

China's AI Stack

Three companies behind the models, servers and network

- **Understanding the AI stack:** Behind an AI chatbot sits a much larger technology stack spanning models, processors, memory and networking. Z.AI, Montage Technology and Zhongji InnoLight provide exposure to different parts of this stack.
- **What does it take to run AI:** Processors perform the calculations, memory and interconnects deliver data to them, and high-speed networks allow many processors and servers to work together. Adding more computing power alone does not make a system proportionately faster.
- **AI models:** Z.AI. Z.AI develops the General Language Model (GLM) family of models. Its business is increasingly focused on usage-based access through its open platform and APIs, which accounted for 86.5% of revenue in the first half of 2026. However, substantial model-development costs mean the company remains loss-making.
- **Memory and server interconnects:** Montage Technology. Montage develops chips that help processors, memory and other server components communicate. Its products include DDR5 memory interface chips, PCIe Retimers and CXL memory controllers. Demand depends on server and memory upgrades and the adoption of new interconnect standards.
- **Data-centre networking:** Zhongji InnoLight. Zhongji supplies high-speed optical transceivers that connect servers and computing clusters. First-half 2026 revenue rose 182.5% year on year as demand and shipments of higher-speed 800G and 1.6T products continued to ramp up.
- **One theme, different businesses:** The three companies are not direct competitors and do not represent every part of the AI supply chain. Their end markets also differ: Z.AI has historically been more China-focused but is expanding globally, Montage serves the global server and memory ecosystem, and Zhongji derives most of its revenue from overseas markets.
- **Valuations reflect different business models and profitability:** Montage and Zhongji are already profitable infrastructure suppliers, while Z.AI remains a loss-making model developer. Their valuation multiples reflect different growth expectations and profitability profiles.
- **Key risks:** Risks include intense model competition and computing costs for Z.AI, technology qualification and adoption cycles for Montage, and customer infrastructure spending, product transitions and geopolitical exposure for Zhongji.
- **Access through SGX SDRs:** Upcoming Singapore Depositary Receipts provide another route to the underlying Hong Kong-listed companies, traded locally in Singapore dollars. Trading in Singapore dollars does not remove foreign-currency exposure.

Research Analyst

Gerald Wong

gerald@growbeansprout.com

What does it take to run AI?

When we type a question into an AI chatbot, an answer can appear within seconds. Less visible is the computing infrastructure working behind the scenes.

AI models need processors to perform huge numbers of calculations. Those processors need fast access to memory and data. When an AI workload is distributed across hundreds or thousands of processors, high-speed connections are also needed so that they can communicate with one another.

Companies described as part of the “AI theme” can therefore have very different businesses.

Z.AI develops the large AI models themselves. Montage Technology develops chips that help memory, processors and other components communicate inside servers. Zhongji InnoLight supplies high-speed optical technology that connects servers and computing clusters in AI data centres.

The three companies are not direct competitors, nor do they represent every part of the AI supply chain. Although all three are based in China, their end markets differ significantly.

Z.AI’s business has historically been more China-focused but is expanding globally. Montage Technology supplies the global server and memory ecosystem, while Zhongji InnoLight derives most of its revenue from overseas markets.

Looking at them together provides a useful way to understand several important layers behind modern AI: how the stack works, where each company fits and what drives its business.

