

ASIA TECHNOLOGY

Understanding the AI Value Chain: Investment opportunities in AI Enablers & Adopters

White Paper ♦ October 2025



Understanding the AI Value Chain: Investment Opportunities in AI Enablers & Adopters

White Paper • October 2025

Rapid technological developments have been a feature of the recent few decades: home computing, the internet, mobile voice, and then mobile data. This has led to a significant transformation in how we as humans live, work, interact socially, and, indeed, how corporations are run and managed. Processing power is now available at an acceptable price point and power consumption level. This, combined with some significant innovations in computing algorithms and the availability of capital, we find ourselves amidst another major wave of technological transformation – the one led by the application of Artificial Intelligence (AI) tools across different aspects of business and, indeed, day-to-day life.

In some ways, AI is not a completely new and shiny ‘thing’. The basics behind the concept have been around for longer than the personal computer – it is just that, as we mentioned earlier, several factors have now come together to make the current time ripe for the rapid expansion and implementation of the concept. Just as the advent of mobile data in the mid-2000s continues to drive innovation and impact businesses almost 20 years later, the ‘age’ of AI will extend for multiple decades into the future. We are currently in step one of that era – the step where one first builds the infrastructure.

Part 1: AI infrastructure build-out

AI is exciting, but it is incredibly capital hungry – which is one of the reasons why it has taken this long for AI themes to step out of the ‘lab’ and into the real world. Training AI models and running queries on them requires immense computing power as well as vast amounts of data, both of which are resource-hungry, necessitating the build-out of data centers suitable to meet those needs. Luckily for the industry, the rapid progress in deep learning and innovation in AI accelerators and graphics processing unit (GPU) technologies occurred at a time when companies already had existing expertise in cloud data centers and the balance sheet & cash flows to undertake the next wave of capital expenditures (capex).

Training compute requirements continue to grow

We believe that the pace of development in this space is likely to continue at a rapid clip for the coming few years before settling down. The size of the foundational and training models, and the data/tokens they are consuming, has seen a step change since 2010 or so (the start of the deep learning era), with training compute (FLOPs¹) growing at ~4.3x per year after having grown at ~1.5x for the previous several decades.² If one were to focus only on the top 30 models in this ‘deep learning era’, the growth rate increases to ~4.9 times per year.³ Evidence coming

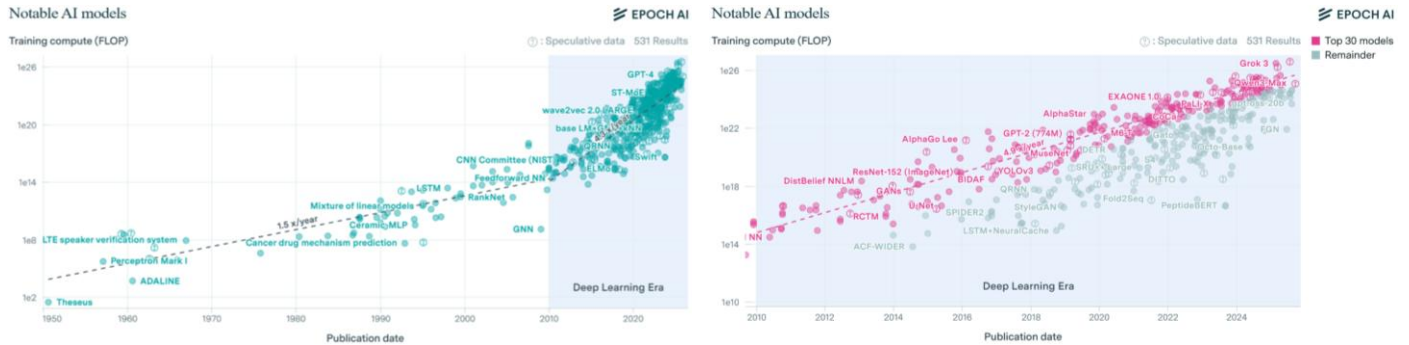
¹ FLOP (Floating Point Operation) measures computational intensity by counting the number of mathematical operations performed. In AI training, it quantifies the computational work required – higher FLOPs typically indicate more complex calculations and powerful models.

² Source: Epoch AI, September 2025

³ Source: Ibid.

