internet, consumer electronics, and automotive sectors. Daily token consumption peaked in December 2025, materially above the 2H25 average.

### Cloud MaaS to drive breadth of adoption

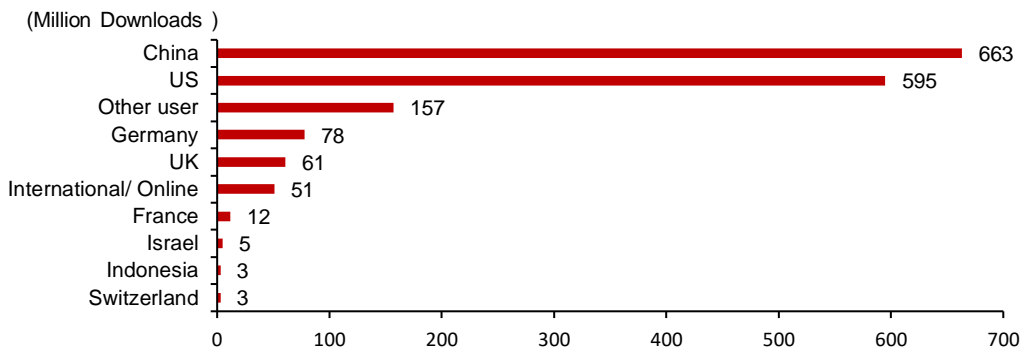
The Chinese enterprise LLM market has crystallized around two complementary deployment modes serving differentiated customer segments.

**On-premise dominates revenue contribution today.** According to F&S, on-premise deployment accounted for RMB3.8bn (~81%) of enterprise LLM revenue in 2024 and is projected to reach RMB69.1bn by 2030 at a 62.2% CAGR, retaining the majority of share throughout the forecast period. The segment is anchored by large enterprises in finance, government, and healthcare, where data sovereignty, compliance, system stability, and industry-specific customization requirements outweigh the cost benefits of cloud delivery. On-premise deployment also enables deep fine-tuning on proprietary data and domain knowledge bases, which materially improves real-world performance in specialized scenarios such as legal document review, financial analytics, and medical diagnostics.

**Cloud MaaS is the faster-growing deployment mode.** F&S projects cloud-based enterprise LLM revenue to increase from RMB0.9bn in 2024 to RMB21.3bn by 2030, implying a 69.4% CAGR and outpacing the overall on-premise market. Based on IDC China data, public-cloud LLM invocations in China reached 536.7tn tokens in 1H25 alone, versus 114tn for all of 2024 — implying nearly 5x growth within six months. Three drivers underpin this trajectory:

- **Radically lower adoption barriers.** Self-deploying a flagship LLM requires substantial upfront capex — a single H100-class GPU costs approximately US\$25,000–40,000, with multi-GPU clusters often required for meaningful workloads. Cloud MaaS converts this capex into opex via pay-as-you-go and tiered pricing, bringing high-quality LLM capabilities within reach of SMEs and innovation teams. According to a F&S, 70% of enterprises choose to deploy or call LLMs through the public cloud, with 71% planning to scale up their public-cloud generative AI services going forward.
- **Agile delivery and rapid iteration.** Cloud APIs enable integration in hours rather than months, with continuous model upgrades pushed by providers — a pace that on-premise deployments structurally struggle to match.
- **Open-source tailwind.** According to F&S, over 80% of enterprises are expected to adopt open-source models going forward. According to Hugging Face, China accounted for ~41% of global open-source model downloads from February 2025 to February 2026 — a dynamic that plays directly to domestic independent providers with strong open-source franchises.

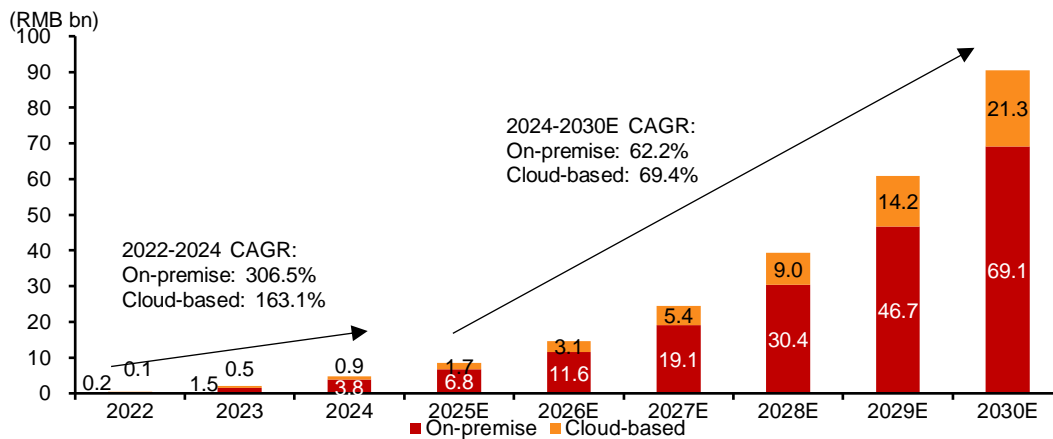
**Figure 15: Hugging Face open-source downloads from Feb 2025 – Feb 2026**



Source: Hugging face, CMBIGM

In our view, on-premise and cloud MaaS are complementary rather than substitutive: on-premise anchors high-value revenue from regulated incumbents, while cloud MaaS drives breadth of adoption across SMEs and rapid-iteration use cases.

**Figure 16: Enterprise LLM revenue by deployment mode, 2022–2030E**



Source: F&S, CAICT, CMBIGM

### Value chain expanding beyond text into multimodal and agentic workflows

China's LLM industry value chain comprises three layers: 1) upstream — infrastructure hardware (GPUs, servers, storage, and networking), infrastructure software such as databases, and training data; 2) midstream — independent and non-independent LLM providers; and 3) downstream — enterprise customers across financial services, internet, education, healthcare, and government, as well as individual consumers.

Two structural shifts are reshaping the value chain. First, modality is expanding well beyond text — manufacturers are integrating image and video analytics for quality inspection, financial institutions are fusing multimodal data for risk assessment and customer interaction, and consumer-facing AI devices such as smartphones, humanoid robots, and AI glasses are driving demand for real-time audio and video generation. Second, agent architectures are becoming the dominant interaction paradigm, with a single business process invoking a model dozen to hundreds of times for planning, tool selection, execution, and verification. In our view, these two shifts imply that competitive positioning will increasingly depend on breadth of modality coverage and platform-level orchestration.

### Z.ai ranked first among independent LLM providers in China in 2024

According to F&S, the LLM market in China remains fragmented, with the top five providers accounting for only 33.2% of 2024 enterprise LLM revenue and no single player having

established clear dominance. Z.ai ranked second overall and first among independent providers, with a 6.6% market share.

**Figure 17: Top 5 LLM providers in China by revenue, 2024**

Ranking	Company	Type	2024 Revenue (RMBbn)	Market share
1	A	Non-independent	0.44	9.4%
2	Z.ai	Independent	0.31	6.6%
3	B	Non-independent	0.30	6.4%
4	C	Non-independent	0.29	6.1%
5	D	Non-independent	0.22	4.7%
Total			1.56	33.2%

Source: F&S, CMBIGM

The market bifurcates into two provider types: independent providers, which are natively built around LLM technology from inception and whose businesses do not compete with enterprise clients' core operations; and non-independent providers, which are affiliated with established platforms extending adjacent advantages — such as search and audio/vision capabilities — into LLM. For enterprise buyers in regulated sectors, particularly financial services and government, vendor selection typically considers factors including data sovereignty, potential overlap between the vendor's affiliated businesses and the customer's own operations, and the vendor's model roadmap priorities. Z.ai's pure-play model-vendor profile means its model development is not subordinated to a parent platform's strategic needs — a factor that management and industry interviews cite as recurring in vendor evaluation by regulated-sector customers.

According to F&S's MaaS benchmark, Z.ai also leads peers in model portfolio breadth, covering all 11 capability dimensions: language, code generation, image generation, video generation, audio generation, real-time video, inference, character, GUI agent, phone/web use, and computer use. In our view, this full-stack coverage is increasingly decisive as agent-based workflows require orchestration across multiple modalities within a single task.

Importantly, Z.ai's MaaS platform should be distinguished from traditional PaaS: while PaaS abstracts infrastructure for general application development and targets developers broadly, MaaS delivers pre-trained foundation models and development toolkits as the product itself, operating at a higher level of abstraction and targeting enterprises and AI-focused developers who do not build models from scratch.