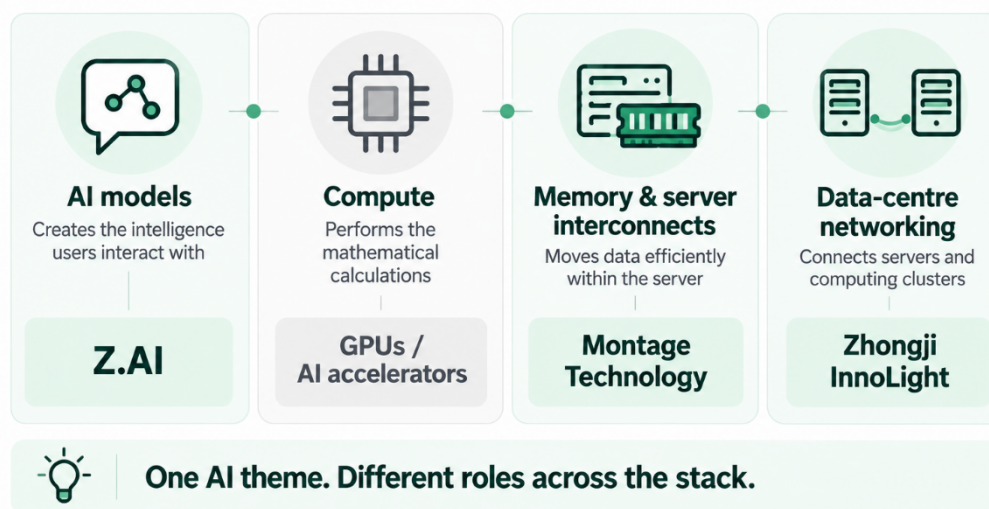
The layers behind modern AI

A useful starting point is to break AI infrastructure into three components: compute, memory and networking. Alongside these sits the model layer, which creates the intelligence users interact with.

Figure 1: Behind every AI answer: the AI stack

Behind every AI answer: the AI stack

Where Z.AI, Montage Technology and Zhongji InnoLight fit



Source: Beansprout

#1 – Compute: doing the calculations

AI models contain large numbers of mathematical parameters that determine how they process information and generate an output.

Z.AI's GLM-5, for example, has 744 billion total parameters, with 40 billion activated at a time.

When a model is trained or used to answer a question, processors such as GPUs perform calculations involving these parameters. GPUs are particularly useful because they can perform many calculations simultaneously.

For increasingly large models, AI workloads can be distributed across many GPUs working together. Meta, for example, trained its 405-billion-parameter Llama 3.1 model using more than 16,000 Nvidia H100 GPUs.

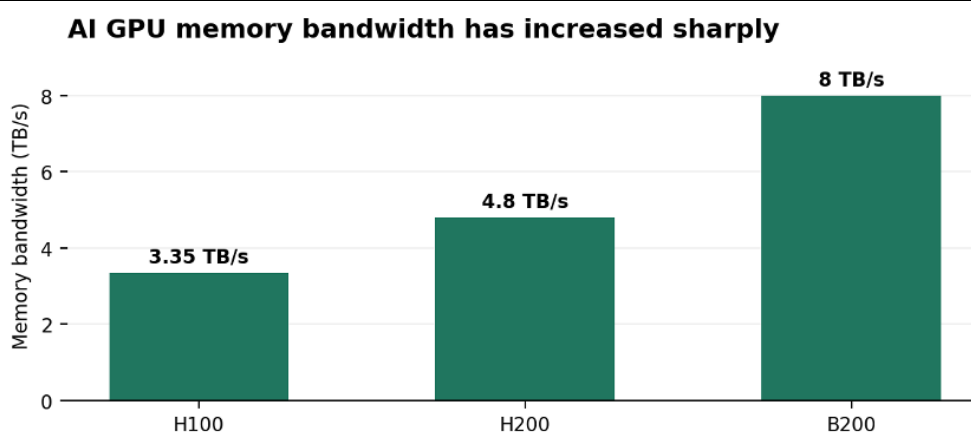
Today's large-scale AI infrastructure consists of clusters of accelerators supported by networking and storage. Google Cloud's clustered GPU architecture, for example, connects thousands of GPUs into a tightly coupled system for large training workloads.

However, adding more GPUs does not make the system proportionately faster. Those GPUs also need data quickly enough to keep them working.

#2 – Memory and interconnects: getting data to the processors

A processor cannot perform a calculation without access to the data it needs. Memory capacity and bandwidth—the speed at which data can be read and moved—are therefore important to AI performance.

The progression in Nvidia's AI GPUs illustrates this. An H100 GPU has 80GB of high-bandwidth memory and 3.35TB per second of memory bandwidth. The H200 increased this to 141GB and 4.8TB per second, while the newer B200 supports 180GB and bandwidth of up to 8TB per second.

Figure 2: AI GPU memory bandwidth

Source: NVIDIA enterprise reference architecture. H100: 80GB HBM3; H200: 141GB HBM3e; B200: 180GB HBM3e

CPUs, GPUs, memory, storage and network cards also need to communicate with one another inside the server.

Technologies such as PCI Express, or PCIe, provide these connections. Newer standards such as Compute Express Link, or CXL, are designed to make it possible to connect and use memory resources more flexibly.

Interconnect chips help data travel quickly and reliably between the components that perform AI calculations.

