Iluvatar CoreX (9903 HK)

A system-level challenger in China's evolving AI compute stack space; Initiate at BUY

We initiate coverage on Shanghai Iluvatar CoreX Semiconductor at BUY, with TP at HK$694. We think the key change in China's AI compute market is no longer just faster localization, but a deeper shift in how compute is deployed and monetized: As AI infrastructure becomes more system-driven, software compatibility, workload specialization, and deployment readiness are beginning to matter as much as peak silicon performance. In that context, the Company stands out as a well-positioned domestic GPGPU (general-purpose computing on graphics processing unit) challenger, with 1) dual product lines across training and inference, 2) a mature hardware-software co-design platform, and 3) growing commercial adoption across multiple industry verticals. We therefore see the Company as a beneficiary of both domestic substitution and a more structural reordering of the AI compute stack.

- **China's GPGPU market is entering a sustained growth phase, with the basis of competition shifting from single-chip performance toward system-level deployment.** Per Frost & Sullivan (F&S), the market is projected to grow at a 29.5% CAGR over 2025-29E, with domestic vendors outpacing global peers as policy support, ecosystem maturity, and uncertainty around access to foreign AI chips continue to foster local substitution. At the same time, AI compute is increasingly evaluated on workload fit, software compatibility, and cluster efficiency rather than peak silicon performance alone. **In our view, this creates a more accessible path for domestic suppliers to gain share through incremental insertion into heterogeneous, workload-specific deployments, especially where deployment readiness matters as much as peak performance.**
- **Within this backdrop, the Company differentiates itself from domestic GPGPU peers along three dimensions.** First, it is one of the few Chinese GPGPU vendors offering both training (TG series) and inference (ZK series) workloads with dedicated silicon. Second, its CUDA-compatible software stack and in-house hardware-software co-design platform lower migration friction for customers transitioning from incumbent accelerators, a capability that takes years of toolchain investment to replicate. Third, the Company has already achieved commercial traction across cloud compute, finance, telecom, and public-sector customers, translating into a visible shipment ramp that validates its go-to-market execution. Combined, these factors position the Company less as a pure performance-gap challenger and more as a system-level alternative for domestic AI infra. buildout.
- **Initiate at BUY with TP at HK$694.** We adopt a peer-based valuation framework and apply a 26.7x 2027E P/S multiple (in line with its peers) to reflect the Company's high-growth trajectory and early-stage earnings profile. We believe this valuation is supported by its strong shipment ramp-up, expanding role in China's AI infrastructure buildout, and positioning within a structurally evolving, system-level compute architecture.
- **Risks:** Supply chain disruptions, weak demand, intensified challenges from domestic or foreign peers, etc.

BUY (Initiate)

Target Price	**HK$694.00**
Up/Downside	**65.6%**
Current Price	**HK$419.20**

China Semiconductors

Kevin ZHANG
(852) 3761 8727
kevinzhang@cmbi.com.hk

Aaron GUO
(852) 3916 3715
aaronguo@cmbi.com.hk

Stock Data

Mkt Cap (HK$ mn)	95,949.0
Avg 3 mths t/o (HK$ mn)	285.9
52w High/Low (HK$)	NA/NA
Total Issued Shares (mn)	228.9

Source: FactSet

Shareholding Structure

3W Fund Management	2.8%
Qiming Venture Partners	1.2%

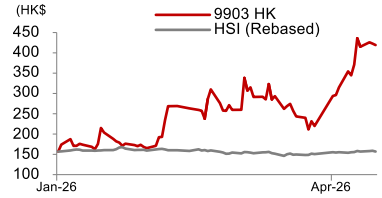
Source: HKEx

Share Performance

	Absolute	Relative
1-mth	42.9%	38.0%
3-mth	95.0%	98.5%
6-mth	NM	NM

Source: FactSet

12-mth Price Performance



Source: FactSet

Earnings Summary

(YE 31 Dec)	FY24A	FY25A	FY26E	FY27E	FY28E
Revenue (RMB mn)	540	1,034	2,482	5,233	8,067
YoY growth (%)	86.7	91.6	140.2	110.8	54.2
Gross margin (%)	49.1	54.0	51.6	50.6	51.5
Net profit (RMB mn)	(892.4)	(1,003.7)	(671.0)	159.6	1,040.9
YoY growth (%)	na	na	na	na	552.2
EPS (Reported) (RMB cents)	(464.00)	(531.00)	(293.17)	69.73	454.77
P/S (x)	154.8	80.8	33.6	16.0	10.4
P/E (x)	ns	ns	ns	523.3	80.2

Source: Company data, Bloomberg, CMBIGM estimates

Contents

Investment Thesis

System-level positioning creates a more realistic path to share gains

The Company's investment case is not solely about catching up on peak chip specifications, but about positioning itself within a changing AI compute architecture. With product lines spanning both training and inference, together with AI computing solutions and a relatively mature hardware-software co-design platform, the Company is better-placed than many same-tiered domestic peers to participate in system-level deployments. As AI infrastructure evolves toward prefill-decode (PD) disaggregation and more heterogeneous compute architectures, we think this broader platform positioning creates a more realistic path to incremental share gains, particularly in deployment scenarios where workload fit, software compatibility, and implementation efficiency matter as much as absolute silicon performance.

Domestic substitution remains a structural growth tailwind

China's push toward semiconductor self-sufficiency continues to drive structural demand for domestic GPGPU vendors. Amid ongoing US export controls on advanced AI chips and persistent uncertainty around access to foreign suppliers, local customers are increasingly prioritizing domestically developed solutions. In this context, the Company stands to benefit from both policy tailwinds and improving ecosystem readiness, with its expanding product portfolio and growing deployment track record supporting broader adoption across training and inference workloads.

Broad commercialization progress strengthens execution credibility

The Company has already moved beyond a purely roadmap-driven story, with growing evidence of commercial adoption across a broad set of end markets. Its customer base has expanded across sectors such as finance, healthcare, retail, and research in addition to cloud compute vendors, while its products and solutions have been deployed in a rising number of real-world projects. In our view, this breadth of commercialization is important because it validates not only the versatility of the platform, but also the Company's ability to convert technology capability into repeatable customer adoption, which should support both near-term growth and longer-term stickiness.

Flexible supply-chain access supports scaling resilience

The Company appears relatively better positioned than some domestic peers in terms of supply-chain flexibility, supported by its ability to leverage both offshore and onshore manufacturing capacity. While advanced-node constraints remain a key industry bottleneck, particularly within China, the Company's product positioning below certain export-control thresholds has historically allowed continued access to external foundry services. At the same time, ongoing domestic capacity buildout provides an additional medium-term buffer, helping mitigate supply risk and supporting more stable production scaling.

Initiate coverage at BUY with TP at HK$694

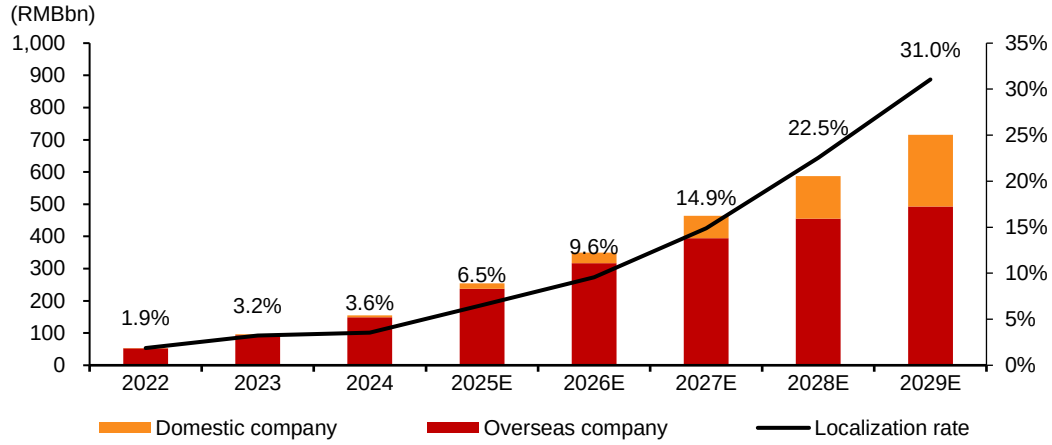
We adopt a peer-based valuation framework and apply a 26.7x 2027E P/S multiple (in line with its peers) to reflect the Company's high-growth trajectory and early-stage earnings profile. We believe this valuation is supported by its strong shipment ramp-up, expanding role in China's AI infrastructure buildout, and positioning within a structurally evolving, system-level compute architecture.

Industry Overview

China's GPGPU market is entering a high-growth phase, with inference and localization driving the next leg of expansion

According to F&S, China's GPGPU market is projected to grow at a **29.5% CAGR over 2025–29E**, supported by expanding AI adoption, continued cloud infrastructure buildout, and steady progress by domestic chipmakers. Within this, domestic vendors are expected to grow materially faster than foreign peers, with market size rising at a **91.4% CAGR** versus **20.0%** for overseas suppliers, implying China's GPGPU localization rate will increase from **6.5% in 2025E** to **31.0% by 2029E**. In our view, the more important takeaway is not only that the market is expanding rapidly, but that the structure of demand is shifting in a way that should gradually improve the competitive position of domestic vendors.