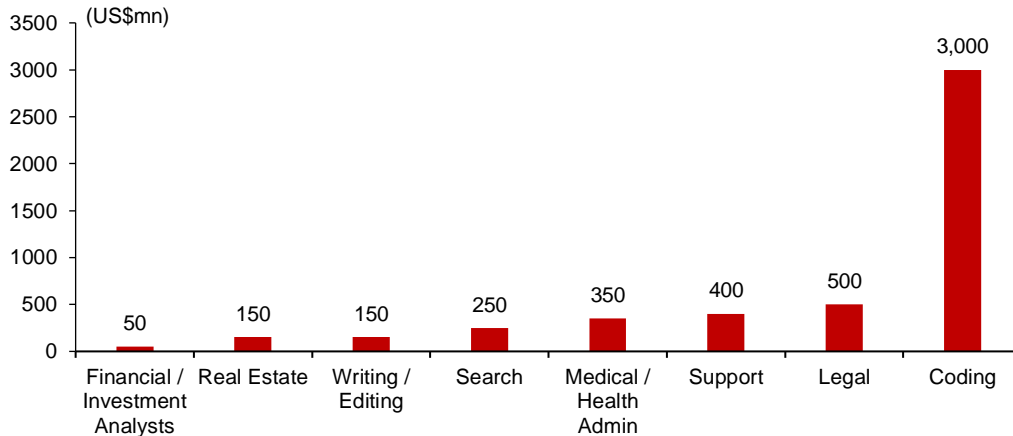
Barriers to entry are rising along four dimensions, which we expect to drive gradual consolidation around scaled incumbents:

- **Technical barriers.** Leading providers maintain a full-spectrum model matrix from edge to flagship — spanning hundreds of billions of parameters and supporting diverse task types — with a positive feedback loop between proprietary pre-training frameworks, client scenario data, and model iteration that is difficult for new entrants to replicate.
- **Flexible business models and delivery.** Serving heterogeneous enterprise needs across data sensitivity, compute infrastructure, and budget requires concurrent support for on-demand, subscription, and one-time deployment — a capability only a handful of providers possess at scale.
- **Ecosystem depth.** Leading vendors have built multi-directional ecosystems spanning open-source developer communities, upstream chip partners for cross-hardware adaptation, and downstream industry clients — creating closed-loop feedback effects that accelerate iteration and raise switching costs.
- **Talent.** Top-tier researchers and engineers with deep LLM expertise remain scarce, with incumbents having secured disproportionate shares of the available talent pool.

## Developer-side market sizing: Z.ai has large potential to further penetrate

Among AI application verticals, coding stands out as the most commercially mature one, with several products having already scaled to billion-dollar ARR. Per a16z (April 2026), annualized revenue across enterprise AI coding tools has reached ~US\$3bn, outpacing all other enterprise AI application categories. While we strongly believe Z.ai has the potential to penetrate the full range of enterprise AI application verticals — all of which offer strong growth runway — we first narrow our initial focus to the developer-facing AI tools segment. In our view, even within this relatively narrow slice of the broader enterprise AI market, Z.ai retains meaningful upside on both penetration and monetization.

**Figure 18: Annualized revenue by enterprise AI application vertical (Apr 2026)**



Source: a16z, CMBIGM

### ■ China Developer AI Subscription TAM: mature-market benchmark implies RMB1,219/month ARPU potential, well above current pricing

Our analysis points to a structural runway in China's developer-facing AI coding subscription market. Applying a mature-market intensity benchmark to local wage data yields an implied per-developer spend ceiling of ~RMB1,219/month even at current level—materially above prevailing price points — supporting our view that the segment remains pre-steady-state, with continued ARPU and volume expansion ahead.

**Multi-tool stacking is now the base case, not the exception.** Industry data consistently shows developers layering multiple AI coding tools rather than consolidating on a single vendor — a dynamic we view as supportive of aggregate subscription spend per seat. NetCorp's *AI-Generated Code Statistics 2026* and Second Talent's *AI Coding Assistant Statistics & Trends 2026* indicate ~59% of developers used three or more AI coding tools in parallel in 1Q25. The Stack Overflow 2025 Developer Survey corroborates this pattern, with adoption rates of 82% for ChatGPT, 68% for GitHub Copilot, 47% for Gemini, and 41% for Claude totalling in excess of 100%, reflecting meaningful overlap. NxCode's 2026 industry guide explicitly identifies "Cursor + Claude Code" and "Cursor + Claude Code + Copilot" as standard stack configurations, reinforcing our view that bundled-tool spend, rather than single-vendor ARPU, is the relevant unit of analysis.

**Benchmarking the U.S.: ~5.5% of median software wage flows to AI tooling.** Per BLS OEWS data (May 2024), the median annual wage for Software and Web Developers, Programmers, and Testers stands at US\$131,450. Assuming a high-end developer subscribes to 2-3 AI tools at a combined ~US\$600/month (US\$7,200/year), implied AI tooling spend equates to ~5.5% of median software-occupation wages. We adopt this ratio as our benchmark for AI subscription intensity in mature markets.

**Applying the framework to China: implied ceiling of RMB1,219/month per developer.** Per the *2025 China Statistical Yearbook*, the average annual wage for professional and technical personnel in the information transmission, software, and IT services sector

reached RMB266,974 in 2024 (~US\$39,000). Applying the 5.5% intensity benchmark yields implied per-developer AI subscription capacity of RMB14,622/year (~US\$2,145), or ~RMB1,219/month. Current market pricing for domestic coding subscriptions sits well below this implied ceiling, which in our view signals that China's developer AI subscription market has yet to reach steady state. We see scope for sustained expansion as pricing, product depth, and multi-tool stacking behaviors converge toward mature-market norms. Applying our per-developer annual AI subscription capacity (RMB14,622, derived earlier) to China's software developer population yields an implied China coding plan TAM of ~US\$20.1bn.

**Figure 19: China developer-side consumer AI subscription TAM**

Step	Item	Value	Formula	Source / Calculation
Derive AI subscription spending ratio from U.S. benchmark				
1	U.S. developer median annual wage	US\$131,450	→A	BLS OEWS, May 2024
	Assumed annual AI subscription spend per developer	US\$7,200	→B	US\$600/month × 12 (2-3 subs)
	→ Implied subscription spend as % of annual wage	5.50%	→C=B / A	= 7,200 / 131,450
Apply ratio to China developer wage to derive per-developer spending capacity				
2	China IT services sector — average annual wage (2024)	~US\$39,000 (RMB266,974)	→D	2025 China Statistical Yearbook
	Applied income-to-subscription ratio	5.50%	→C	From Step 1
	→ Per-developer annual AI subscription capacity	~US\$ 2,145 (RMB14,622)	→E=C × D	US\$39,000 × 5.5%
Scale by China software developer population to derive total TAM				
3	China software developer population (2024)	9.4mn	→F	MIIT
	→ Implied developer-side AI subscription TAM, China	US\$20.1bn	→G=E × F	US\$ 2,145 × 9.4mn

Source: BLS, MIIT, 2025 China Statistical Yearbook, CMBIGM

Note: our USD:RMB exchange rate assumption is 1:6.85; same below

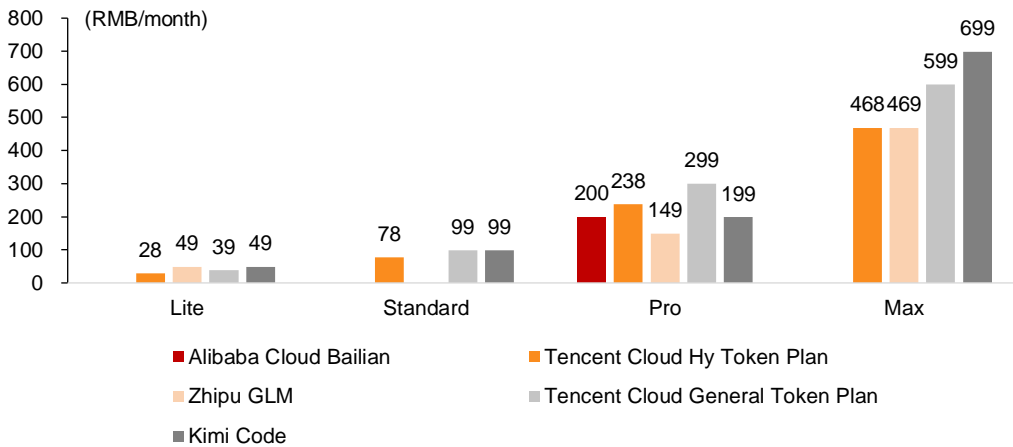
We view Z.ai as a credible share gainer in China's developer AI subscription market, underpinned by three reinforcing pillars:

**1) Pricing headroom at the plan level.** Z.ai's GLM Coding Plan – Max is currently priced at RMB469/month before discounts — a meaningful ~62% discount to our RMB1,219/month implied per-developer ceiling at a current stage. This positioning leaves room for either standalone ARPU expansion or comfortable inclusion within a 2-3 tool stack, consistent with the multi-tool adoption pattern we observe in mature markets. We view this gap as an option on future monetization.

**2) Competitive model capability.** GLM-5 delivers benchmark performance broadly in line with leading domestic and overseas peers on coding-specific evaluations (see Fig. 8&9), neutralizing the "capability gap" objection that has historically weighed on domestic AI model providers, in our view. We see this as a necessary — though not sufficient — condition for sustained share capture.

**3) Demonstrated pricing power.** Most notably, Z.ai implemented an 83% price increase between Dec 2025 and Mar 2026 without observable demand compression. Capacity has remained supply-constrained post-hike, which we read as a tangible signal of inelastic willingness-to-pay within the installed base and demonstrated pricing power of Z.ai.

**Figure 20: China individual coding plan pricing by tier (RMB/month)**



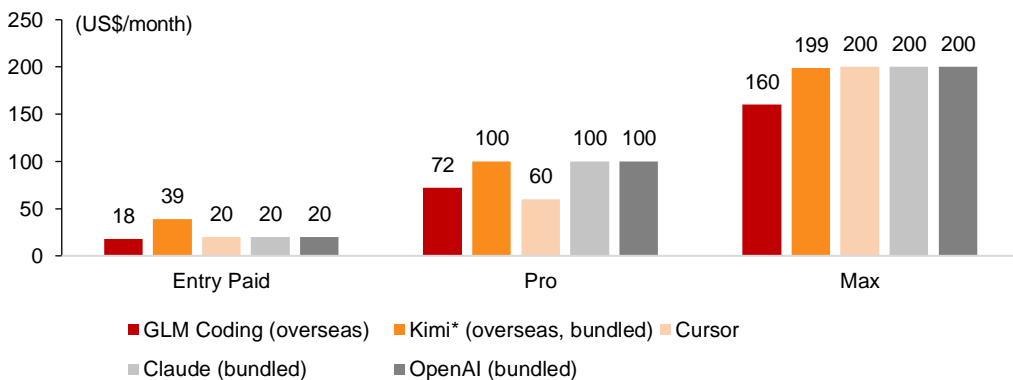
Source: Company website, CMBIGM

AI coding and productivity tools are predominantly consumed in parallel rather than on a substitutive basis, which in our view argues against a strict winner-takes-all market structure. The implication is important: a model provider does not need to be best-in-class across every task to monetize at scale — securing a slot among the 2–3 tools developers routinely select is, in our view, sufficient to capture a meaningful share of wallet.