#3 – Networking: allowing many processors to work together

Once an AI workload extends beyond a single server, the challenge becomes larger. Thousands of accelerators may need to repeatedly exchange information as they work on different parts of a model.

When processors have to wait for data from other machines, they spend less time doing useful calculations. High-performance GPU networks reduce this communication overhead, allowing GPUs to spend more time performing calculations.

At these scales, optical networking becomes increasingly important. Optical transceivers convert electrical signals into light so that large amounts of data can travel quickly over fibre, before converting the signal back at the other end.

Together, these components form a simplified AI stack.

Figure 3: A simplified view of the AI stack

Layer	What it does	Example in this article
AI model	Creates the intelligence users interact with	Z.AI
Compute	Performs the mathematical calculations	GPUs / AI accelerators
Memory & server interconnect	Moves data efficiently within the server	Montage Technology
Data-centre networking	Connects servers and computing clusters	Zhongji InnoLight

Source: Beansprout

Note: This is a simplified representation. AI infrastructure also includes storage, software, power, cooling and other components.

How Singapore investors can access the theme

Z.AI, Montage Technology and Zhongji InnoLight are listed in Hong Kong under stock codes 2513, 6809 and 3308, respectively.

Singapore Depository Receipts, or SDRs, provide another way for Singapore investors to access selected overseas-listed companies through SGX.

An SDR gives its holder a beneficial interest in an underlying foreign-listed security. SDRs trade in Singapore dollars during SGX market hours and are custodied with CDP.

However, trading an SDR in Singapore dollars does not remove foreign-currency exposure. Its price may still be affected by movements in the underlying share, exchange rates, liquidity, and supply and demand.

Figure 4: SDR details

Company	SGX SDR code	SDR ratio / effective date
Z.AI	HZAD	100:1
Montage Technology	HMED	30:1
Zhongji InnoLight	HZJD	100:1

Source: SGX

#1 – AI Models: Z.AI

Developing the intelligence users interact with

Z.AI sits at the model layer. The company develops the General Language Model (GLM) family of general-purpose AI models, which can power applications such as chatbots, coding assistants and AI agents across tasks including coding and reasoning.

Developing a capable model is only one part of the business. The other question is how that model is commercialised.

One route is through an application programming interface, or API. Rather than developing an AI model themselves, developers can send requests to Z.AI's models and pay based on usage.

Revenue is increasingly driven by platform and API usage

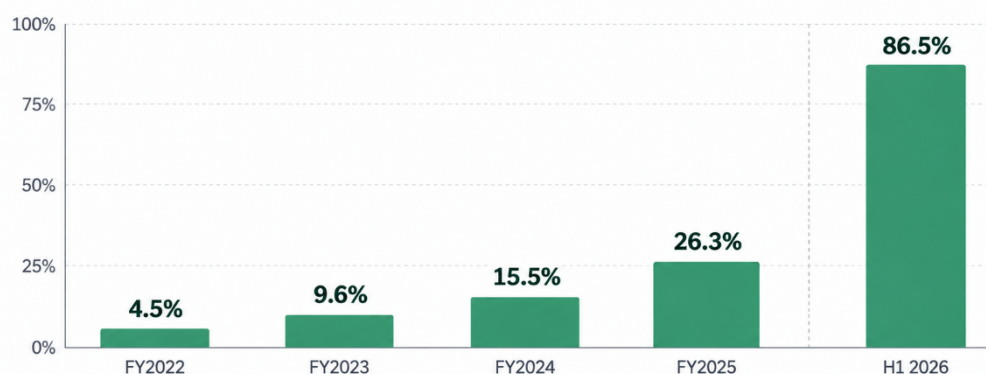
Z.AI has increasingly shifted towards this model.

In the first half of 2026, revenue from its open platform and API reached RMB825 million, accounting for 86.5% of its RMB954 million total revenue, compared with 15.2% a year earlier.

Figure 5: Z.AI's revenue mix shifts towards APIs

Z.AI's revenue mix shifts towards APIs

Cloud-based / open-platform & API revenue as a share of total revenue



Source: Z.AI's IPO Prospectus, FY2025 & 1H26 Results.

Its GLM Model-as-a-Service platform had more than 5.8 million enterprise and developer users at the end of June 2026.

Model development remains costly

Producing that intelligence is expensive. Z.AI spent about RMB2.13 billion on R&D in the first half of 2026 and remained loss-making.