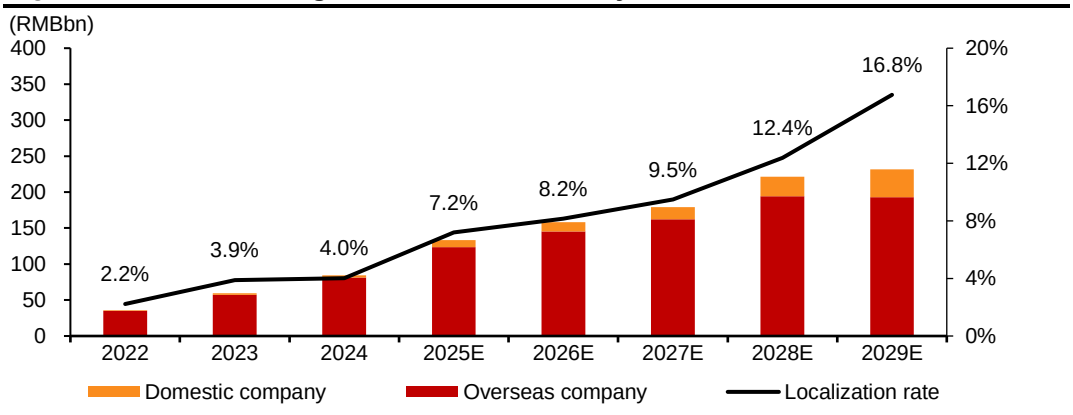
Figure 1: China's GPGPU market size, by revenue



Source: F&S, CMBIGM

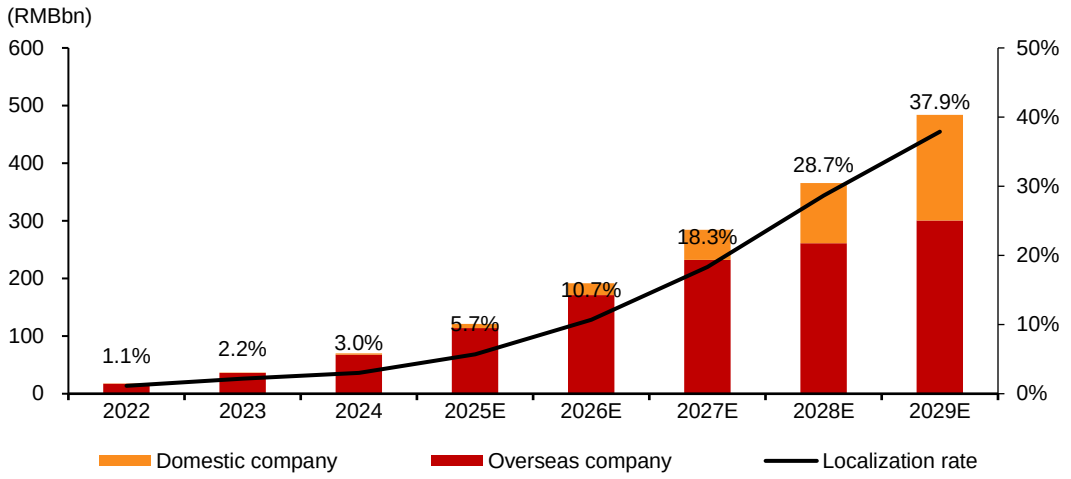
From a product perspective, the market is increasingly bifurcated between training and inference, with each rewarding different capabilities. Training GPGPUs remain focused on compute-intensive model development, where memory bandwidth, parallel processing capability, and cluster performance are the key competitive metrics. Inference GPGPUs, by contrast, are increasingly tied to real-world deployment, where low latency, energy efficiency, software compatibility, and implementation flexibility matter more. This distinction is becoming increasingly important because the fastest-growing part of the market is also the one where the basis of competition is gradually shifting away from pure peak silicon performance toward broader system-level capability.

Figure 2: China's training GPGPU market size, by revenue



Source: F&S, CMBIGM

Figure 3: China's inference GPGPU market size, by revenue



Source: F&S, CMBIGM

This shift also has important implications for localization. Domestic substitution is likely to progress more slowly in training, where performance requirements remain closer to the global frontier, but should have a more realistic path in inference, where deployment breadth, ecosystem adaptation, and workload-specific optimization play a larger role.

According to F&S, localization in China's training GPGPU market is expected to rise from **4.0% in 2024** to **16.8% by 2029E**, while inference localization is projected to increase much faster, from **3.0%** to **37.9%** over the same period. In our view, this is an important signal that domestic vendors are more likely to gain share first through broader inference and deployment-driven workloads before meaningfully narrowing the gap in high-end training.

Figure 4: Classification of GPGPU chips: Training purpose vs. inference purpose

AI GPGPU chips	Primary Function	Key Requirements	Typical Applications	Description
Training	Model training, learning from large datasets	Large memory bandwidth; parallel processing capability	LLM training; complex model developments	Handle extremely intensive computational workloads, such as large-scale data processing and complex matrix calculations required to train AI models.
Inference	Real-time prediction; model deployments	Low latency; energy efficiency	Cloud inference services; edge computing	Focus on efficiency, enabling fast, low-latency execution of trained models in real-world applications.

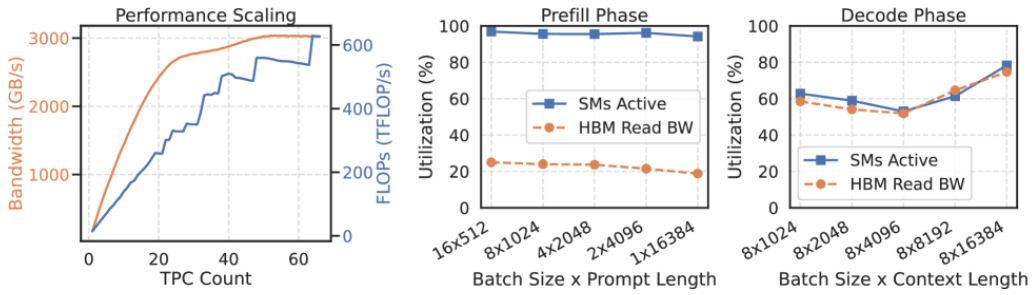
Source: F&S, CMBIGM

On the volume level, F&S states that overall GPU shipments reached 1.6mn units in 2024, representing a 72.8% 2022-2024 CAGR.

PD disaggregation is turning inference from a single-product market into a workload-specialized cluster opportunity

Large-model inference is increasingly being deployed as two separate workloads rather than one unified compute task. **Prefill, which processes the input prompt, is primarily compute-intensive, while decode, which generates output tokens sequentially, is more constrained by memory bandwidth and cache management.** As model responses become longer, especially in reasoning use cases, this mismatch becomes more pronounced, making a single hardware configuration less efficient to serve both prefill and decode on the same hardware configuration.

Figure 5: Left: Profiled HBM bandwidth and FLOPs vs. active TPCs; right: Resource utilization during prefill and decode phases



Source: DuetServe: Harmonizing Prefill and Decode for LLM Serving via Adaptive GPU Multiplexing, arXiv:2511.04791 (2025)
Note: TPC – Texture Processing Cluster

The physical asymmetry between prefill and decode is also translating into measurable economic impact, visible in frontier API pricing. Input and output token prices are the commercial surface through which prefill and decode costs are monetized, so the spread between them offers a useful market signal of the relative economics of the two workloads.

That spread has widened materially over time. Based on provider's published API pricing, in the GPT-3.5 era of 2023, output tokens were typically priced at roughly 1.3–2.0x the rate of input tokens, implying only modest cost differentiation between the two phases. **By 2026, the spread has widened structurally to roughly 5–8x across major frontier providers**. While pricing is not a pure cost proxy, the consistency of this widening across leading platforms suggests that prefill and decode have become increasingly distinct economic workloads in production inference.

Figure 6: Output-to-input token pricing ratio across major frontier LLM providers

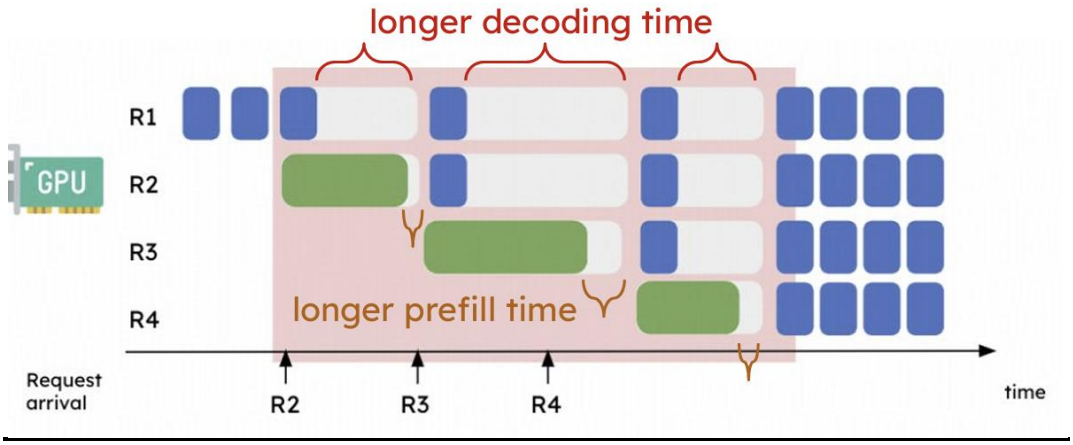
Provider	Model	Input ($/1M Token)	Output ($/1M Token)	Output/Input Ratio
Anthropic	Opus 4.6	$5.00	$25.00	5.0×
	Sonnet 4.6	$3.00	$15.00	5.0×
	Haiku 4.5	$1.00	$5.00	5.0×
OpenAI	GPT-5.4	$2.50	$15.00	6.0×
	GPT-5.4 mini	$0.75	$4.50	6.0×
	GPT-5.4 nano	$0.20	$1.25	6.25×
	GPT-5.4 pro	$30.00	$180.00	6.0×
Google	Gemini 3.1 Pro Preview	$2.00	$12.00	6.0×
	Gemini 2.5 Pro	$1.25	$10.00	8.0×
	Gemini 2.5 Flash	$0.30	$2.50	8.3×
	Gemini 2.5 Flash-Lite	$0.10	$0.40	4.0×

Source: Company data, CMBIGM

Such economic separation is now translating into physical separation at the deployment layer. By late 2025, production inference systems across major global and Chinese

platforms had already begun adopting disaggregated serving frameworks, including deployments disclosed by SGLang, Moonshot AI, Perplexity, Alibaba Cloud, and NVIDIA's Dynamo ecosystem. The implication is clear: **Inference infrastructure is moving from a single-pool model toward workload-specialized clusters, where prefill and decode resources are deployed, scheduled, and optimized separately.**

Figure 7: A single GPU pool creates queueing inefficiency when prefill and decode are mixed



Source: DistServe: Disaggregating Prefill and Decoding for Goodput-optimized Large Language Model Serving, arXiv:2401.09670

The broader industry direction reinforces this view. Global platform leaders are increasingly moving toward workload-specialized inference infrastructure, led by NVIDIA and Huawei. In China, Huawei has laid out the clearest silicon-level roadmap: **Ascend 950PR and 950DT use the same 950 die, but are split by workload**, with 950PR targeted at prefill and recommendation using lower-cost HiBL memory and already launched in 1Q26, while 950DT is aimed at decode and training with 144GB HiZQ 2.0 memory, 4 TB/s bandwidth, with expected availability in 4Q26.