■ **Overseas ARPU also carries headroom, although China remains the near-term focus**

Z.ai's overseas GLM Coding Plan is currently priced at US\$18/US\$72/US\$160 per month for Lite/Pro/Max, versus overseas peers clustering at US\$10-20 (entry), US\$60-100 (mid), and ~US\$200 (top). The Max tier is the most relevant anchor in our view: with model capability broadly matching leading overseas peers, GLM's US\$160/month implies a ~20% discount to the US\$200 top-end benchmark. We read the gap as latent pricing headroom that could support upward revisions to overseas plan pricing over time, although we note management's near-term focus remains on the domestic opportunity.

**Figure 21: Overseas individual coding plan pricing by tier (USD/month)**



Source: Company website, CMBIGM

Note: Kimi (overseas) also offers a mid-tier plan at US\$39/month

■ **Medium-to-long-term global TAM framing: bear/base/bull cases at US\$6.5bn/US\$90.5bn/US\$348.5bn**

**Scenario analysis frames the monetization opportunity.** Our scenario framework points to a global developer-facing AI subscription TAM of **US\$6.5bn / US\$90.5bn / US\$348.5bn** under bear / base / bull cases, underscoring what we view as a substantial long-duration monetization runway for Z.ai's AI coding franchise.

**ARPU assumptions:** We frame the ARPU axis as follows:

- **Bear case (US\$20/month):** assumes each developer subscribes only to the cheapest entry tier — i.e., minimal monetization depth and no multi-tool stacking.
- **Base case (US\$200/month):** anchored to Anthropic's disclosed enterprise telemetry for Claude Code (US\$150–250 per developer per month), with the midpoint taken as a representative benchmark for mainstream developer spend. We view this as the most empirically grounded of the three anchors.
- **Bull case (US\$600/month):** assumes a power-user configuration of three Max-tier subscriptions in aggregate, consistent with the multi-tool stacking patterns documented earlier in this note.

**Developer-base assumptions: bracketed by professional vs. broader definitions.** We bound the user-base axis using two widely cited industry estimates: 27.0mn (Evans Data, 2024 — full-time professional software developers worldwide) at the low end, and 48.4mn (SlashData, *State of the Developer Nation 2024*, using a broader definition inclusive of part-time and freelance contributors) at the high end. The base case takes the 37.7mn midpoint. Our bear/base/bull cases framing the global TAM at US\$6.5bn/US\$90.5bn/US\$348.5bn.

**Figure 22: Global developer-side AI subscription TAM**

	Bear case	Base case	Bull case
Subscription tier assumption	Entry	Enterprise Avg	Power-User Stack
Monthly price per sub (US\$/mo)	20 (1x Entry Paid Subscription)	200 (Midpoint of Claude Code enterprise's cost range)	600 (3x Max Subscriptions at US\$200/mo)
x Months per year	12	12	12
<b>Annual ARPU (US\$)</b>	<b>240</b>	<b>2,400</b>	<b>7,200</b>
Global developer base (mn)	27.0 (2024 full-time professional software developers worldwide)	37.7 (Midpoint)	48.4 (broader definition including part-time and freelance contributors in 2024)
x Annual ARPU (US\$)	240	2,400	7,200
<b>Implied TAM (US\$ bn)</b>	<b>6.5</b>	<b>90.5</b>	<b>348.5</b>

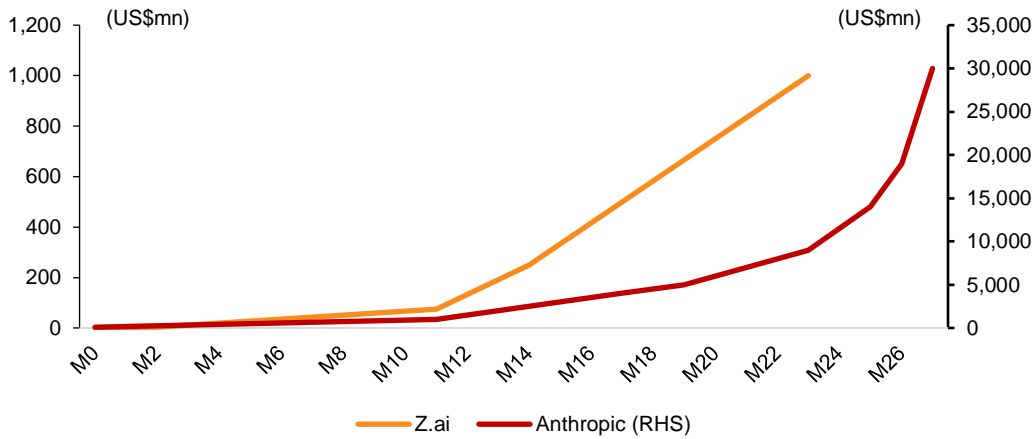
Source: Anthropic, Stack Overflow, Evans Data, CMBIGM

## Benchmarking Against Anthropic: Sizable Runway to Drive ARR Growth

We view Anthropic as the most analytically useful comparable for Z.ai given the shared focus on coding, tool use, and agentic workloads. The benchmarking exercise points to a credible path for Z.ai to traverse a similar ARR trajectory — at higher velocity, from a lower base, and at a materially lower per-token cost.

**ARR trajectory: faster ramp from a lower base.** Anthropic's commercial inflection (M0) occurred in January 2024 at ~US\$87mn in revenue, scaling to ~US\$1bn ARR by year-end. Z.ai's equivalent inflection came 12 months later, in January 2025, from a ~US\$1mn base; ARR has since reached ~US\$250mn as of March 2026, with management targeting approach to US\$1bn by year-end 2026. In our view, the directional read-across is supportive: while Z.ai starts from a smaller revenue base, its growth trajectory appears steeper. We remain positive on Z.ai's ability to achieve its year-end ARR target and believe ongoing model enhancements could further accelerate ARR growth.

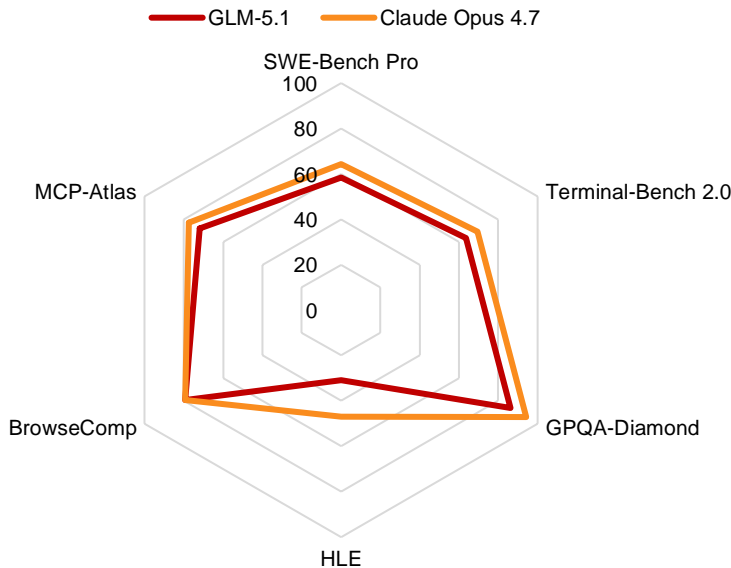
**Figure 23: Z.ai vs. Anthropic: ARR**



Source: Company data, Reuters, The Information, Sacra, CMBIGM  
 Note: The M0 of Z.ai/ Anthropic are respectively Jan 2025/ Jan 2024, under our definition.

**Model capability: residual gap is narrower than the pricing gap.** GLM-5 scores 77.8% on SWE-Bench Verified — the highest among open-source models at launch— versus Claude Opus 4.7 at 87.6%. On SWE-Bench Pro, GLM-5.1 scored 58.4, ahead of GPT-5.4 (57.7%) and Gemini 3.1 Pro (54.2%) among generally available models, though trailing Opus 4.7 at 64.3%. A wider gap persists on graduate-level reasoning (GPQA-Diamond: 86.2% vs. 94.2%). The analytically important point, in our view: for enterprise workloads centered on coding, tool use, and agentic execution, the residual capability gap to Anthropic's frontier is meaningfully smaller than the per-token pricing gap — the empirical underpinning of Z.ai's value-for-performance positioning in the global market.