For investors, the relevant question is not only whether its latest model is capable, but whether Z.AI can turn that capability into sustained usage and revenue while managing the cost of developing and running its models.

#2 – Memory & Server Interconnects: Montage Technology

Helping data move inside the AI server

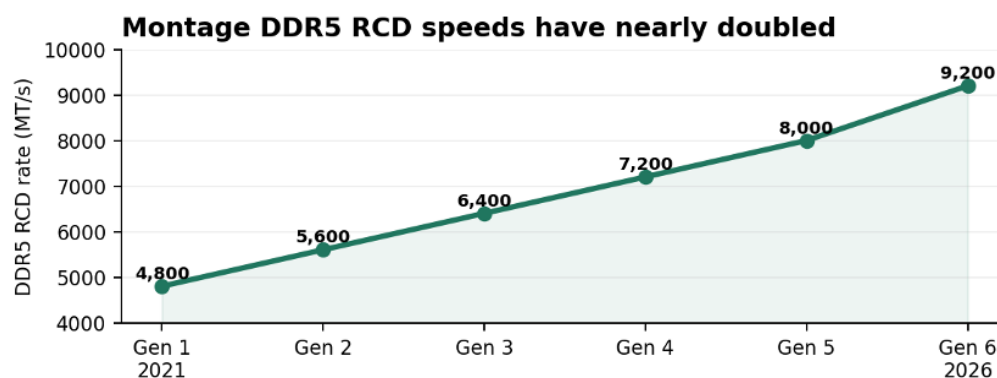
Montage Technology operates further down the infrastructure stack. It is a fabless semiconductor company focused on interconnect chips, which help processors, memory and other components inside servers communicate.

Memory interface chips

As server memory becomes faster, maintaining reliable communication between processors and memory becomes harder. Montage's DDR5 Registering Clock Driver, or RCD, chips help manage these high-speed signals.

The technology has moved quickly. Montage's first DDR5 RCD generation supported speeds of up to 4,800 MT/s in 2021. By June 2026, it had sampled its sixth-generation product supporting 9,200 MT/s.

Figure 6: Montage DDR5 RCD speeds have nearly doubled



Source: Montage Technology 2026 Interim Report.

PCIe Retimers

As high-speed electrical signals travel between GPUs, storage and network devices, they can weaken or become distorted.

A Retimer reconstructs the signal so that data can continue travelling reliably.

These chips can be used in significant numbers. Montage says a typical AI server containing eight GPUs may require eight to 16 PCIe Retimers, while some configurations use as many as 24.

CXL memory controllers

Montage is also developing Compute Express Link (CXL) memory controllers, which can enable memory expansion and pooling.

Instead of being limited to memory directly attached to one processor, future server architectures could use CXL to access memory resources more flexibly.

This technology is still earlier in its adoption cycle. Montage said in its 2026 interim report that CXL memory-pooling solutions were in the industry trial and validation phase among cloud providers and server manufacturers.

Interconnect products account for most of revenue

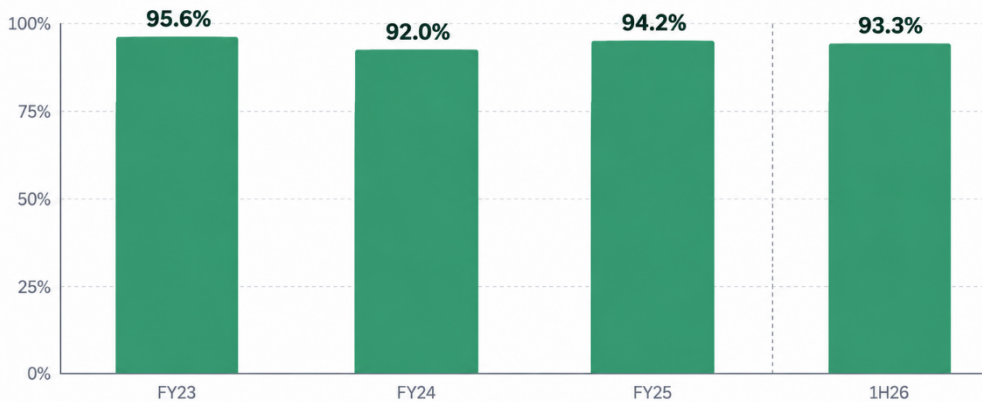
Interconnect products already make up most of Montage's business.