NVIDIA is moving in the same strategic direction, albeit through a different implementation path. The Vera Rubin platform now points toward a more heterogeneous inference architecture, pairing general-purpose GPUs with dedicated low-latency decode acceleration. In effect, both Huawei and NVIDIA are converging on the same underlying conclusion: As inference workloads become more specialized, value increasingly shifts toward architectures that separate workload roles across different compute engines.

Figure 8: Huawei's prefill and decode products

Specification	910C	950 PR	950 DT
Target workload	General	Prefill	Decode
Production date	1Q25	1Q26	4Q26
Memory size	128 GB HBM2e	128GB HiBL 1.0	144GB HiZQ 2.0
Memory bandwidth	3.2 TB/s	1.6 TB/s	**4 TB /s**

Source: Company data, Tom's Hardware, CMBIGM

Figure 9: Nvidia's CPX vs. LPU

Specification	B200	Rubin CPX	LPU
Target workload	General	Prefill	Decode
Production date	2024	N/A	1Q26
Memory size	180GB HBM3e	128GB GDDR7	500 MB SRAM
Memory bandwidth	8 TB/s	1.2 TB/s	**150 TB/s**

Source: Company data, CMBIGM

Within this evolving landscape, Iluvatar CoreX disclosed DeepSeek R1 671B benchmark results suggest it is already competitive in prefill-oriented inference workloads, delivering lower TTFT [Time to First Token] and higher throughput than NVIDIA's A800 under the disclosed setup. While the comparison is not precision-matched, the result is directionally important in our view because it supports Iluvatar's relevance as a potential prefill-side participant in disaggregated inference clusters.

Figure 10: DeepSeek R1 671B prefill performance (NVIDIA A800 baseline)

TP	PP	Parameters	Tokens	TTFT (ms)	Throughput (tokens/s)
2	8	671B	1024	**735.67**	**11,135**
2	8	671B	2048	**1,343.85**	**12,192**
2	8	671B	4096	**2,749.54**	**11,918**

Source: Company data, CMBIGM

Note: TP – tensor parallelism, PP – pipeline parallelism; the benchmark comparison is not precision-matched: the Company's chip ran the DeepSeek model under W4A8 quantization, while the A800 ran W4A16 quantization

Figure 11: DeepSeek R1 671B prefill performance (Iluvatar platform, optimized configuration)

TP	PP	Parameters	Tokens	TTFT (ms)	Throughput (tokens/s)
2	8	671B	1024	**397.93**	**20,586**
2	8	671B	2048	**770.20**	**21,272**
2	8	671B	4096	**1627.90**	**20,129**

Source: Company data, CMBIGM

For the domestic GPU sector, PD disaggregation matters because it changes the basis of competition. **We think the opportunity is not only about replacing imported accelerators one for one. It is increasingly about whether a local vendor can fit into a more complex inference stack that separates compute pools, requires tighter software integration, and rewards deployment readiness as much as peak chip specifications.**

That matters for market structure. In a single-SKU procurement model, domestic vendors often compete mainly on compliance, availability, and entry pricing. In a disaggregated procurement model, vendors that can address specific workload roles effectively, while integrating into production frameworks with lower migration friction, should be able to capture a larger portion of the deployment. This raises the strategic value of software compatibility and multi-product breadth across the domestic GPGPU landscape.

A deployment model that captures this opportunity is the mixed-vendor disaggregated cluster, in which domestic hardware is introduced on the prefill side while hardware with enhanced performance is retained on the decode side. For Chinese hyperscalers, the appeal is that it delivers a meaningful cluster-level improvement in both total cost of ownership and effective throughput per dollar of hardware spend, without forcing a full ecosystem migration.

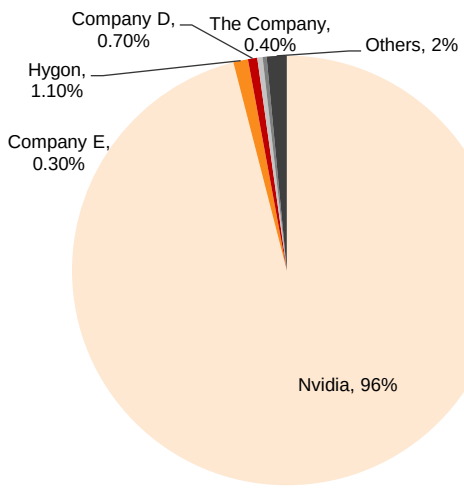
In our view, this is one of the clearest ways to link the broader domestic GPU thesis with the Company's product positioning: **Domestic substitution provides the demand floor, while PD disaggregation creates the possibility of share redistribution inside that demand pool.** The prefill side offers the most accessible near-term participation route for domestic hardware in mixed-vendor clusters, with the opportunity to expand into broader inference roles as disaggregated architectures become the procurement default.

Competitive Landscape

While overseas companies currently dominate, domestic manufacturers are gaining shares

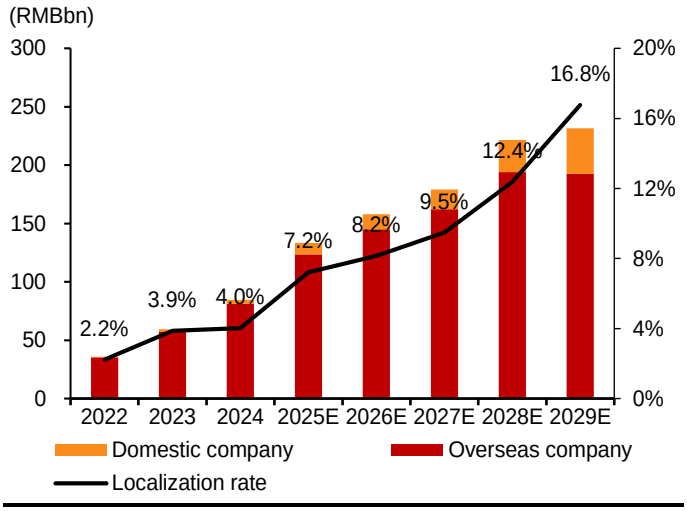
In the training GPGPU segment, foreign companies still dominate China's market, with the leading overseas player holding 96.0% share by revenue in 2024, but domestic vendors are gradually gaining traction as localization improves. According to Frost & Sullivan, the localization rate of China's training GPGPU market is expected to rise from 4.0% in 2024 to 16.8% by 2029, with domestic training GPGPU revenue projected to grow at a 52.0% CAGR over 2025–2029, versus 11.7% for foreign companies. Importantly, four of the top five participants in China's training GPGPU market in 2024 were domestic companies, showing the rising relevance of local players despite the market's still highly concentrated competitive structure (F&S).

Figure 12: China's training GPGPU market: Dominated by overseas suppliers in 2024



Source: F&S, CMBIGM estimates

Figure 13: China's training GPGPU market size and localization rate trend, 2022–2029E



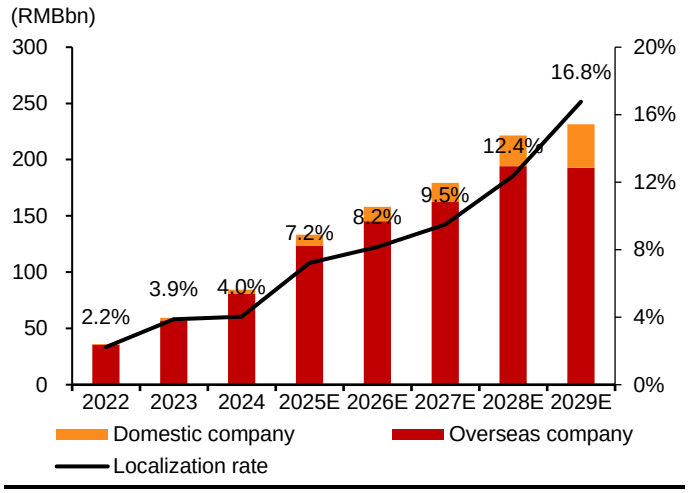
Source: F&S, CMBIGM

The inference GPGPU segment presents an even larger long-term opportunity for domestic manufacturers. While foreign companies still dominate the market today, F&S estimates the localization rate of China's inference GPGPU market was only 3.0% in 2024, implying foreign vendors accounted for roughly 97% of market revenue. That said, domestic players are expected to scale much faster as AI deployment broadens and local technological competitiveness improves.

We believe the primary battleground for domestic semiconductor substitution has shifted decisively toward inference, as China's AI ecosystem transitions from initial model training to large-scale commercial deployment. This view is substantiated by F&S, which forecasts an expansion in the domestic inference GPGPU segment, projecting a 126.9% revenue CAGR from 2025 to 2029. This domestic growth trajectory significantly outpaces the 27.4% CAGR estimated for foreign peers over the same period, with the domestic penetration rate climbing from a mere 3.0% in 2024 to a 37.9% by 2029, reinforcing the case for localized inference as the most critical growth driver in the mid-to-long term.