**Figure 23: Z.ai vs. Anthropic: model capability benchmarks**



Source: Company official website, CMBIGM

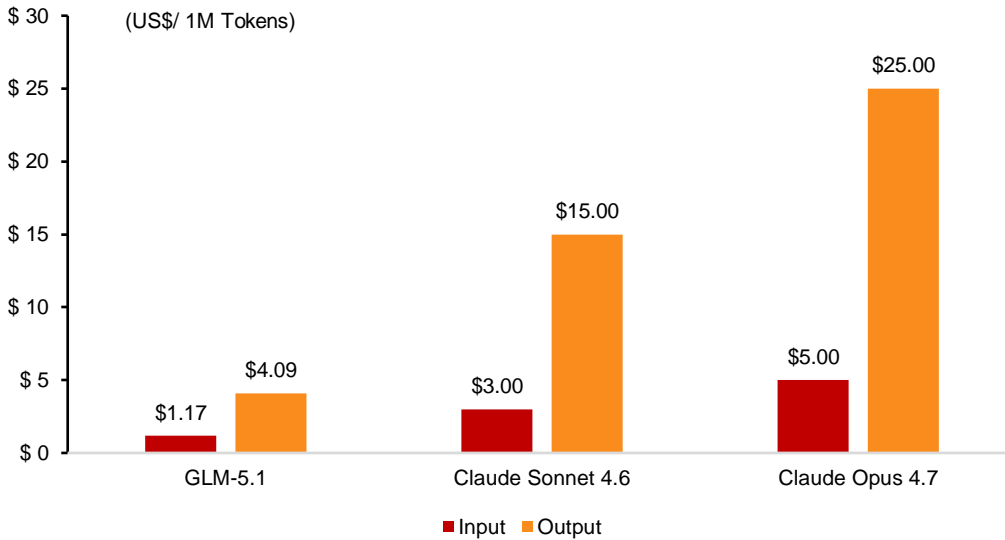
**Pricing: a sizable per-token discount across both models.** Z.ai prices (GLM-5.1's 32K+ context tier) materially below Anthropic across input and output tokens:

- **vs. Claude Opus 4.7:** Opus is 328% more expensive on input (US\$5.00 vs. US\$1.17 per 1M tokens) and 512% more expensive on output (US\$25.00 vs. US\$4.09 per 1M tokens).

- **vs. Claude Sonnet 4.6:** Sonnet is 157% more expensive on input (US\$3.00 vs. US\$1.17 per 1M tokens) and 267% more expensive on output (US\$15.00 vs. US\$4.09 per 1M tokens).

**Two pricing dynamics worth flagging.** First, Z.ai raised GLM Coding Plan pricing by 30% in February 2026, citing usage growth and compute reinvestment — a tangible signal of pricing power within the installed base (consistent with the 83% hike absorbed earlier, discussed above). Second, Anthropic cut Claude Opus 4.5 pricing by ~66% in November 2025, narrowing the absolute pricing gap but, in our view, not eliminating Z.ai's relative cost advantage.

**Figure 24: Z.ai vs. Anthropic: token price**



Source: Company data, CMBIGM

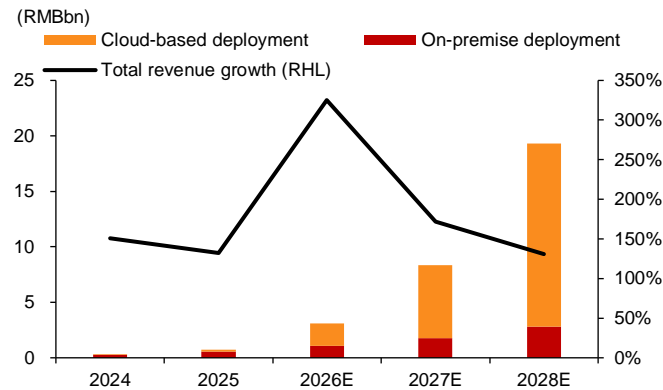
Note: Z.ai's pricing is based on GLM-5.1's 32K+ context tier.

## Business segment analysis and financial forecast

We forecast Z.ai's total revenue to achieve a 199% CAGR over 2025-2028E, reaching RMB19.3bn in 2028E, driven by the rapid increase in token usage and potential price increases aided by improvements in model intelligence capabilities.

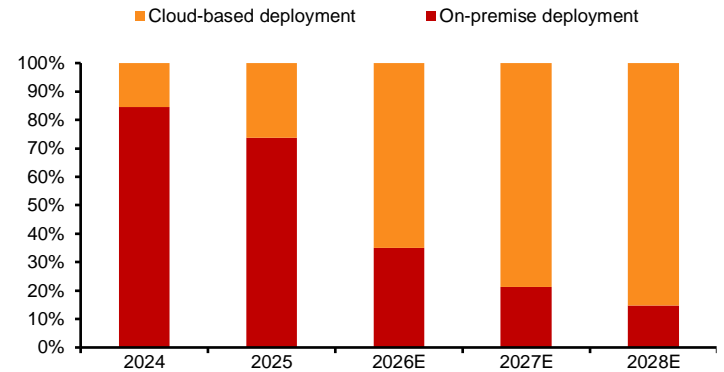
Breaking down the revenue forecast by deployment method, we project a 2025-2028E CAGR of 74% for on-premise deployment revenue and 342% for cloud-based deployment revenue. By business form and core product lines, we anticipate revenue CAGRs of 342%, 117%, 46%, and 20% for Open Platform and API, enterprise-level agents, enterprise-level general-purpose large models, and technical services and others, respectively.

**Figure 25: Z.ai: revenue growth and forecast**



Source: Company data, CMBIGM estimates

**Figure 26: Z.ai: revenue breakdown**



Source: Company data, CMBIGM estimates

Driven by Z.ai's strategic shift toward prioritizing the development of its Open Platform and API business, alongside strong scalable demand, we expect the Open Platform and API segment to contribute 85.3% of revenue in 2028E, up from 26.3% in 2025.

**Figure 27: Z.ai: key financial forecast**

(RMBmn)	2024	2025	2026E	2027E	2028E
<b>1. Revenue breakdown - by deployment method</b>					
<b>1.1 On-premise deployment</b>	<b>264</b>	<b>534</b>	<b>1,080</b>	<b>1,775</b>	<b>2,834</b>
Growth (yoy%)	134.4%	102.3%	102.4%	64.3%	59.7%
As % of total revenue (%)	84.5%	73.7%	35.1%	21.2%	14.7%
<b>1.2 Cloud-based deployment</b>	<b>48</b>	<b>190</b>	<b>1,999</b>	<b>6,597</b>	<b>16,492</b>
Growth (yoy%)	306.6%	292.6%	950.0%	230.0%	150.0%
As % of total revenue (%)	15.5%	26.3%	64.9%	78.8%	85.3%
<b>2. Revenue breakdown - by business form and core product lines</b>					
<b>2.1 Open platform and API</b>	<b>48</b>	<b>190</b>	<b>1,999</b>	<b>6,597</b>	<b>16,492</b>
Growth (yoy%)		292.6%	950.0%	230.0%	150.0%
As % of total revenue (%)	15.5%	26.3%	64.9%	78.8%	85.3%
<b>2.2 Enterprise-level agents</b>	<b>47</b>	<b>166</b>	<b>505</b>	<b>942</b>	<b>1,700</b>
Growth (yoy%)		248.8%	205.0%	86.3%	80.6%
As % of total revenue (%)	15.2%	22.9%	16.4%	11.2%	8.8%
<b>2.3 Enterprise-level general-purpose large models</b>	<b>215</b>	<b>366</b>	<b>572</b>	<b>830</b>	<b>1,130</b>
Growth (yoy%)		70.5%	56.4%	45.1%	36.1%
As % of total revenue (%)	68.7%	50.4%	18.6%	9.9%	5.8%
<b>2.4 Technical service and others</b>	<b>2</b>	<b>3</b>	<b>3</b>	<b>4</b>	<b>4</b>
Growth (yoy%)		31.6%	25.1%	20.1%	16.1%
As % of total revenue (%)	0.6%	0.4%	0.1%	0.0%	0.0%
<b>Total revenue</b>	<b>312</b>	<b>724</b>	<b>3,079</b>	<b>8,372</b>	<b>19,326</b>
Growth (yoy%)		150.9%	131.9%	325.1%	171.9%
<b>Gross profit</b>	<b>176</b>	<b>297</b>	<b>1,076</b>	<b>2,997</b>	<b>7,560</b>
- On-premise deployment	174	261	536	886	1,458
- Cloud-based deployment	2	36	540	2,111	6,102
<b>Operating profit (loss)</b>	<b>(2,538)</b>	<b>(3,787)</b>	<b>(4,848)</b>	<b>(4,935)</b>	<b>(3,022)</b>
Selling and marketing expenses	387	391	528	686	892
As % of total revenue	124.0%	54.0%	17.1%	8.2%	4.6%

Research and development expenses	2,195	3,180	4,802	6,483	8,752
As % of total revenue	702.7%	439.1%	156.0%	77.4%	45.3%
General and administration expenses	134	505	632	790	957
As % of total revenue	42.8%	69.8%	20.5%	9.4%	5.0%
<b>Net profit (loss) attributable to equity shareholders</b>	<b>(2,956)</b>	<b>(4,698)</b>	<b>(4,802)</b>	<b>(4,883)</b>	<b>(2,969)</b>
<b>Adjusted net profit (loss)</b>	<b>(2,466)</b>	<b>(3,182)</b>	<b>(4,097)</b>	<b>(4,032)</b>	<b>(2,074)</b>
<b>GPM (%)</b>	<b>56.3%</b>	<b>41.0%</b>	<b>34.9%</b>	<b>35.8%</b>	<b>39.1%</b>
- On-premise deployment	66.0%	48.8%	49.6%	49.9%	51.4%
- Cloud-based deployment	3.4%	18.9%	27.0%	32.0%	37.0%
<b>OPM (%)</b>	<b>-812.5%</b>	<b>-522.8%</b>	<b>-157.4%</b>	<b>-58.9%</b>	<b>-15.6%</b>
<b>NPM (%)</b>	<b>-946.3%</b>	<b>-648.6%</b>	<b>-155.9%</b>	<b>-58.3%</b>	<b>-15.4%</b>
<b>Adjusted NPM (%)</b>	<b>-789.2%</b>	<b>-439.3%</b>	<b>-133.0%</b>	<b>-48.2%</b>	<b>-10.7%</b>