In the first half of 2026, its interconnect-chip product line generated RMB3.11 billion of its RMB3.34 billion total revenue. Revenue from newer products, including MRCD/MDB, PCIe Retimer, CKD and CXL MXC chips, reached RMB538 million, up 80.7% year on year.

Figure 7: Interconnect chips dominate Montage's revenue

Interconnect chips dominate Montage's revenue

Interconnect-chip revenue as a share of total revenue



Source: Montage Technology annual and interim reports.

Montage's role in AI is therefore less about performing the computation and more about ensuring that data can reach the components doing the computation efficiently.

#3 – Data-Centre Networking: Zhongji InnoLight

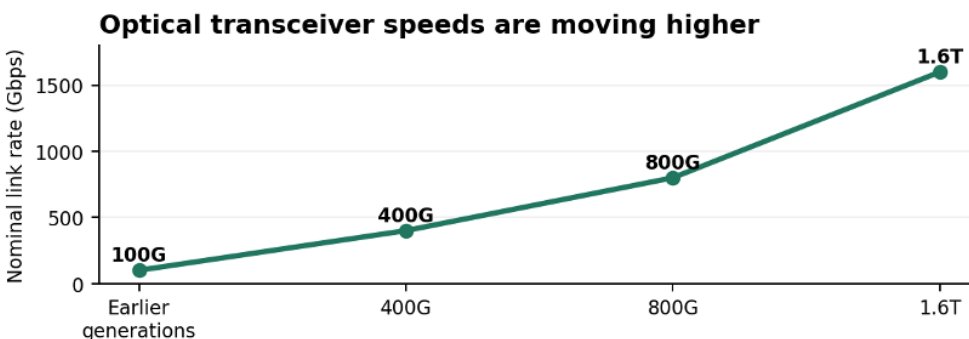
Moving data between AI servers

Montage helps explain what happens inside a server. Zhongji InnoLight addresses another problem: moving data between servers and computing clusters.

The company produces high-speed optical transceivers. These convert electrical signals into optical signals and back again, allowing data to travel over fibre between servers, switches and other network equipment.

Larger AI clusters require faster optical links

As AI clusters become larger, these links have had to become faster. Zhongji's product portfolio spans 100G and 200G through 400G, 800G and 1.6T optical transceivers.

Figure 8: Optical transceiver speeds are moving higher

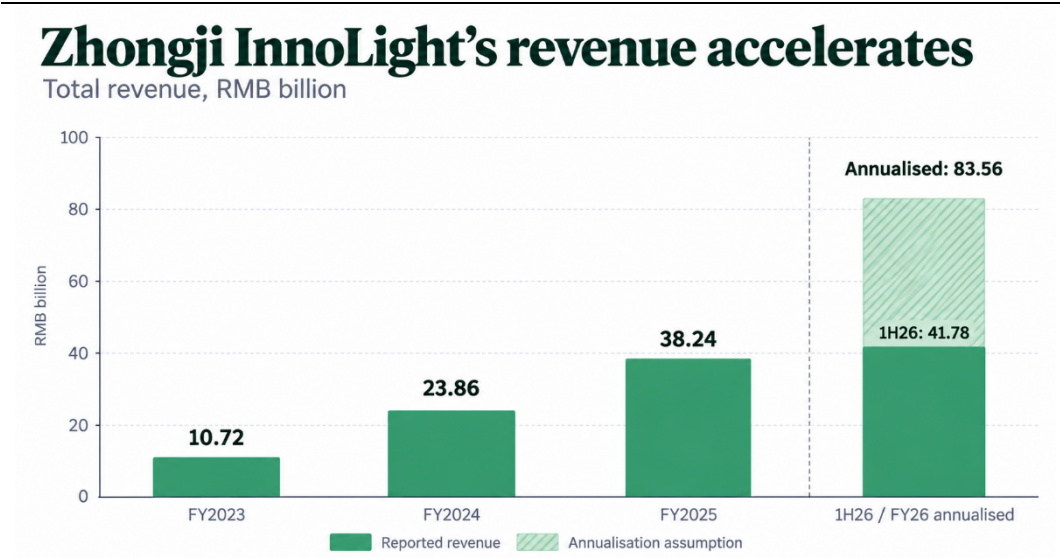
Source: Zhongji InnoLight 2026 Interim Report. 1.6T = 1,600 Gbps.