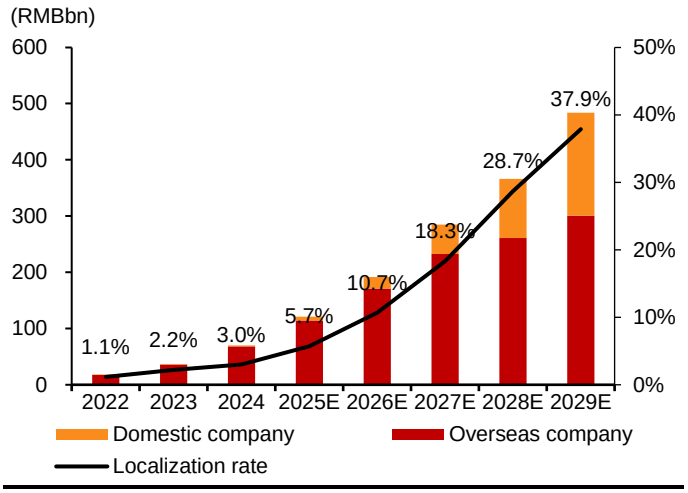
According to F&S, domestic inference GPGPU revenue is projected to grow at a 126.9% CAGR from 2025 to 2029, far outpacing the 27.4% CAGR for foreign companies, with the localization rate expected to rise sharply from 3.0% in 2024 to 37.9% by 2029. This suggests inference should be the most important battleground for domestic substitution as China's AI market shifts from model training toward large-scale commercial deployment.

Figure 14: China's training GPGPU market: Domestic suppliers to take 17% share by 2029E



Source: F&S, CMBIGM

Figure 15: China's inference GPGPU market: Domestic suppliers to take 38% share by 2029E



Source: F&S, CMBIGM

Heterogeneous compute as an emerging direction for inference infrastructure

The prefill-decode split is, in our view, only one expression of a broader shift taking shape across the inference infrastructure stack. As model architectures, context lengths, and deployment patterns continue to diverge, the assumption that a single class of accelerator can remain the most efficient unit of compute for the entire inference pipeline is being reexamined. Evidence supporting this direction has begun to accumulate. NVIDIA released Dynamo at GTC 2025, an open-source framework that splits prefill and decode across distinct GPU pools and has since been integrated by major hyperscaler platforms. In April 2026, SambaNova and Intel jointly disclosed a heterogeneous inference blueprint that maps prefill, decode, and agentic operations onto three different classes of silicon, with production availability targeted for 2H26. These efforts suggest that future inference clusters are likely to be composed of multiple accelerator types, each tuned to the workload it is structurally better suited to handle.

Heterogeneous inference today sits closer to an architectural and research direction than to an established procurement model, and several key uncertainties remain, including the durability of specific workload boundaries, the role each accelerator class will ultimately play, and the maturity of the software stacks needed to orchestrate across them. The long-term opportunity for specialized compute, including for domestic GPGPU vendors capable of occupying specific workload niches, extends beyond simply replacing foreign chips one for one.

Bottleneck remains on the manufacturing end for domestic AI chip designers

Domestic GPGPU performance still trails global leaders by several years at the manufacturing frontier, in our view, largely because access to leading-edge manufacturing capacity remains structurally constrained. For most Chinese fabless AI chipmakers, the key bottleneck is no longer only architecture or software capability, but whether they can secure stable advanced-node supply. This constraint is particularly acute domestically, where SMIC (981 HK, NR) is currently the only Chinese foundry reported by Reuters to have demonstrated 7nm-class production capability through its N+2 technology, while a second local supplier, Hua Hong Group, is only beginning to move toward 7nm (link).

Figure 16: US and allies' export controls and the rise of China's domestic chip manufacturing capabilities

Date	US and allies' actions	China's progress
Sep 2022	NSA Jake Sullivan declares the US must maintain "as large a lead as possible" in AI and force-multiplier technologies.	**Pre-controls stockpiling:** Chinese firms begin accelerating purchases of semiconductor equipment and chips in anticipation of restrictions. Equipment imports surge. **Big Fund II:** Continued deployment of US$29bn fund targeting mature processes and equipment localization.
Oct 7, 2022	**First Export Controls Package:** BIS restricts advanced chips (TPP ≥4,800 GOPS + interconnect speed ≥600 GB/s), EDA software, SME, and critical components	**YMTC "Wudang Mountain" project:** Following Entity List designation (Dec 2022), YMTC launches systematic plan to remove US tools from production lines and transition to domestic equipment
Early 2023	**Legal circumvention exposed:** NVIDIA introduces H800 (300 GB/s) and A800 (400 GB/s), modified chips engineered to fall just below the 600 GB/s interconnect speed limit	**Equipment import surge:** Chinese semiconductor equipment imports jump from US$2.9bn (Jun–Jul 2022) to US$5.0bn (Jun–Jul 2023), primarily from the Netherlands and Japan
Mar 2023	**Netherlands restricts DUV exports:** Following U.S. diplomatic pressure, the Netherlands blocks exports of ASML's most advanced DUV machines (NXT:2000i) capable of sub-14 nm fabrication	**SMEE 28nm lithography:** Shanghai Micro Electronics Equipment (SMEE) reportedly developing 28 nm-process lithography machines
Sep 2023		**Huawei Mate 60 Pro:** Contains SMIC-fabricated 7nm Kirin 9000S chip using multi-patterning DUV lithography, well below the 14/16 nm threshold that export controls assumed China couldn't surpass
Oct 2023	**Updated export controls:** Drops interconnect speed criterion; slashes TPP threshold from 4,800 to 1,600 GOPS, and adds performance density and datacenter criteria	
Dec 2023	**Raimondo calls out NVIDIA:** Commerce Secretary publicly criticizes NVIDIA at Reagan National Defense Forum for designing chips specifically to circumvent export controls	**Huawei sales surge:** Huawei projects 60mn Mate-series smartphones for 2024 — all using domestically fabricated SMIC 7 nm chips
Mid 2024	US still allowed capped-performance, China-specific AI GPUs such as NVIDIA's H20 to be shipped	**Big Fund Phase III:** China's largest chip investment yet, US$47.5bn, focused on advanced packaging, alternative materials, and equipment localization to bypass EUV restrictions
2025-2026	US moved from broad tightening to selective licensing, permitting some H200 exports to China in Jan 2026 subject to case-by-case review and usage/supply conditions	**CXMT begins HBM2 production:** CXMT starts mass production of HBM2, with plans for HBM3 (2026E) and HBM3E (2027E)

Source: US Center for Strategic and International Studies, TechInsights, Reuters, CMBIGM

The Company appears better positioned than some local peers because its products fall below the US export-control performance thresholds introduced in 2022, which has historically given it greater flexibility to continue taping out at overseas fabs. Even so, we would still expect the Company to pursue domestic manufacturing alternatives over time to reduce geopolitical risk and secure a more sustainable supply path should export controls tighten further.

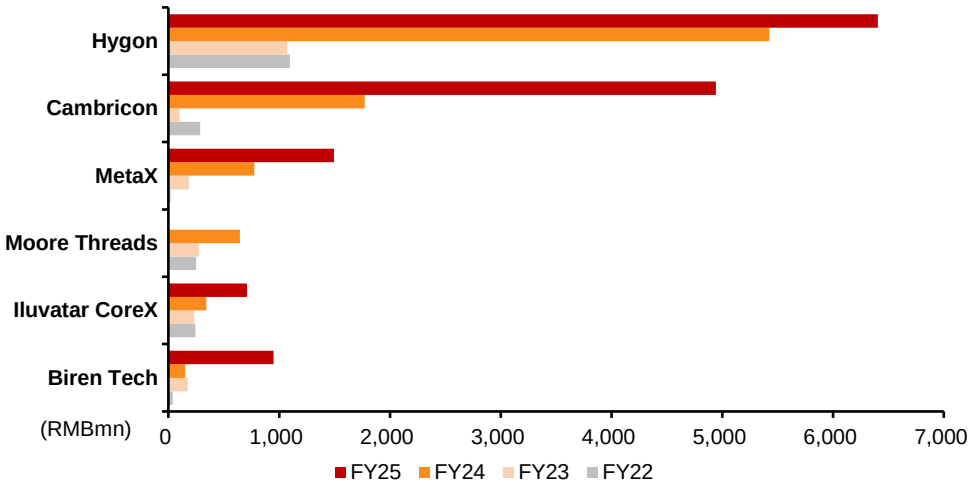
Figure 17: US export controls

Regulatory Tier	Official Control Thresholds
ECCN 3A090.a / 3A090.b core control rule	**Total Processing Performance (TPP) is ≥ 4,800 or if TPP is ≥ 1,600 and Performance Density (PD) is ≥ 5.92.**
2026 tiered review rule (effective January 15, 2026)	For exports to end-users in the mainland or Macau of China, chips with: **TPP below 21,000 and total DRAM bandwidth below 6,500 GB/s** may qualify for a narrower review track, subject to additional certifications and testing requirements. All other configurations, including products above either threshold, remain outside this carve-out

Source: Federal Register, BIS, CMBIGM

Inventory build-up across domestic AI chip players reflects a combination of production ramp-up, strategic stockpiling under tighter US export controls, and still-developing demand visibility. From 2022 to 2025, most domestic GPGPU vendors reported materially higher inventory levels, with larger players such as Hygon and Cambricon (688256 CH, NR) carrying the highest absolute balances, likely benefiting from stronger procurement capacity and a greater ability to secure key components ahead of supply-chain disruptions. Meanwhile, smaller-scale players such as **Moore Threads (688795 CH, NR), MetaX (688802CH, NR), Biren Tech (6082 HK, NR)**, and the Company also saw inventory rise in 2025, indicating that supply-chain buffering has become a broad industry response as China's AI infrastructure investment and domestic substitution efforts continue to build.