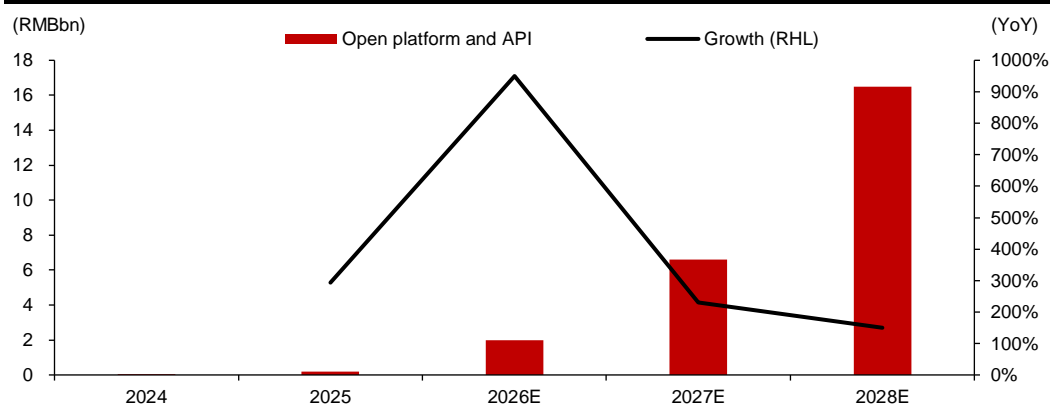
Source: Company data, CMBIGM estimates

## Open platform and API (64.9% of 2026E revenue)

The Open Platform and API segment represents the standardized, platform-based cloud services Z.ai provides to developers and enterprise customers, built on its general-purpose large model capabilities. These services transform the complex algorithmic capabilities of underlying large models into accessible cloud offerings through Application Programming Interfaces (APIs), Software Development Kits (SDKs), and Model-as-a-Service (MaaS) platform services. Open Platform and API revenue grew 292.6% YoY to RMB190.4mn in 2025, driven by the elevation of Z.ai's model intelligence ceiling and the increase in model invocations.

Driven by continued growth in model invocations and ongoing improvements in model capabilities, we forecast Z.ai's Open Platform and API revenue to grow to RMB2.0bn in 2026E, up 950% YoY, and to reach RMB16.5bn in 2028E - implying a 342% CAGR over 2025-2028E.

**Figure 28: Z.ai: open platform and API revenue and forecast**



Source: Company data, CMBIGM estimates

## On-premise deployment (35.1% of 2026E revenue)

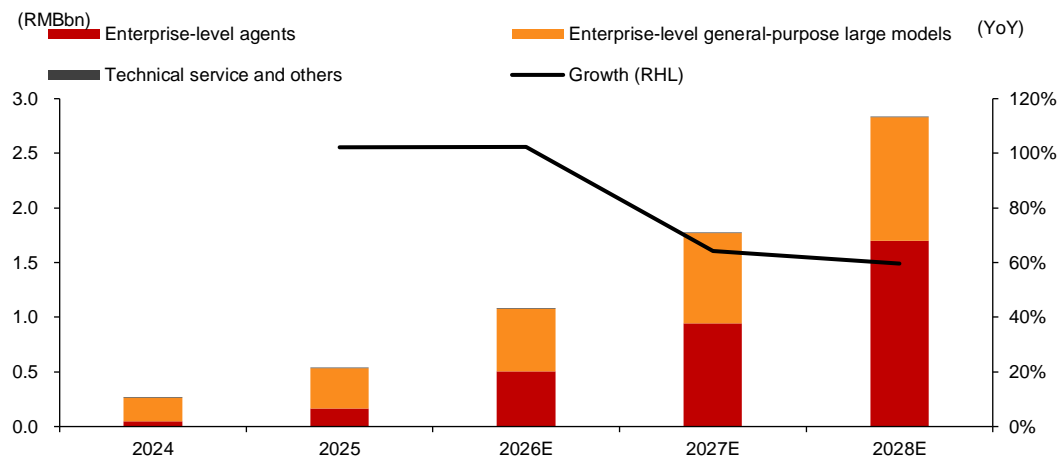
As laid out in Fig. 1, on-premise deployment revenue comprises three sub-segments: enterprise-level general-purpose large models, enterprise-level agents, and technical services and others. We forecast overall on-premise deployment revenue to achieve a 74% CAGR over 2025-2028E, reaching RMB2.8bn in 2028E. For 2026E, we project YoY revenue growth of 205% for enterprise-level agents, 56% for enterprise-level general-purpose large models, and 25% for technical services and others, translating to total on-premise deployment revenue growth of 102%.

- **Enterprise-level agents** refer to autonomous intelligent systems built for complex enterprise scenarios, with general-purpose large models serving as the core control unit, combined with enterprise knowledge bases and tool invocation capabilities. These products integrate memory storage, task planning, and execution feedback mechanisms, enabling them to emulate the thought patterns of human experts.

According to Z.ai, these agents autonomously understand instructions, decompose tasks, and invoke external tools (APIs) to complete business loops, addressing the need for automated execution and auxiliary decision-making across enterprise business processes — thereby improving operational efficiency and the level of digital management. Revenue from enterprise-level agents reached RMB165.7mn in 2025, up 248.8% YoY, driven by the elevation of Z.ai's model intelligence ceiling and surging market demand for enterprise-level agent products.

- **Enterprise-level general-purpose large models** refer to a matrix of pre-trained models independently developed by Z.ai, featuring large-scale parameters and strong generalization capabilities, and serving as the foundation for the company's core technological innovations and business development. These products are primarily delivered through private on-premise deployment, where the models and related software environments are installed directly on customers' local computing infrastructure or private cloud environments. Built on self-developed pre-trained foundation models with large-scale parameter sizes and deep learning capabilities, Z.ai has constructed a core model system encompassing language models, code models, multimodal models, and contemplation and reflection models. Through pre-training and instruction fine-tuning with massive volumes of multi-source data, these models address AI application needs across a range of vertical scenarios. Revenue from enterprise-level general-purpose large models was RMB365.7mn in 2025, up 70.5% YoY, primarily attributable to the elevation of Z.ai's model intelligence ceiling, coupled with sustained strong market demand.
- **Technical services and others** refer to specialized technical services provided by Z.ai to developers and enterprise clients in the AI field, based on the underlying technologies of general-purpose large models. This business focuses on the commercialization and application of the latest technological advances in the large model domain. Through in-depth technical enablement, Z.ai offers professional technical consultancy on large models, as well as resource matching, technical exchange, and collaborative innovation support. Revenue from technical services and others reached RMB2.5mn in 2025, up 31.6% YoY, primarily driven by increased customer demand.

**Figure 29: Z.ai: on-premise deployment revenue and forecast**



Source: Company data, CMBIGM estimates

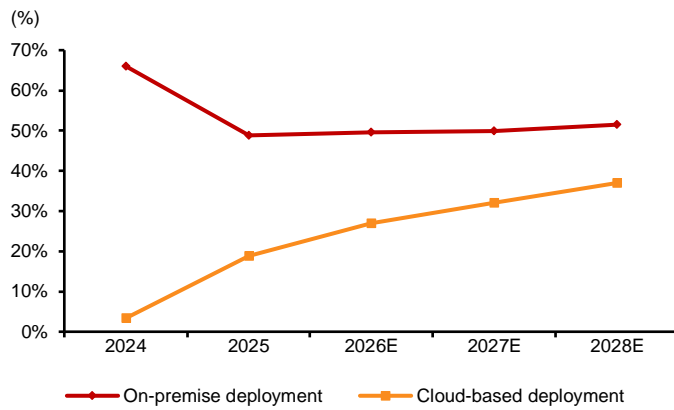
## Margin analysis

Z.ai's cost of revenue mainly consists of: 1) payroll costs, primarily representing the wages and benefits of service personnel involved in the deployment and maintenance of Z.ai's services; 2) computing service fees paid to providers of computing power required for service delivery; 3) provisions for warranties; 4) depreciation and amortization; 5) technical and consultation fees, as Z.ai outsources certain standardized, labor-intensive tasks —

such as data annotation — to third-party service providers to manage costs; and 6) taxes and surcharges.

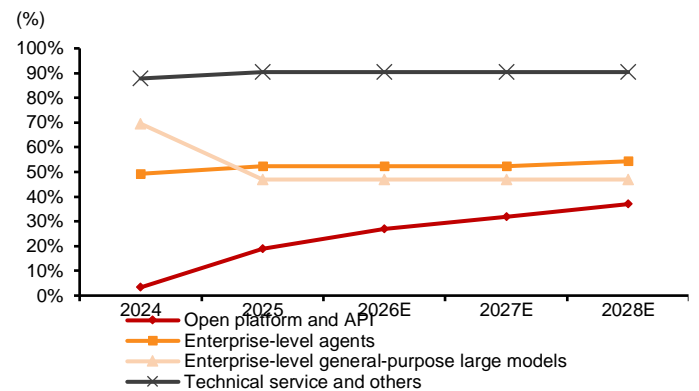
Z.ai's overall gross profit margin (GPM) declined from 56.3% in 2024 to 41.0% in 2025, owing to a higher revenue contribution from the cloud deployment business and a decline in the gross margin of the on-premise deployment business. Looking ahead, we expect overall GPM to decline further to 34.9% in 2026E, driven by the rising share of cloud-based deployment, which carries a lower gross margin than the group average. We do, however, anticipate an improvement in cloud-based deployment GPM, supported by improving model inference efficiency, economies of scale from expanded computing capacity that lower marginal costs, and price increases. As a result, driven by continued GPM improvement in cloud-based deployment, we estimate group-level GPM will rise to 35.8% in 2027E and 39.1% in 2028E.