Higher-speed products support revenue growth

Revenue reached about RMB41.8 billion in the first half of 2026, up 182.5% year on year.

The company said demand and shipments of its higher-speed 800G and 1.6T products continued to ramp up as customers increased investment in AI computing infrastructure.

Figure 9: Zhongji InnoLight’s revenue accelerates



Source: Zhongji InnoLight Annual and Interim reports. Annualised number taken as 1H26 multiplied by 2, not company guidance.

For investors, Zhongji’s business is more closely connected to the expansion of AI computing clusters and upgrades to faster data-centre networks than to which individual AI model becomes most popular.

Same AI theme, different business drivers

The three companies can all be linked to increased use of AI, but through different mechanisms. The same growth in AI usage therefore does not affect each company in the same way.

Figure 10: Comparing the three companies across the AI stack

	Z.AI	Montage Technology	Zhongji InnoLight
Part of the AI stack	AI models	Memory and server interconnect	Data-centre networking
What it provides	AI models and access to them	Chips helping server components exchange data	Optical links connecting servers and clusters
Simple way to think about it	Creates the intelligence	Moves data inside the server	Moves data between servers
What drives the business	Model usage and monetisation	Server and memory upgrades; new interconnect standards	AI cluster expansion and network-speed upgrades

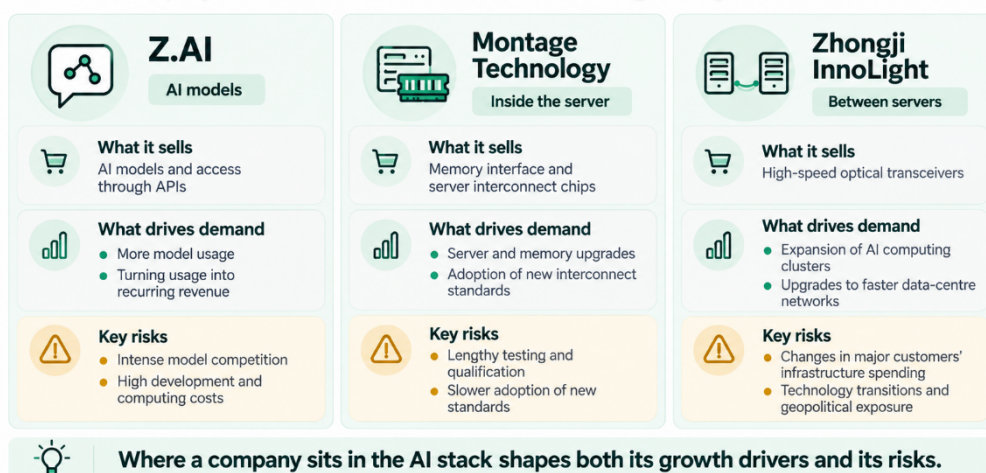
Source: Beansprout

Describing all three simply as “AI stocks” tells investors relatively little. A more useful question is where each company sits in the stack and what actually drives demand for that part.

Figure 11: Same AI theme, three different business engines

Same AI theme, three different business engines

What each company sells, what drives demand, and what could go wrong



Source: Beansprout

Valuations reflect different business models and profitability

Figure 12: Peer comparison

Company Name	Ticker	Currency	Price	Market Value (US\$b)	P/E	NTM (x)	P/BV (x)	Dividend Yield (%)	ROE (%)
Z.AI Co., Ltd. -H- Reg S	2513-HK	HKD	696.00	23.3	n.a.	n.a.	n.a.	n.a.	n.a.
Montage Technology Co., Ltd. Class H	6809-HK	HKD	291.00	37.3	78.2	n.a.	n.a.	n.a.	18.3
Zhongji InnoLight Co., Ltd. Class A	300308-CN	CNY	808.44	142.3	26.5	22.8	0.2	44.2	
Semiconductor Manufacturing International Corp.	981-HK	HKD	60.90	89.2	44.0	3.4	-	3.3	
Alibaba Group Holding Limited	9988-HK	HKD	105.40	267.3	16.7	1.8	0.9	10.2	
Tencent Holdings Ltd	700-HK	HKD	420.60	487.3	12.4	4.2	0.9	21.0	
Baidu, Inc. Class A	9888-HK	HKD	84.75	23.9	14.9	1.2	-	1.8	
Average					32.1x	6.7x	0.4%	16.5	
Median					21.6x	3.4x	0.2%	14.2	

Source: Factset, data as of 8 October 2026. Average based on comparison-set average. Zhongji InnoLight valuation data shown refers to its Shenzhen-listed A shares (300308-CN).