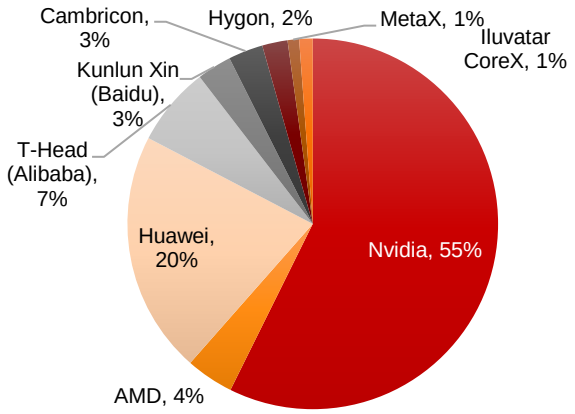
Figure 18: Domestic AI chip players' inventory levels from 2022-2025



Source: Bloomberg, Company data, CMBIGM

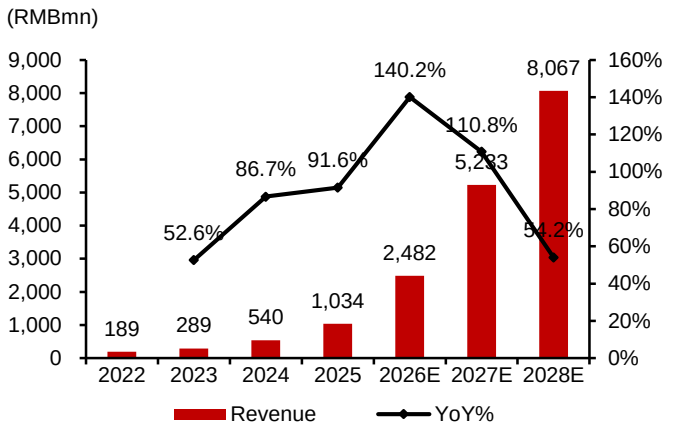
At the same time, the market backdrop in 2025 was supportive for domestic players. Uncertainty around Nvidia's China product roadmap and the timing of H200 shipments created room for local AI chip vendors to ramp. **Reuters, citing IDC, reported that Chinese GPU and AI chipmakers captured about 41% of China's AI accelerator server market in 2025**, while total AI accelerator shipments in China reached roughly 4.0mn units. **Domestic vendors collectively shipped about 1.65mn units, versus 2.2mn for Nvidia**, implying that local adoption has already moved well beyond a niche substitution story. **Against that backdrop, the Company's training GPU series saw shipment growth of 131.5% YoY while its inference GPU series increased by 194.8% YoY in 2025, per mgmt..**

Figure 19: China's AI GPU market share by shipment in 2025



Source: Company data, IDC, Reuters, CMBIGM

Figure 20: The Company's revenue and growth



Source: Company data, CMBIGM estimates

Company Overview

A first-mover domestic GPGPU company spanning training, inference and AI computing solutions

The Company offers GPGPU chips, accelerators, and AI computing solutions, including GPGPU servers and clusters. According to F&S, among China's chip designers, it was the first to achieve mass production of inference GPGPU chips, the first to mass produce training GPGPU chips, and the first to accomplish these milestones using advanced 7nm process technology.

Founded in December 2015, the Company has developed around a hardware-software co-design philosophy and maintains a three-generation R&D cadence of one generation in mass production, one in design, and one in pre-research. On the product side, it has commercialized multiple generations of training- and inference-focused GPGPUs: TG Gen 1 achieved mass production in September 2021; the ZK series was launched in December 2022, with both ZK Gen 1 and ZK Gen 1X entering mass production in February 2023; and TG Gen 2 achieved mass production in 4Q23. During 2025, the Company continued to iterate its TG and ZK product lines, while advancing next-generation training and inference products. As of end-2025, it had served over 340 customers across various industries, with products and solutions deployed in over 1,000 projects across sectors including financial services, healthcare, transportation, manufacturing, retail, research, and education.

Figure 21: Iluvatar's milestone

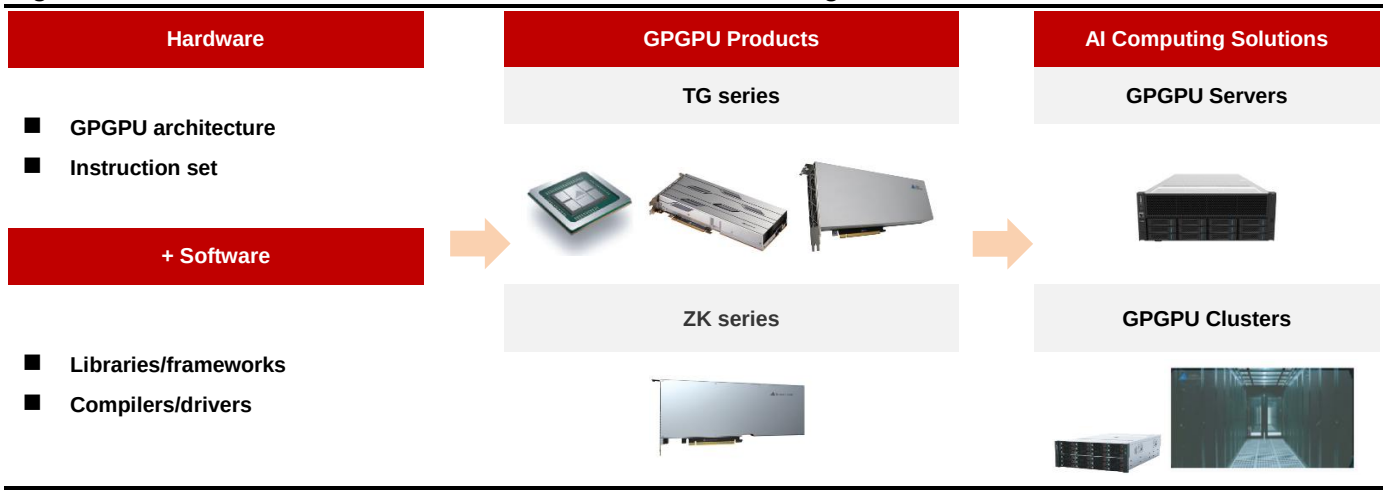
2015	2018	2021	2022	2023	2024
Company established	Launched the design of GPGPU	**Launched China's first GPGPU product: TG Gen 1 (training)**, and achieved mass production	**Launched inference series GPGPU: ZK Gen 1 and ZK Gen 1X**	**Launched TG Gen 2 (training)**, and achieved mass production	**Launched TG Gen 3 (training)**

Source: Company data, CMBIGM

Broad product portfolio supported by hardware-software integration

Iluvatar offers GPGPU products and AI computing solutions across both training and inference scenarios, centered on its TG training series and ZK inference series. Beyond standalone chips and accelerators, it also provides GPGPU servers and scalable clusters that integrate its proprietary products with third-party infrastructure and software, allowing deployment across a wide range of customer environments. The Company's key differentiation lies in its hardware-software co-design capability: A proprietary software stack spanning compilers, drivers, libraries and frameworks enables compatibility with mainstream GPGPU programming ecosystems and lowers migration friction. In 2025, Iluvatar further upgraded this platform with native compatibility with mainstream GPU programming models, improving code migration efficiency by more than 80%, while its proprietary acceleration libraries and large-model toolkit improved performance, utilization and cluster communication efficiency materially.

Figure 22: Iluvatar's business model: software-hardware co-design

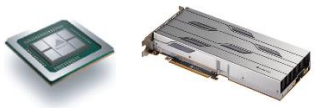


Source: Company data, F&S, CMBIGM

Iluvatar provides two series of GPGPU products, targeting different computing needs for training and inference scenarios. The former requires intensive computational power to ensure model flexibility while maximizing performance, while the latter emphasizes efficient model deployment and execution, delivering optimal performance under broader constraints. **Through multiple generations of architectural evolution, Iluvatar has enhanced product performance, compatibility and cross-scenario adaptability.**

- **TG Series (training-focused, commercialized two generations):** As Iluvatar's flagship training product line, the TG series is designed for AI model training and features optimized performance computing cores, memory configurations and architectural enhancements to support multi-GPU cluster systems. According to Frost & Sullivan, Iluvatar was the first among China's chip designers to mass produce training GPGPU chips. TG Gen 1 achieved mass production in September 2021, while TG Gen 2 entered mass production in 4Q23 and became the main revenue driver within the training series in 2024. **In 2025, TG series revenue rose 116.7% YoY to RMB583.7mn, supported by stronger demand**, especially for TG Gen 2, as well as accelerated sales of TG Gen 1 inventories. The next-generation TG series remains under active development, with a focus on improving performance for large-scale AI training applications.
- **ZK Series (inference-focused, multiple commercialized products):** The ZK series is Iluvatar's dedicated inference line, designed for cloud and edge inference applications with enhanced integer computing units and optimized data channels to deliver efficient, low-latency execution. According to Frost & Sullivan, it is China's first GPGPU product specifically designed for inference tasks. ZK Gen 1 and ZK Gen 1X both entered mass production in February 2023. By 2024, ZK Gen 1 had become the dominant revenue contributor within the inference series as its shipment volume increased and exceeded that of ZK Gen 1X. **In 2025, ZK series revenue increased 238.2% YoY to RMB338.9mn**, reflecting growing inference demand and stronger sales execution. The next-generation ZK series is also under active development, with targeted optimizations for emerging large language models, lower-precision data types and mixed-precision computing.