**Figure 30: Z.ai: gross margin by deployment method**



Source: Company data, CMBIGM estimates

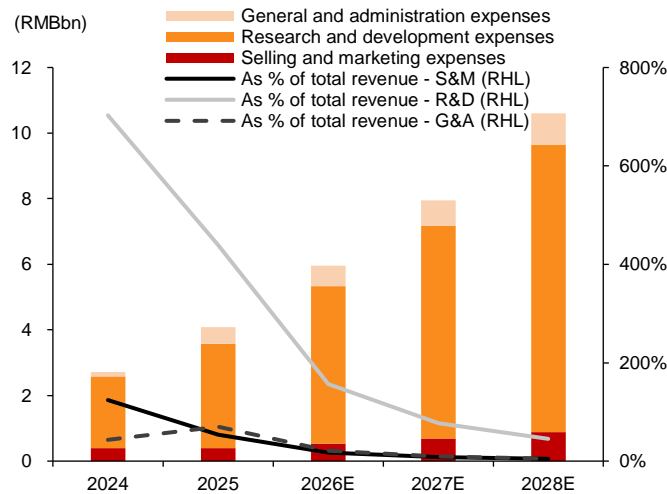
**Figure 31: Z.ai: gross margin by business form and core product lines**



Source: Company data, CMBIGM estimates

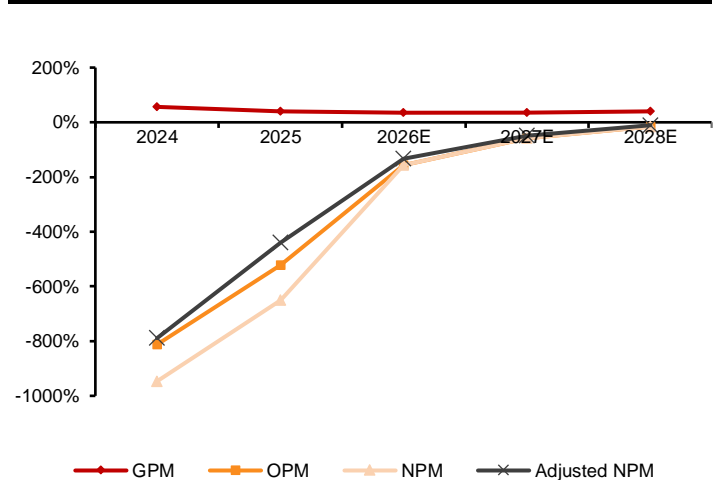
As a percentage of total revenue, S&M/R&D/G&A expenses reached 54/439/70% of total revenue in 2025. With robust growth in invocation volume, which will drive solid revenue growth and unleash of operating leverage, we anticipate overall operating expense ratio (defined as the sum up of S&M, R&D, G&A expenses as % of total revenue) to optimize to 55% in 2028E from 563% in 2025, which should drive a narrowing operating loss margin to 16% in 2028E from 523% in 2025.

**Figure 32: Z.ai: Opex and operating expense ratio**



Source: Company data, CMBIGM estimates

**Figure 33: Z.ai: margin trend**



Source: Company data, CMBIGM estimates

## Valuation: HK\$1,503.9 per share based on 40.0x 2028E PS

We value Z.ai at RMB584.5bn (HK\$1,503.9 per share), based on 40.0x 2028E P/S discounted back to 2026E at a 15% WACC. Our 40.0x 2028E target multiple represents a premium to the comparable peer average of 12.0x P/S, which in our view is justified by Z.ai's strong 199% revenue CAGR over 2025-2028E (versus the industry average of 77%). Also, our positive view is anchored on Z.ai's: 1) frontier model capability, underpinned by deep research that sustains rapid iteration; 2) strategic foresight in pinpointing monetization scenarios, positioning it to capture opportunities yet to unfold; and 3) a large institutional base (8k+ customers as of Jun 2025), served with tailor-made solutions.

**Figure 34: Z.ai: HK\$1,503.9 per share based on 40.0x 2028E PS**

(RMBmn)	2026E
2028E Revenue	19,326
Target 2028E P/S	40.0x
Equity value (2028E)	773,041
Equity value (2026E; discount back at 15% WACC)	584,530
No. of shares outstanding (diluted; mn)	446
Target price (RMB)	1,311
<b>Target price (HKD)</b>	<b>1,503.9</b>

Source: Company data, CMBIGM estimates

Note: our HKD: RMB exchange rate assumption is 1.15:1

Compared to the valuations of leading private industry peers: 1) Anthropic completed its US\$65bn Series H financing in May 2026 at a post-money valuation of US\$965bn. Based on Anthropic's latest disclosed ARR of US\$47bn in earlier May 2026, the implied post money valuation multiple was 20.5x ARR; 2) OpenAI announced the completion of a US\$122bn funding round on 31 March 2026 — one of the largest private fundraises in history, bringing the post-money valuation to US\$852bn. Based on Sacra's estimated ARR of approximately US\$25bn for OpenAI, the latest round implies a post-money valuation multiple of roughly 34x.

**Figure 35: Z.ai: comps table**

Company	Ticker	Price (Local)	Market cap (US\$mn)	PS(x)				EV/EBITDA (x)			Revenue CAGR 25-28E
				2025	2026E	2027E	2028E	2026E	2027E	2028E	
MiniMax	100 HK	464.4	18,443	258.3	62.0	22.1	9.9	NA	NA	NA	197%
Xunce	3317 HK	157.4	6,531	36.8	19.4	11.7	7.7	280.0	75.3	39.1	69%
Haizhi	2706 HK	67.5	3,449	39.9	26.9	20.1	15.1	1204.5	333.8	199.7	38%
Mininglamp	2718 HK	179.7	3,341	15.8	12.5	9.7	7.5	156.0	92.3	NA	28%
Palantir	PLTR US	136.5	327,161	74.4	42.3	29.1	20.0	74.6	51.0	35.8	55%
<b>Total average</b>				<b>173.8</b>	<b>32.6</b>	<b>18.6</b>	<b>12.0</b>	<b>428.8</b>	<b>138.1</b>	<b>91.5</b>	<b>77%</b>

Source: Bloomberg, CMBIGM

Note: data as of 9 Jun 2026

## Appendix

Figure 36: Z.ai: management profile

Name	Position	Roles and responsibilities	Prior experience
Dr. Liu Debing (刘德兵)	Co-founder, executive director and chairman of the board	Responsible for the strategic planning, business direction and overall management of Z.ai	Dr. Liu has nearly 18 years of experience in the computing technology industry. From September 2007 to December 2012, he worked at the Beijing Institute of Technicolor (China) Technology Co., Ltd. with his last position as a research engineer. Dr. Liu subsequently worked as senior engineer at Tsinghua University. Dr. Liu obtained a bachelor's degree in computer science and technology from Beijing Jiaotong University in the PRC in July 1999 and a Ph.D. degree in computer science and technology from Institute of Computing Technology, Chinese Academy of Sciences in the PRC in July 2007. In October 2013, Dr. Liu received the Science and Technology Progress Award—First Prize from the Chinese Association for Artificial Intelligence. In November 2017, he was awarded the title of Senior Engineer by the Chinese Academy of Sciences and the Beijing Science and Technology Progress Award—First Prize by the Beijing Municipal People's Government. In September 2021, Dr. Liu received the National Science and Technology Progress Award—Second Prize by the State Council. In March 2024, Dr. Liu received Science and Technology Progress—First Prize from the China Institute of Electronics.
Dr. Zhang Peng (张鹏)	Co-founder, executive director, chief executive officer and general manager	Responsible for the business development, R&D and the daily operations and management of Z.ai	Dr. Zhang has nearly 20 years of experience in the field of computer science. He worked at Tsinghua University from August 2005 to December 2020. Dr. Zhang obtained a bachelor's degree in computer science and technology, a master's degree in computer science and technology and a Ph.D. degree in electronics and information in July 2002, July 2005 and June 2025, respectively, from Tsinghua University in the PRC. In November 2009, Dr. Zhang received the Fourth Wang Xuan News Science and Technology Award from the China Association of Press and Technology Professionals. In March 2024, Dr. Zhang was awarded the Science and Technology Progress Award—First Prize by the China Institute of Electronics. In April 2025, Dr. Zhang was awarded the title of Model Worker of the National Industry and Information System by the Ministry of Human Resources and Social Security and the MIIT.
Ms. Zhang Xiaohan (张笑涵)	Executive director	Participating in the Board as an employee director	Ms. Zhang joined Z.ai in October 2021 and has served as core manager of the Company's data labeling operations and Z.ai QingYan since July 2022. Ms. Zhang obtained a bachelor's degree in engineering mechanics in July 2019 and a master's degree in data science and information technology in July 2022 from Tsinghua University in the PRC.
Dr. Li Juanzi (李涓子)	Co-founder and non-executive director	Responsible for providing guidance for the R&D, strategy and business development of Z.ai	From July 1989 to July 1996, Dr. Li was a lecturer in the Department of Computer Science and Technology at Shanxi University. Dr. Li served as lecturer from December 2001 to December 2002, associate professor from December 2002 to December 2008 and tenured professor since December 2008 in the Department of Computer Science and Technology at Tsinghua University. Since December 2014, Dr. Li has served as director of the Special Committee on Language and Knowledge Computing of the Chinese Information Processing Society of China. Since January 2019, Dr. Li has served as director of the Knowledge Intelligence Center at the Institute for Artificial Intelligence of Tsinghua University. Dr. Li also serves as director of the Joint Research Center between Tsinghua University's Department of Computer Science and Technology and Siemens (China) Co., Ltd for Industrial Intelligence and Internet of Things. Dr. Li obtained a bachelor's degree and a master's degree in computer science and technology from Shanxi University in the PRC in July 1986 and July 1989, respectively. She received her Ph.D. degree in computer science and technology from Tsinghua University in the PRC in January 2000. Dr. Li also conducted postdoctoral research in the same field from December 2000 to December 2001. In October 2013, Dr. Li received the Scientific and Technological Progress—First Prize from the Chinese Association for Artificial Intelligence. In November 2017, Dr. Li received the Beijing Scientific and Technological Progress—First Prize from the Beijing People's Municipal Government. In March 2020, Dr. Li was recognized by the China Language Resources Protection Outstanding Individual Award. In April 2021, Dr. Li received the Beijing Invention Patent Award—First Prize from the Beijing People's Municipal Government. In September 2021, Dr. Li received the National Scientific and Technological Progress—Second Prize from the State Council. In March 2024, Dr. Li received Scientific and Technological Progress—First Prize from the China Institute of Electronics. In April 2025, Dr. Li was recognized as Outstanding Mentor from Tsinghua University.