The three companies are at different stages of profitability, which is reflected in their valuation metrics.

Montage Technology traded at 78.2x forward P/E, the highest among the companies in the comparison and well above the peer average of 32.1x and median of 21.6x. Its return on equity (ROE) stood at 18.3%.

Zhongji InnoLight traded at 26.5x forward P/E, above the peer median but below the average. Its ROE was 44.2%, the highest in the comparison, while its price-to-book ratio of 22.8x was also significantly above the peer average of 6.7x.

Z.AI does not currently have a meaningful forward P/E or ROE, as the company remains loss-making. It is still investing heavily in model development and computing infrastructure as it seeks to grow API and platform usage.

The comparison highlights why the three companies should not be viewed as direct peers. Z.AI is an earlier-stage AI model developer, while Montage Technology and Zhongji InnoLight are already profitable infrastructure suppliers.

Their valuation multiples reflect different business models, growth expectations and profitability profiles.

Key risks

Z.AI: model competition and development costs

Model competition is intense, and developing and running frontier models requires significant R&D and computing expenditure. Model capability therefore needs to translate into recurring usage and revenue.

Montage Technology: evolving standards and adoption cycles

Technology standards continue to evolve. Products built around DDR, PCIe and CXL must be developed, tested and qualified before they are adopted at scale.

A new standard being technologically promising does not guarantee rapid commercial adoption.

Zhongji InnoLight: infrastructure spending and product transitions

Demand is closely tied to spending by large cloud and AI infrastructure customers. Rapid transitions from 400G to 800G and 1.6T also require manufacturers to keep developing and ramping successive generations of technology.

Zhongji InnoLight: geopolitical and regulatory exposure

Its global customer exposure introduces geopolitical and regulatory risks.

In June 2026, Zhongji disclosed that it had been included on the US Section 1260H list, while saying at the time that its orders, production and supply chain remained normal.

Stronger AI adoption does not automatically produce the same outcome for every company linked to the theme. Understanding where a company sits in the AI stack also helps investors understand what could go wrong.

Disclosure Appendix

Analyst Certification and Disclosures

The analyst(s) named in this report certifies that (i) all views expressed in this report accurately reflect the personal views of the analyst(s) with regard to any and all of the subject securities and companies mentioned in this report and (ii) no part of the compensation of the analyst(s) was, is, or will be, directly or indirectly, related to the specific views expressed by that analyst herein. The analyst(s) named in this report (or their associates) does not have a financial interest in the corporation(s) mentioned in this report.

An associate is defined as (i) the spouse, or any minor child (natural or adopted) or minor step-child, of the analyst; (ii) the trustee of a trust of which the analyst, his spouse, minor child (natural or adopted) or minor step-child, is a beneficiary or discretionary object; or (iii) another person accustomed or obliged to act in accordance with the directions or instructions of the analyst.

Company Disclosure

Global Wealth Technology Pte Ltd ("Beansprout") does not have any financial interest in the corporation(s) mentioned in this report.

Beansprout received monetary compensation from Singapore Exchange (SGX) to provide independent research on Singapore Depository Receipts (SDRs).

Disclaimer

This report is provided by Beansprout for the use of intended recipients only and may not be reproduced, in whole or in part, or delivered or transmitted to any other person without our prior written consent. By accepting this report, the recipient agrees to be bound by the terms and limitations set out herein.

You acknowledge that this document is provided for general information purposes only. Nothing in this document shall be construed as a recommendation to purchase, sell, or hold any security or other investment, or to pursue any investment style or strategy. Nothing in this document shall be construed as advice that purports to be tailored to your needs or the needs of any person or company receiving the advice. The information in this document is intended for general circulation only and does not constitute investment advice. Nothing in this document is published with regard to the specific investment objectives, financial situation and particular needs of any person who may receive the information.