Figure 23: Iluvatar's products and solutions offerings

Products	Product image	Launch date	Mass production	Key notes
GPGPU Products				
Training GPGPU products – TG series				
TG Gen 1		Mar. 2021	Sep. 2021	China's first domestically mass-produced GPGPU product. Supports a general-purpose instruction set, mixed-precision computing, vector and tensor computing, and peer-to-peer chip communication.
TG Gen 2		Sep. 2023	4Q23	Enhanced training-focused GPGPU with improved application performance and architecture efficiency. Supports mainstream deep learning frameworks, model acceleration operators and large-scale cluster scheduling.
TG Gen 3		*3Q24*	*Expected Late-2026*	*Further enhances computing performance, with significant improvements in large-scale cluster efficiency and connectivity. Supports advanced precision requirements for large AI models, increased memory capacity, PCIe Gen5, higher peer-to-peer bandwidth and multi-card architecture.*
Inference GPGPU products – ZK series				
ZK Gen 1		Dec. 2022	Feb. 2023	China's first domestic inference-focused GPGPU product. Positioned at a higher price point than ZK Gen 1X, with shipment volume increasing significantly and exceeding ZK Gen 1X in 2024
ZK Gen 1X		Dec. 2022	Feb. 2023	Lower-power inference product targeting edge and client-edge scenarios. Commercialized alongside ZK Gen 1 in February 2023
ZK Gen 3		*Expected 1Q26*	*Expected 2Q26*	*Next-generation inference-focused GPGPU optimized for cloud inference workloads. Selected features include significant performance improvement over the previous generation and stronger performance-to-cost ratio for mainstream large-model inference and training applications*
AI Computing Solutions				
GPGPU Servers		Combine multiple GPGPU accelerators with integrated software stack, typically used for enterprise AI workloads and large-model deployment		
GPGPU Clusters		Integrate Iluvatar GPGPU products and software stack with third-party servers, storage and networking infrastructure to support large-scale AI training and inference deployments. In FY25, AI computing solutions generated RMB96.1mn revenue and the Company disclosed proprietary cluster management, scheduling and communication technologies		

Source: Company data, CMBIGM

Taken together, the Company's TG and ZK product lines, together with its AI computing solutions, indicate that the Company is trying to compete not just as a standalone chip vendor, but as a broader hardware-software platform provider. The next question is whether its current-generation hardware can deliver competitive performance in the workloads that matter most for actual deployment. In our view, third-party benchmarking is particularly helpful because it provides an external check on whether the Company's system-level positioning is supported by underlying silicon performance.

Third-party benchmarking supports competitive real-world performance

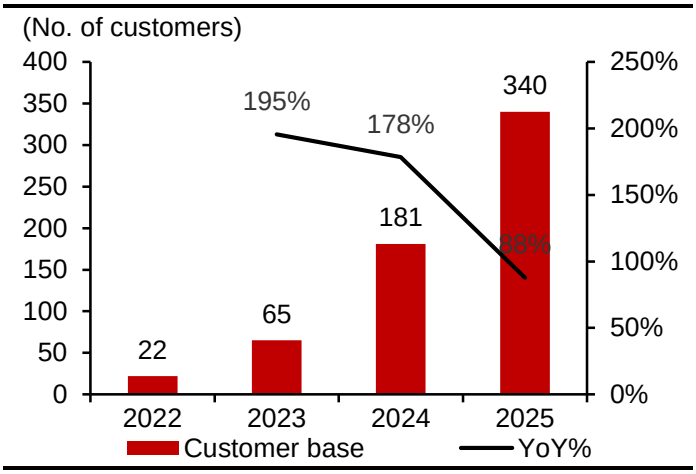
ByteMLPerf, an AI accelerator benchmark that focuses on evaluating AI accelerators from practical production perspective from ByteDance, provides a useful third-party perspective on the Company's current-generation hardware relative to NVIDIA's A800 under production-oriented workloads (link). Based on ByteMLPerf testing results, the Company's product can outperform A800 on selected tests, and appears structurally stronger in the parts of the inference stack most relevant to prefill-oriented deployment, while still showing visible limitations in interconnect and lower-precision support.

- **Compute efficiency on production inference workloads**. On INT8 GEMM and FlashAttention BF16 (the two operators that dominate production prefill workloads), the Company's chip achieved compute utilization (MFU) at or above A800 across most tested shapes and sequence lengths, with the gap most pronounced at large matrix dimensions and longer contexts.
- **Memory bandwidth and instruction throughput**. On a representative decode-side memory-bound workload (AWQ GEMV), memory bandwidth utilization (MBU) reached over 90% across batch sizes, supported by the chip's asynchronous SME Engine DMA design. Vector and special-function units also reached close to theoretical peak on standard FP32 and half-precision instructions.
- **Interconnect and parallelism strategy.** The chip's PCIe Gen 4 inter-chip bandwidth sits materially below A800's NVLink-class interconnect, which constrains tensor-parallel scaling and shapes the Company's preference for pipeline-parallel deployment. Communication–computation overlap optimizations on the GEMM + All-Reduce path partially offset this gap by hiding collective latency inside compute time.
- **End-to-end model performance.** On LLaMA2 7B prefill, the chip's TTFT (time-to-first-token) was roughly twice that of A800 at short input lengths, while throughput surpassed A800 at longer token counts. On DeepSeek R1 671B prefill, the Company's chip delivered materially lower TTFT and higher throughput than A800 across all tested input lengths, though we note this comparison is not precision-matched, with the Company's chip running W4A8 (INT8 compute) versus A800 on W4A16 (BF16 compute).
- Taken together, these results suggest that the Company's chip is already capable of delivering meaningful hardware efficiency on the production inference workloads most relevant to actual deployment. The chip's SIMT-based GPGPU architecture also provides a degree of generality and flexibility that should help it adapt to future shifts in large-model architecture, while keeping developer learning curves and migration costs relatively low. But interconnect bandwidth and native FP8/FP4 support remain visible architectural gaps, which the Company has indicated it intends to address in its next-generation product line.

A broad and expanding customer base across multiple industries

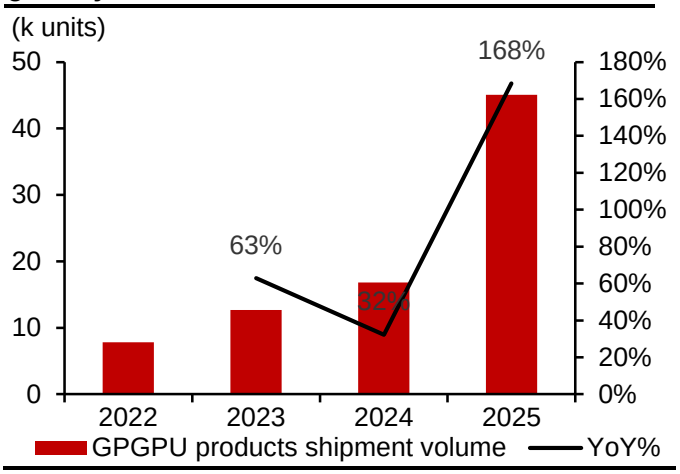
Iluvatar has built a broad customer base across a wide range of end markets, reflecting growing commercial adoption of its GPGPU products and AI computing solutions. According to our analysis, by the end of 2025, the Company had shipped GPGPUs to more than **230** customers and supported over **800** deployments across sectors including finance, healthcare, transportation, manufacturing, retail, and research. As of 2025 year-end, the Company had served over **340** customers, with products and solutions deployed in over **1,000** projects.

Figure 24: Iluvatar's customer base increased from 22 in 2022 to 340 in 2025



Source: Company data, CMBIGM

Figure 25: Iluvatar's total GPGPU shipment volume grew by 168% YoY in 2025



Source: Company data, CMBIGM estimates

Iluvatar's real-world applications, from quantitative trading and medical imaging to autonomous retail and research simulations, **demonstrate not only the versatility and scalability of its technology but also the company's strong R&D capabilities and maturity in delivering industrially validated AI compute solutions.**

Figure 26: Application of Iluvatar's GPGPU products and solutions across key industries

	Financial Service	Healthcare	Retail	Education
Industry challenges	Growing need to process complex data and support intelligent decision-making under high performance and reliability requirements	Growing need to modernize diagnostic capabilities and clinical workflows through advanced AI computing	Increasing demand for digital transformation to enhance customer experience and optimize store operations	Rising demand to modernize both teaching and research capabilities through advanced computing power
Application scenarios	Iluvatar's products and solutions have been deployed across the financial sector, including intelligent consultation and decision-support type workloads, and FY25 also saw continued growth in projects across finance and large-model applications	Iluvatar supports two key healthcare scenarios: medical imaging analysis and intelligent clinical support. Its integrated platforms combine deep learning with specialized hardware acceleration for RT imaging, landmark detection and PET/MR correction, while its large-model-based solutions support intelligent consultation, decision support, structured EHR management and automated medical documentation. FY25 also witnessed deep adaptation in healthcare and collaboration with Grade A tertiary hospitals to promote clinical deployment of AI-assisted diagnosis	Iluvatar supports smart-store operations and intelligent shopping experiences. Its solutions enable store digitalization through IoT platforms and business analytics, with significant improvement in inspection efficiency, while smart-cart solutions can achieve over 99.5% autonomous product-recognition accuracy, enabling automatic identification and contactless checkout. FY25 further noted the ZK series has been applied in commercial retail	Iluvatar supports both scientific research computing and classroom teaching applications. For research, its products incorporate PINN-based capabilities to support complex simulations such as computational fluid dynamics and Navier-Stokes processing, as well as broader numerical simulations in molecular dynamics, protein folding, atmospheric science and geological exploration. For teaching, it integrates development tools and simulation modules to enable hands-on learning, interactive experiments and computational visualization. FY25 also recorded continued growth in projects across education and scientific research

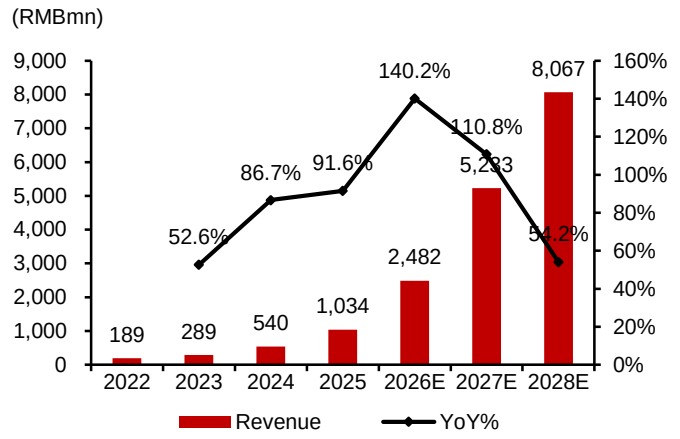
Source: Company data, CMBIGM

Financial Analysis

Strengthening earnings profile on growing demand and improving scale

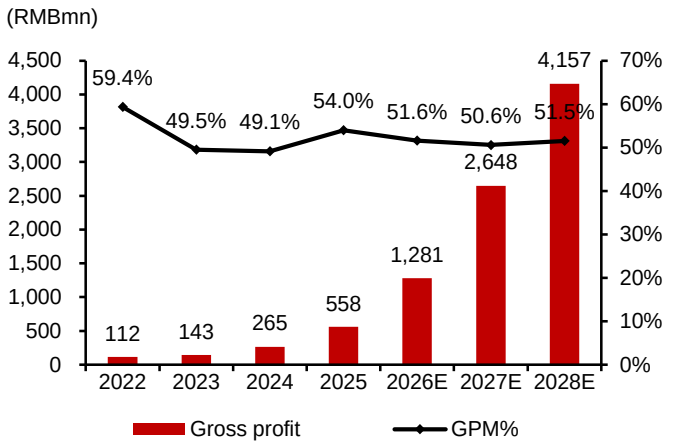
We forecast a significant revenue inflection for the Company, with YoY growth of **140%/111%/54% over FY26-28E** driven by sustained momentum in both training and inference GPU shipments. While 2026 growth will be largely supported by robust inference demand, the **2Q26 launch of the TG Gen 3** platform represents a key medium-term catalyst; we expect shipments to scale in late-2026 and contribute meaningfully to the top line from 2027 onward. We project **GPM to remain resilient at 51.6% in 2026E**, and stabilize in the **50.6%-51.5% range through 2028E**, as the Company absorbs typical margin volatility associated with new product ramps. In our view, the key earnings inflection is less about sharp margin expansion and more about operating leverage, as shipment scale begins to absorb the Company's still-elevated R&D and commercialization base.