Name	Position	Roles and responsibilities	Prior experience
Mr. Wang Shaolan (王绍兰)	Co-founder and deputy general manager	Responsible for the business development, financing and the daily operations and management of Z.ai	Mr. Wang has more than 20 years of experience in the technology industry. From August 2001 to May 2012, he worked at Beijing O2Micro Microelectronics Development Co., Ltd. From July 2012 to June 2013, he worked at Nebula Sunac (Beijing) Technology Co., Ltd. From July 2013 to December 2017, he served as chief executive officer at Wankang Century Technology (Beijing) Co., Ltd. From March 2018 to June 2019, he worked at Tsinghua University. Mr. Wang obtained a master's degree in electronics and electrical drive from Anhui University of Technology in the PRC in June 1999.
Mr. Xiao Lei (肖磊)	Board secretary	Responsible for the corporate governance and financing operations of Z.ai	Mr. Xiao served as a legal consultant of COFCO Land Management Co., Ltd. from August 2012 to August 2014. From May 2014 to January 2020, Mr. Xiao held several positions at Shoutai Jinxin (Beijing) Equity Investment Fund Management Co., Ltd., including deputy general manager and general manager of the legal department, secretary to the board, general manager of the investment banking department, and deputy general manager. Mr. Xiao served as director of finance and investment, joint company secretary, board secretary, chief financial officer, and vice president of AlInnovation Technology Group Co., Ltd., a company listed on the Stock Exchange with stock code 2121. He served as vice president of Shanghai Soybean Network Technology Co., Ltd. from February 2024 to October 2025. Mr. Xiao obtained his bachelor's degree in law from China University of Political Science and Law in the PRC in July 2009 and his master's degree in civil and commercial law from China University of Political Science and Law in June 2012.

Source: Company data, CMBIGM

## Financial Summary

INCOME STATEMENT	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec (RMB mn)						
Revenue	125	312	724	3,079	8,372	19,326
Cost of goods sold	44	137	428	2,004	5,375	11,766
Others	0	0	0	0	0	0
<b>Gross profit</b>	<b>80</b>	<b>176</b>	<b>297</b>	<b>1,076</b>	<b>2,997</b>	<b>7,560</b>
<b>Operating expenses</b>	<b>(696)</b>	<b>(2,717)</b>	<b>(4,077)</b>	<b>(5,962)</b>	<b>(7,959)</b>	<b>(10,602)</b>
Selling expense	(101)	(387)	(391)	(528)	(686)	(892)
SG&A expense	(66)	(134)	(505)	(632)	(790)	(957)
R&D expense	(529)	(2,195)	(3,180)	(4,802)	(6,483)	(8,752)
Others	10	(2)	7	(39)	(28)	(20)
<b>Operating profit</b>	<b>(626)</b>	<b>(2,538)</b>	<b>(3,787)</b>	<b>(4,848)</b>	<b>(4,935)</b>	<b>(3,022)</b>
Interest income	11	13	15	68	72	86
Interest expense	(26)	(38)	(74)	(30)	(24)	(21)
Others	(147)	(395)	(872)	(13)	(17)	(31)
<b>Pre-tax profit</b>	<b>(788)</b>	<b>(2,958)</b>	<b>(4,718)</b>	<b>(4,822)</b>	<b>(4,903)</b>	<b>(2,989)</b>
Income tax	0	0	0	0	0	0
<b>After tax profit</b>	<b>(788)</b>	<b>(2,958)</b>	<b>(4,718)</b>	<b>(4,822)</b>	<b>(4,903)</b>	<b>(2,989)</b>
Minority interest	0	(2)	(20)	(20)	(20)	(20)
<b>Net profit</b>	<b>(788)</b>	<b>(2,956)</b>	<b>(4,698)</b>	<b>(4,802)</b>	<b>(4,883)</b>	<b>(2,969)</b>
<b>Adjusted net profit</b>	<b>(621)</b>	<b>(2,466)</b>	<b>(3,182)</b>	<b>(4,097)</b>	<b>(4,032)</b>	<b>(2,074)</b>
<b>BALANCE SHEET</b>						
YE 31 Dec (RMB mn)						
<b>Current assets</b>	<b>1,863</b>	<b>3,016</b>	<b>3,571</b>	<b>3,762</b>	<b>4,890</b>	<b>8,851</b>
Cash & equivalents	1,249	2,268	2,255	2,394	2,869	5,786
Restricted cash	0	1	4	4	4	4
Account receivables	416	667	699	1,131	1,673	2,538
Inventories	29	32	126	229	339	514
ST bank deposits	0	0	109	0	0	0
Financial assets at FVTPL	159	43	377	0	0	0
Contract assets	10	5	2	3	5	8
<b>Non-current assets</b>	<b>997</b>	<b>1,360</b>	<b>1,282</b>	<b>1,320</b>	<b>1,440</b>	<b>1,688</b>
PP&E	788	866	656	514	422	365
Investment in JVs & assos	13	201	338	393	448	503
Intangibles	55	50	51	53	56	61
Goodwill	39	39	39	39	39	39
Other non-current assets	102	203	198	321	474	720
<b>Total assets</b>	<b>2,860</b>	<b>4,376</b>	<b>4,854</b>	<b>5,083</b>	<b>6,330</b>	<b>10,539</b>
<b>Current liabilities</b>	<b>3,609</b>	<b>7,838</b>	<b>12,406</b>	<b>2,637</b>	<b>3,838</b>	<b>5,935</b>
Short-term borrowings	0	269	605	0	0	0
Account payables	288	603	1,329	2,115	3,186	5,077
Other current liabilities	3,180	6,677	10,073	0	0	0
Lease liabilities	67	213	251	251	251	251
Contract liabilities	74	75	149	271	400	607
<b>Non-current liabilities</b>	<b>234</b>	<b>493</b>	<b>559</b>	<b>581</b>	<b>660</b>	<b>845</b>
Long-term borrowings	0	0	85	85	85	85
Deferred income	30	35	180	292	431	654
Other non-current liabilities	204	458	294	205	144	106
<b>Total liabilities</b>	<b>3,843</b>	<b>8,331</b>	<b>12,965</b>	<b>3,217</b>	<b>4,497</b>	<b>6,780</b>
Share capital	28	36	40	45	45	45
Capital surplus	0	0	0	14,069	18,069	22,069
Retained earnings	(998)	(3,954)	(8,133)	(12,935)	(17,819)	(20,787)
Other reserves	(13)	(39)	0	726	1,597	2,511
<b>Total shareholders equity</b>	<b>(983)</b>	<b>(3,957)</b>	<b>(8,093)</b>	<b>1,904</b>	<b>1,891</b>	<b>3,837</b>
Minority interest	0	1	(18)	(38)	(58)	(78)
<b>Total equity and liabilities</b>	<b>2,860</b>	<b>4,376</b>	<b>4,854</b>	<b>5,083</b>	<b>6,330</b>	<b>10,539</b>