Nothing in this document shall be construed as, or form part of, any offer for sale or subscription of or solicitation or invitation of any offer to buy or subscribe for any securities. The data and information made available in this document are of a general nature and do not purport, and shall not in any way be deemed, to constitute an offer or provision of any professional or expert advice, including without limitation any financial, investment, legal, accounting or tax advice, and shall not be relied upon by you in that regard. You should at all times consult a qualified expert or professional adviser to obtain advice and independent verification of the information and data contained herein before acting on it. Any financial or investment information in this document are intended to be for your general information only. You should not rely upon such information in making any particular investment or other decision which should only be made after consulting with a fully qualified financial adviser. Such information do not nor are they intended to constitute any form of financial or investment advice, opinion or recommendation about any investment product, or any inducement or invitation relating to any of the products listed or referred to. Any arrangement made between you and a third party named on or linked to from these pages is at your sole risk and responsibility.

You acknowledge that Beansprout is under no obligation to exercise editorial control over, and to review, edit or amend any data, information, materials or contents of any content in this document. You agree that all statements, offers, information, opinions, materials, content in this document should be used, accepted and relied upon only with care and discretion and at your own risk, and Beansprout shall not be responsible for any loss, damage or liability incurred by you arising from such use or reliance.

This document (including all information and materials contained in this document) is provided “as is”. Although the material in this document is based upon information that Beansprout considers reliable and endeavours to keep current, Beansprout does not assure that this material is accurate, current or complete and is not providing any warranties or representations regarding the material contained in this document. All opinions contained herein constitute the views of the analyst(s) named in this report, they are subject to change without notice and are not intended to provide the sole basis of any evaluation of the subject securities and companies mentioned in this report. Any reference to past performance should not be taken as an indication of future performance. To the fullest extent permissible pursuant to applicable law, Beansprout disclaims all warranties and/or representations of any kind with regard to this document, including but not limited to any implied warranties of merchantability, non-infringement of third-party rights, or fitness for a particular purpose.

Beansprout does not warrant, either expressly or impliedly, the accuracy or completeness of the information, text, graphics, links or other items contained in this document. Neither Beansprout nor any of its affiliates, directors, employees or other representatives will be liable for any damages, losses or liabilities of any kind arising out of or in connection with the use of this document. To the best of Beansprout’s knowledge, this document does not contain and is not based on any non-public, material information. The information in this document is not intended for distribution to, or use by, any person or entity in any jurisdiction where such distribution or use would be contrary to law or regulation, or which would subject Beansprout to any registration requirement within such jurisdiction or country. Beansprout is not licensed or regulated by any authority in any jurisdiction or country to provide the information in this document.

As a condition of your use of this document, you agree to indemnify, defend and hold harmless Beansprout and its affiliates, and their respective officers, directors, employees, members, managing members, managers, agents, representatives, successors and assigns from and against any and all actions, causes of action, claims, charges, cost, demands, expenses and damages (including attorneys’ fees and expenses), losses and liabilities or other expenses of any kind that arise directly or indirectly out of or from, arising out of or in connection with violation of these terms, use of this document, violation of the rights of any third party, acts, omissions or negligence of third parties, their directors, employees or agents. To the extent permitted by law, Beansprout shall not be liable to you, any other person, or organization, for any direct, indirect, special, punitive, exemplary, incidental or consequential damages, whether in contract, tort (including negligence), or otherwise, arising in any way from, or in connection with, the use of this document and/or its content. This includes, without limitation, liability for any act or omission in reliance on the information in this document. Beansprout expressly disclaims and excludes all warranties, conditions, representations and terms not expressly set out in this User Agreement, whether express, implied or statutory, with regard to this document and its content, including any implied warranties or representations about the accuracy or completeness of this document and the content, suitability and general availability, or whether it is free from error.

If these terms or any part of them is understood to be illegal, invalid or otherwise unenforceable under the laws of any state or country in which these terms are intended to be effective, then to the extent that they are illegal, invalid or unenforceable, they shall in that state or country be treated as severed and deleted from these terms and the remaining terms shall survive and remain fully intact and in effect and will continue to be binding and enforceable in that state or country.

These terms, as well as any claims arising from or related thereto, are governed by the laws of Singapore without reference to the principles of conflicts of laws thereof. You agree to submit to the personal and exclusive jurisdiction of the courts of Singapore with respect to all disputes arising out of or related to this Agreement. Beansprout and you each hereby irrevocably consent to the jurisdiction of such courts, and each Party hereby waives any claim or defence that such forum is not convenient or proper.

© 2025 Global Wealth Technology Pte Ltd