Figure 27: Revenue and growth



Source: Company data, CMBIGM estimates

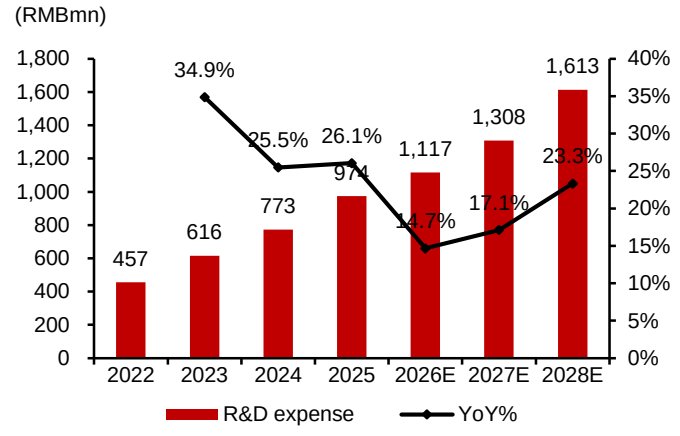
Figure 28: Gross profit and GPM%



Source: Company data, CMBIGM estimates

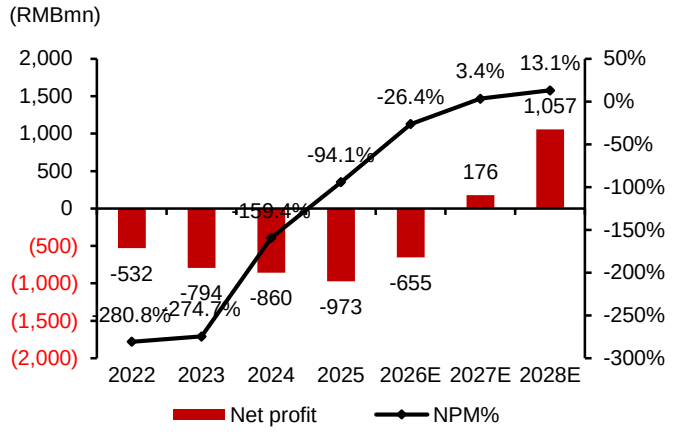
We expect the Company to maintain high R&D intensity to support its aggressive product roadmap and continued architectural iteration. Accordingly, we model R&D expense growth of 15%/17%/23% over FY26–28E. As next-generation GPGPU shipments scale and fixed-cost absorption improves, we project the Company to reach a net profit inflection point in 2027E.

Figure 29: R&D expenses and YoY%



Source: Company data, CMBIGM estimates

Figure 30: Net profit and NPM%



Source: Company data, CMBIGM estimates

The company's revenue breakdown by segment

We project the Company's GPGPU product segment revenue to increase by 143%/112%/54% over FY26–28E, driven primarily by shipment growth and product iteration across both training and inference. We expect AI computing solutions to scale alongside the core GPGPU business, but to remain on a smaller, project-driven commercialization layer rather than the primary growth engine.

Figure 31: The Company's revenue breakdown by segment (2022-2028E)

RMBmn	2022	2023	2024	2025	2026E	2027E	2028E
Revenue by segment							
GPGPU products	**189**	**267**	**370**	**923**	**2,243**	**4,743**	**7,320**
YoY%		*41.6%*	*38.5%*	*149.6%*	*143.1%*	*111.5%*	*54.3%*
%	*100%*	*92%*	*69%*	*89%*	*90%*	*91%*	*91%*
AI computing solutions	**0**	**16**	**166**	**96**	**224**	**474**	**732**
YoY%			*970.8%*	*-42.2%*	*133.5%*	*111.5%*	*54.3%*
% GPGPU products	*0.0%*	*5.8%*	*45.0%*	*10.4%*	*10.0%*	*10.0%*	*10.0%*
Others	**1**	**7**	**4**	**15**	**15**	**15**	**15**
YoY%		*716.3%*	*-44.5%*	*306.7%*	*0.0%*	*0.0%*	*0.0%*
%	*0.4%*	*2.3%*	*0.7%*	*1.4%*	*1%*	*0%*	*0%*
Total	**189**	**289**	**540**	**1,034**	**2,482**	**5,233**	**8,067**
YoY%		*53%*	*87%*	*92%*	*140%*	*111%*	*54%*

Source: Company data, CMBIGM estimates

We expect the Company's consolidated GPM to remain broadly resilient above 50% over the medium term. While new-product ramps may create some near-term mix volatility, we think continued design optimization, improving scale, and a higher contribution from next-generation products should help preserve overall margin stability.

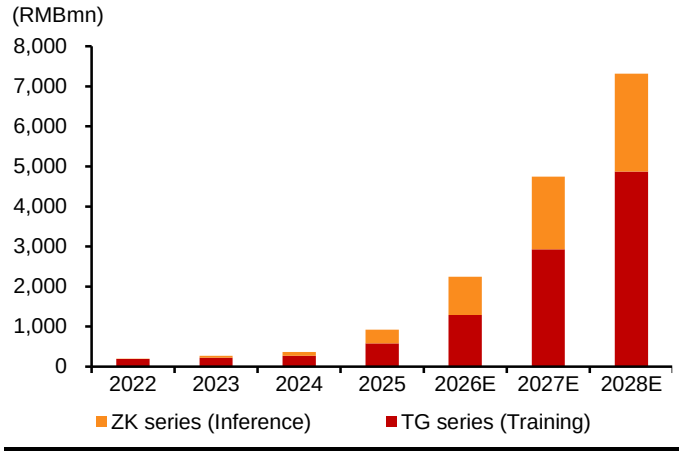
Figure 32: The Company's gross profit breakdown by segment (2022-2028E)

RMBmn	2022	2023	2024	2025	2026E	2027E	2028E
GP by segment							
General GPU products	**112**	**134**	**209**	**508**	**1,181**	**2,448**	**3,854**
YoY%		*20%*	*56%*	*143%*	*133%*	*107%*	*57%*
GPM%	*59.4%*	*50.2%*	*56.6%*	*55.0%*	*52.6%*	*51.6%*	*52.7%*
AI compute power solutions	**0**	**4**	**53**	**40**	**90**	**190**	**293**
YoY%			*1210.3%*	*-24.3%*	*125%*	*111%*	*54%*
GPM%		*25.9%*	*31.7%*	*41.5%*	*40.0%*	*40.0%*	*40.0%*
Others	**0**	**5**	**3**	**10.4**	**10**	**10**	**10**
YoY%		*1127.3%*	*-34.5%*	*213.0%*	*0%*	*0%*	*0%*
GPM%	*51.3%*	*77.0%*	*90.9%*	*69.9%*	*70.0%*	*70.0%*	*70.0%*
Total	**112**	**143**	**265**	**558**	**1,281**	**2,648**	**4,157**
YoY%		*27.3%*	*85.2%*	*110.5%*	*129.6%*	*106.7%*	*57.0%*
GPM%	***59.4%***	***49.5%***	***49.1%***	***54.0%***	***51.6%***	***50.6%***	***51.5%***

Source: Company data, CMBIGM estimates

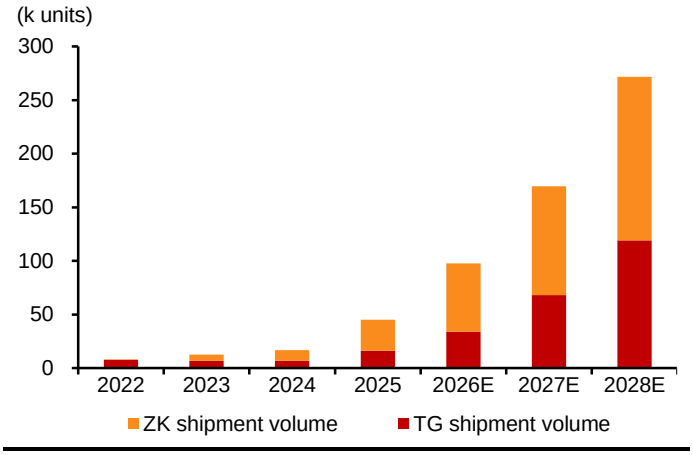
We expect the **TG series (training GPUs)** to emerge as a primary revenue driver, benefiting from **premium ASPs** that reflect its superior compute performance. Conversely, we anticipate the **ZK series** will drive **volume-led growth**, capturing an increasing share of the market as demand pivotally shifts toward inference-centric workloads.