<b>CASH FLOW</b>	<b>2023A</b>	<b>2024A</b>	<b>2025A</b>	<b>2026E</b>	<b>2027E</b>	<b>2028E</b>
<b>YE 31 Dec (RMB mn)</b>						
<b>Operating</b>						
<b>Profit before taxation</b>	<b>(788)</b>	<b>(2,958)</b>	<b>(4,718)</b>	<b>(4,822)</b>	<b>(4,903)</b>	<b>(2,989)</b>
Depreciation & amortization	69	280	281	226	188	166
Tax paid	0	0	0	0	0	0
Change in working capital	(96)	(0)	704	331	489	966
Others	167	433	1,487	729	884	947
<b>Net cash from operations</b>	<b>(648)</b>	<b>(2,245)</b>	<b>(2,246)</b>	<b>(3,537)</b>	<b>(3,343)</b>	<b>(909)</b>
<b>Investing</b>						
Capital expenditure	(509)	(133)	(23)	(27)	(31)	(35)
Acquisition of subsidiaries/ investments	(110)	(131)	(40)	0	0	0
Others	(166)	215	(312)	486	0	0
<b>Net cash from investing</b>	<b>(785)</b>	<b>(49)</b>	<b>(376)</b>	<b>459</b>	<b>(31)</b>	<b>(35)</b>
<b>Financing</b>						
Net borrowings	0	266	1,385	(607)	0	0
Proceeds from share issues	98	0	0	4,000	4,000	4,000
Others	2,365	3,047	1,229	(176)	(153)	(138)
<b>Net cash from financing</b>	<b>2,463</b>	<b>3,312</b>	<b>2,614</b>	<b>3,217</b>	<b>3,847</b>	<b>3,862</b>
<b>Net change in cash</b>						
Cash at the beginning of the year	219	1,249	2,268	2,255	2,394	2,869
Others	0	0	(5)	0	0	0
<b>Cash at the end of the year</b>	<b>1,249</b>	<b>2,268</b>	<b>2,255</b>	<b>2,394</b>	<b>2,869</b>	<b>5,786</b>
<b>GROWTH</b>	<b>2023A</b>	<b>2024A</b>	<b>2025A</b>	<b>2026E</b>	<b>2027E</b>	<b>2028E</b>
<b>YE 31 Dec</b>						
Revenue	116.9%	150.9%	131.9%	325.1%	171.9%	130.8%
Gross profit	156.6%	118.5%	68.7%	262.6%	178.6%	152.3%
<b>PROFITABILITY</b>	<b>2023A</b>	<b>2024A</b>	<b>2025A</b>	<b>2026E</b>	<b>2027E</b>	<b>2028E</b>
<b>YE 31 Dec</b>						
Gross profit margin	64.6%	56.3%	41.0%	34.9%	35.8%	39.1%
Operating margin	(502.4%)	(812.5%)	(522.8%)	(157.4%)	(58.9%)	(15.6%)
Adj. net profit margin	(498.6%)	(789.2%)	(439.3%)	(133.0%)	(48.2%)	(10.7%)
Return on equity (ROE)	na	na	na	na	(257.4%)	(103.6%)
<b>GEARING/LIQUIDITY/ACTIVITIES</b>	<b>2023A</b>	<b>2024A</b>	<b>2025A</b>	<b>2026E</b>	<b>2027E</b>	<b>2028E</b>
<b>YE 31 Dec</b>						
Current ratio (x)	0.5	0.4	0.3	1.4	1.3	1.5
Inventory turnover days	162.1	81.9	67.5	32.3	19.3	13.2
<b>VALUATION</b>	<b>2023A</b>	<b>2024A</b>	<b>2025A</b>	<b>2026E</b>	<b>2027E</b>	<b>2028E</b>
<b>YE 31 Dec</b>						
P/E	ns	ns	ns	ns	ns	ns
P/E (diluted)	ns	ns	ns	ns	ns	ns
P/B	47.8	33.1	29.8	79.4	63.8	38.3
P/CFPS	ns	ns	ns	ns	ns	ns

Source: Company data, CMBIGM estimates. Note: The calculation of net cash includes financial assets.

# Disclosures & Disclaimers

## Analyst Certification

The research analyst who is primary responsible for the content of this research report, in whole or in part, certifies that with respect to the securities or issuer that the analyst covered in this report: (1) all of the views expressed accurately reflect his or her personal views about the subject securities or issuer; and (2) no part of his or her compensation was, is, or will be, directly or indirectly, related to the specific views expressed by that analyst in this report. Besides, the analyst confirms that neither the analyst nor his/her associates (as defined in the code of conduct issued by The Hong Kong Securities and Futures Commission) (1) have dealt in or traded in the stock(s) covered in this research report within 30 calendar days prior to the date of issue of this report; (2) will deal in or trade in the stock(s) covered in this research report 3 business days after the date of issue of this report; (3) serve as an officer of any of the Hong Kong listed companies covered in this report; and (4) have any financial interests in the Hong Kong listed companies covered in this report. CMBIGM or its affiliate(s) have investment banking relationship with the issuers covered in this report in preceding 12 months.

## CMBIGM Ratings

**BUY** : Stock with potential return of over 15% over next 12 months  
**HOLD** : Stock with potential return of +15% to -10% over next 12 months  
**SELL** : Stock with potential loss of over 10% over next 12 months  
**NOT RATED** : Stock is not rated by CMBIGM

**OUTPERFORM** : Industry expected to outperform the relevant broad market benchmark over next 12 months  
**MARKET-PERFORM** : Industry expected to perform in-line with the relevant broad market benchmark over next 12 months  
**UNDERPERFORM** : Industry expected to underperform the relevant broad market benchmark over next 12 months

## CMB International Global Markets Limited

Address: 45/F, Champion Tower, 3 Garden Road, Hong Kong, Tel: (852) 3900 0888 Fax: (852) 3900 0800

CMB International Global Markets Limited ("CMBIGM") is a wholly owned subsidiary of CMB International Capital Corporation Limited (a wholly owned subsidiary of China Merchants Bank)

## Important Disclosures

There are risks involved in transacting in any securities. The information contained in this report may not be suitable for the purposes of all investors. CMBIGM does not provide individually tailored investment advice. This report has been prepared without regard to the individual investment objectives, financial position or special requirements. Past performance has no indication of future performance, and actual events may differ materially from that which is contained in the report. The value of, and returns from, any investments are uncertain and are not guaranteed and may fluctuate as a result of their dependence on the performance of underlying assets or other variable market factors. CMBIGM recommends that investors should independently evaluate particular investments and strategies, and encourages investors to consult with a professional financial advisor in order to make their own investment decisions.

This report or any information contained herein, have been prepared by the CMBIGM, solely for the purpose of supplying information to the clients of CMBIGM or its affiliate(s) to whom it is distributed. This report is not and should not be construed as an offer or solicitation to buy or sell any security or any interest in securities or enter into any transaction. Neither CMBIGM nor any of its affiliates, shareholders, agents, consultants, directors, officers or employees shall be liable for any loss, damage or expense whatsoever, whether direct or consequential, incurred in relying on the information contained in this report. Anyone making use of the information contained in this report does so entirely at their own risk.

The information and contents contained in this report are based on the analyses and interpretations of information believed to be publicly available and reliable. CMBIGM has exerted every effort in its capacity to ensure, but not to guarantee, their accuracy, completeness, timeliness or correctness. CMBIGM provides the information, advices and forecasts on an "AS IS" basis. The information and contents are subject to change without notice. CMBIGM may issue other publications having information and/or conclusions different from this report. These publications reflect different assumption, point-of-view and analytical methods when compiling. CMBIGM may make investment decisions or take proprietary positions that are inconsistent with the recommendations or views in this report.

CMBIGM may have a position, make markets or act as principal or engage in transactions in securities of companies referred to in this report for itself and/or on behalf of its clients from time to time. Investors should assume that CMBIGM does or seeks to have investment banking or other business relationships with the companies in this report. As a result, recipients should be aware that CMBIGM may have a conflict of interest that could affect the objectivity of this report and CMBIGM will not assume any responsibility in respect thereof. This report is for the use of intended recipients only and this publication, may not be reproduced, reprinted, sold, redistributed or published in whole or in part for any purpose without prior written consent of CMBIGM. Additional information on recommended securities is available upon request.

For recipients of this document in the United Kingdom

This report has been provided only to persons (I) falling within Article 19(5) of the Financial Services and Markets Act 2000 (Financial Promotion) Order 2005 (as amended from time to time) ("The Order") or (II) are persons falling within Article 49(2) (a) to (d) ("High Net Worth Companies, Unincorporated Associations, etc.") of the Order, and may not be provided to any other person without the prior written consent of CMBIGM.

For recipients of this document in the United States

CMBIGM is not a registered broker-dealer in the United States. As a result, CMBIGM is not subject to U.S. rules regarding the preparation of research reports and the independence of research analysts. The research analyst who is primary responsible for the content of this research report is not registered or qualified as a research analyst with the Financial Industry Regulatory Authority ("FINRA"). The analyst is not subject to applicable restrictions under FINRA Rules intended to ensure that the analyst is not affected by potential conflicts of interest that could bear upon the reliability of the research report. This report is intended for distribution in the United States solely to "major US institutional investors", as defined in Rule 15a-6 under the US, Securities Exchange Act of 1934, as amended, and may not be furnished to any other person in the United States. Each major US institutional investor that receives a copy of this report by its acceptance hereof represents and agrees that it shall not distribute or provide this report to any other person. Any U.S. recipient of this report wishing to effect any transaction to buy or sell securities based on the information provided in this report should do so only through a U.S.-registered broker-dealer.

For recipients of this document in Singapore

This report is distributed in Singapore by CMBI (Singapore) Pte. Limited (CMBISG) (Company Regn. No. 201731928D), an Exempt Financial Adviser as defined in the Financial Advisers Act (Cap. 110) of Singapore and regulated by the Monetary Authority of Singapore. CMBISG may distribute reports produced by its respective foreign entities, affiliates or other foreign research houses pursuant to an arrangement under Regulation 32C of the Financial Advisers Regulations. Where the report is distributed in Singapore to a person who is not an Accredited Investor, Expert Investor or an Institutional Investor, as defined in the Securities and Futures Act (Cap. 289) of Singapore, CMBISG accepts legal responsibility for the contents of the report to such persons only to the extent required by law. Singapore recipients should contact CMBISG at +65 6350 4400 for matters arising from, or in connection with the report.