Figure 33: GPGPU revenue breakdown by series



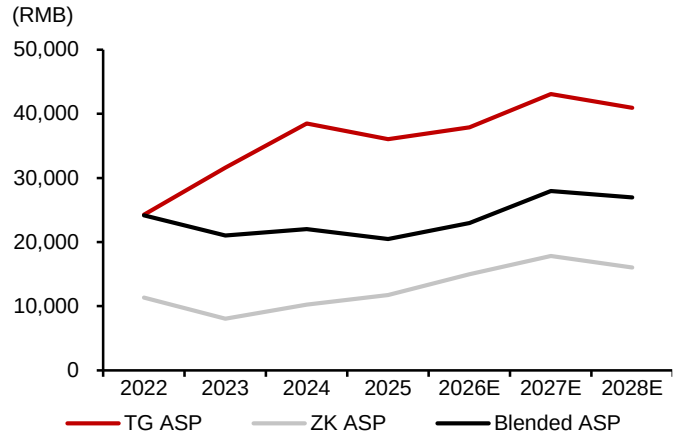
Source: Company data, CMBIGM estimates

Figure 34: GPGPU shipment breakdown by series



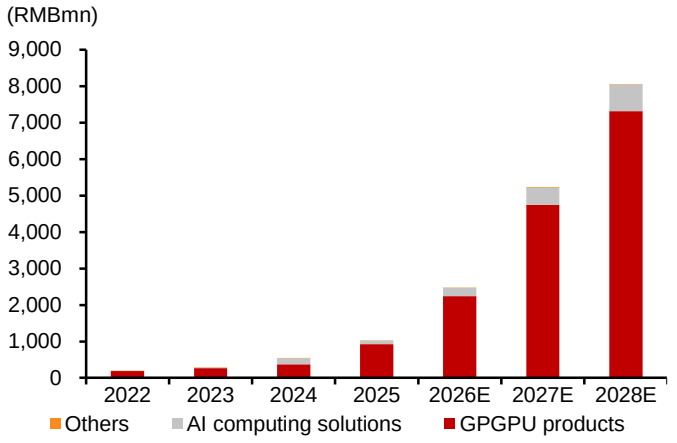
Source: Company data, CMBIGM estimates

Figure 35: ASP trend by series



Source: Company data, CMBIGM estimates

Figure 36: Revenue breakdown by segment



Source: Company data, CMBIGM estimates

Valuation and risks

We adopt a **peer-based valuation framework** to assess the Company's value. Given that domestic AI chip designers are largely in the early stages of product commercialization, we believe **price-to-sales** is the most appropriate metric to capture growth potential.

Applying the peer group's **average 2027E P/S multiple of 26.7x to the Company**, we derive a 2027E equity value of **HK$158.8bn (RMB139.7bn)**, implying a **TP of HK$694** (using a CNY/HKD exchange rate of 0.88). We view this valuation as well-supported, given the Company's rapid revenue scaling and the secular tailwind of **localized AI infrastructure procurement** in China.

Figure 37: Peers table and valuation

		Mkt Cap	P/E (x)		P/S (x)		GPM%	
Company	**Ticker**	**RMB/US$mn**	**FY26E**	**FY27E**	**FY26E**	**FY27E**	**FY26E**	**FY27E**
Overseas and domestic AI chip players								
Cambricon	688256 CH	552,956	110.0	61.4	39.4	23.1	54.8	54.9
Hygon	688041 CH	583,432	124.6	85.5	26.3	19.0	60.6	60.6
MetaX	688802 CH	275,209	4,168.8	272.5	74.6	47.1	54.3	55.1
Moore Threads	688795 CH	293,481	N/A	967.4	109.6	62.6	63.5	61.6
Biren Tech	6082 HK	99,673	N/A	436.1	43.3	18.3	51.2	52.7
Nvidia	NVDA US	4,910,058	25.2	18.4	13.7	10.1	74.4	74.3
AMD	AMD US	448,262	40.7	24.6	9.5	6.6	55.0	55.0
		Average	***893.9***	***266.5***	***45.2***	***26.7***	***59.1***	***59.2***
Iluvatar CoreX's valuation								
2027E revenue		**RMB5,233mn**						
2027E P/S (x)		*26.7x*						
Equity value		**RMB139.7bn**						
CNY/HKD		*0.88*						
Equity value		**HK$158.8bn**						
# Shares (mn)		*228.9*						
Target price		**HK$694**						

Source: Bloomberg and Wind consensus as of 20 April market close, CMBIGM estimates

Risks to our rating and TP

Potential risks to our rating and TP include: 1) supply chain disruption from foundries, 2) deteriorating demand for domestic GPGPU products, 3) intensified competition from both domestic and overseas players, etc.

Financial Summary

INCOME STATEMENT	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec (RMB mn)						
Revenue	**289**	**540**	**1,034**	**2,482**	**5,233**	**8,067**
Cost of goods sold	(146)	(274)	(476)	(1,201)	(2,585)	(3,910)
Gross profit	**143**	**265**	**558**	**1,281**	**2,648**	**4,157**
Operating expenses	**(926)**	**(1,107)**	**(1,511)**	**(1,901)**	**(2,430)**	**(2,870)**
Selling expense	(88)	(122)	(152)	(265)	(453)	(578)
Admin expense	(242)	(257)	(482)	(621)	(785)	(807)
R&D expense	(616)	(773)	(974)	(1,117)	(1,308)	(1,613)
Others	20	45	97	101	116	128
Operating profit	**(783)**	**(842)**	**(953)**	**(620)**	**217**	**1,288**
Other income	(35)	(50)	(51)	(51)	(58)	(63)
EBIT	**(817)**	**(892)**	**(1,004)**	**(671)**	**160**	**1,225**
Income tax	0	0	0	0	0	(184)
After tax profit	**(817)**	**(892)**	**(1,004)**	**(671)**	**160**	**1,041**
Minority interest	(26)	0	0	0	0	0
Net profit	**(791)**	**(892)**	**(1,004)**	**(671)**	**160**	**1,041**

BALANCE SHEET	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec (RMB mn)						
Current assets	**1,287**	**1,262**	**3,432**	**5,688**	**8,883**	**12,496**
Cash & equivalents	308	314	1,505	2,526	3,759	6,217
Restricted cash	0	0	0	0	0	0
Account receivables	200	377	577	1,328	2,113	2,307
Inventories	233	343	710	936	1,614	2,028
Prepayment	339	203	630	796	1,237	1,723
Financial assets at FVTPL	0	0	0	0	0	0
Other current assets	207	26	11	102	159	221
Non-current assets	**306**	**423**	**480**	**923**	**1,397**	**1,818**
PP&E	106	128	188	330	560	838
Right-of-use assets	9	41	10	107	215	334
Intangibles	75	141	190	346	447	431
Financial assets at FVTPL	91	97	76	76	76	76
Other non-current assets	26	17	16	64	99	138
Total assets	**1,593**	**1,685**	**3,912**	**6,611**	**10,280**	**14,314**
Current liabilities	**686**	**880**	**1,112**	**1,767**	**2,582**	**3,085**
Short-term borrowings	492	566	644	650	650	650
Account payables	18	46	31	61	137	162
Other current liabilities	159	222	304	873	1,510	1,876
Lease liabilities	4	18	7	71	110	153
Contract liabilities	14	29	127	113	175	244
Non-current liabilities	**28**	**117**	**446**	**532**	**596**	**667**
Long-term borrowings	0	42	365	350	350	350
Other non-current liabilities	28	75	81	182	246	317
Total liabilities	**715**	**997**	**1,559**	**2,299**	**3,178**	**3,752**
Share capital	186	194	229	229	229	229
Other reserves	692	495	2,124	4,083	6,873	10,334
Total shareholders equity	**878**	**689**	**2,353**	**4,312**	**7,102**	**10,562**
Total equity and liabilities	**1,593**	**1,685**	**3,912**	**6,611**	**10,280**	**14,314**

CASH FLOW	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec (RMB mn)						
Operating						
Depreciation & amortization	118	145	173	314	529	801
Change in working capital	(264)	(169)	(798)	(814)	(1,361)	(857)
Others	(561)	(594)	(443)	(102)	744	1,596
Net cash from operations	**(707)**	**(618)**	**(1,068)**	**(603)**	**(89)**	**1,539**
Investing						
Capital expenditure	(71)	(85)	(139)	(369)	(573)	(798)
Others	(83)	(81)	7	(661)	(720)	(716)
Net cash from investing	**(153)**	**(166)**	**(132)**	**(1,030)**	**(1,293)**	**(1,514)**
Financing						
Net borrowings	340	116	401	(9)	0	0
Proceeds from share issues	565	728	639	639	639	639
Others	42	(54)	1,451	1,557	1,511	1,519
Net cash from financing	**946**	**789**	**2,491**	**2,187**	**2,150**	**2,158**
Net change in cash						
Cash at the beginning of the year	219	308	314	1,505	2,526	3,759
Exchange difference	3	0	(6)	0	0	0
Others	86	6	1,197	1,022	1,233	2,457
Cash at the end of the year	**308**	**314**	**1,505**	**2,526**	**3,759**	**6,217**

GROWTH	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec						
Revenue	52.6%	86.7%	91.6%	140.2%	110.8%	54.2%
Gross profit	27.3%	85.2%	110.5%	129.6%	106.7%	57.0%
Operating profit	na	na	na	na	na	492.6%
EBIT	na	na	na	na	na	667.3%
Net profit	na	na	na	na	na	552.2%

PROFITABILITY	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec						
Gross profit margin	49.5%	49.1%	54.0%	51.6%	50.6%	51.5%
Operating margin	(270.9%)	(156.1%)	(92.2%)	(25.0%)	4.2%	16.0%
Return on equity (ROE)	(105.6%)	(113.9%)	(66.0%)	(20.1%)	2.8%	11.8%

GEARING/LIQUIDITY/ACTIVITIES	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec						
Current ratio (x)	1.9	1.4	3.1	3.2	3.4	4.1

VALUATION	2023A	2024A	2025A	2026E	2027E	2028E
YE 31 Dec						
P/E	ns	ns	ns	ns	523.3	80.2

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.