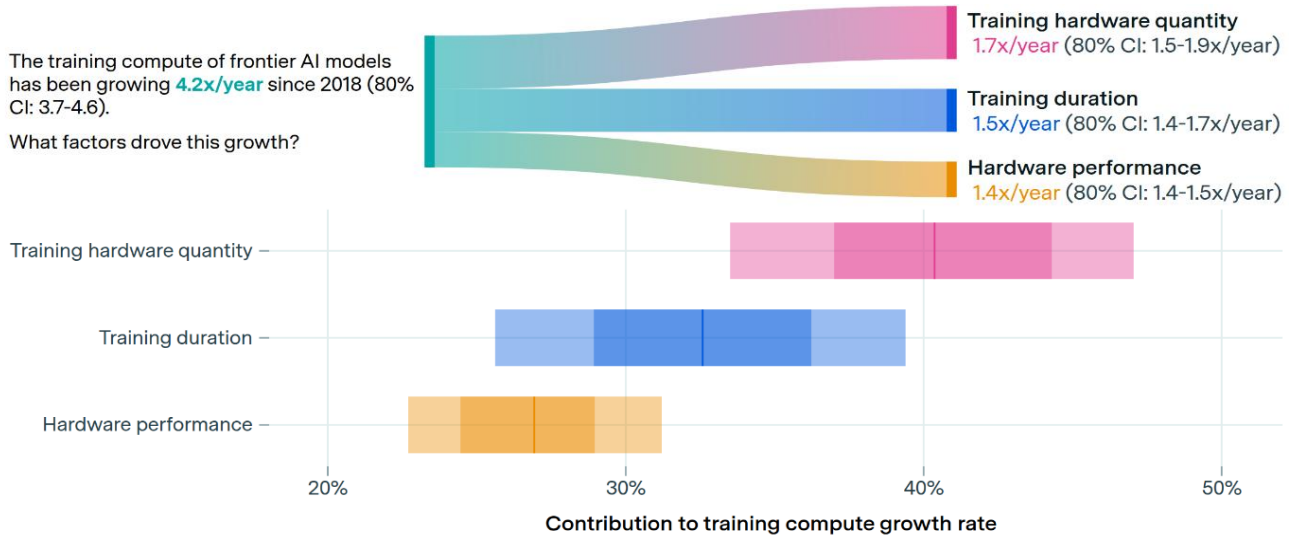
from most of the leading players in this area supports the view that the pace of growth for training compute requirements is likely to sustain itself over the coming few years.

Exhibit 1: Growth in training compute has accelerated in the ‘deep learning era’ (left); For the top 30 models, training compute has expanded by almost 5x per year over the last 15 years (right)



Source: Epoch AI, ‘Data on AI Models’, accessed 16 September 2025

Exhibit 2: Training hardware quantity has been the biggest driver of growth in training compute



Source: Epoch AI, September 2025

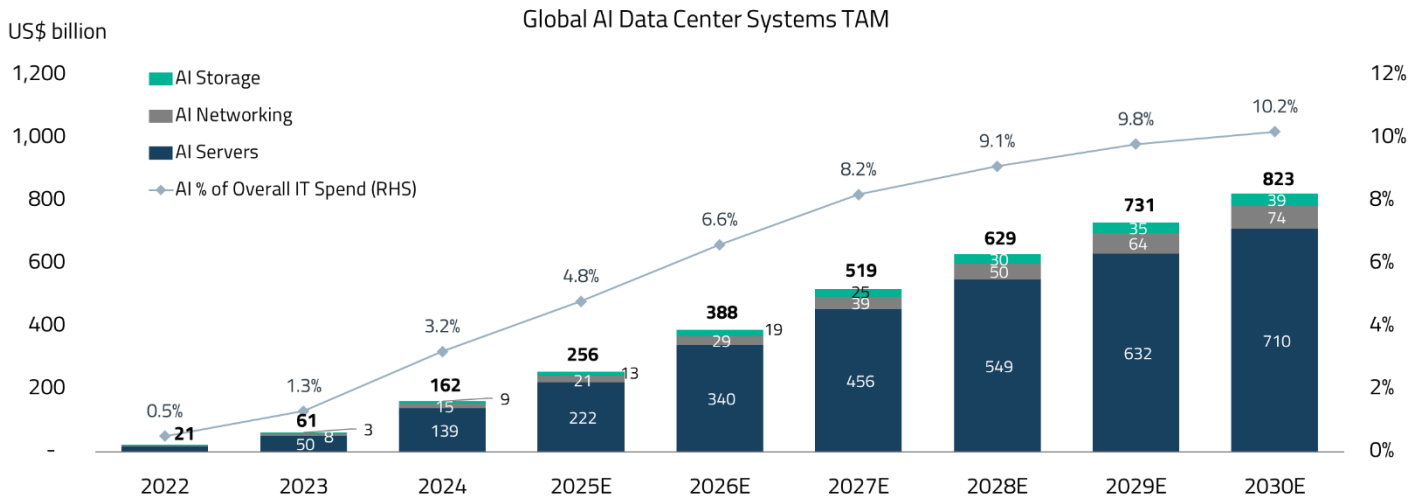
Higher compute requirements are driving dramatic capex growth

The growth in hardware compute requirements for AI infrastructure is a primary driver of global IT spending. That spend has been turbocharged since the release of ChatGPT at the end of 2022, with just the share of AI data center spend as a percent of total IT spending expected to rise from about 0.5% in 2022 to over 10% by 2030 – higher, if one were to take some other forecasts, as we discuss later.⁴

⁴ Source: BofA Global Research estimates, June 2025

This heavy spend on computing hardware needs to be, and is being, accompanied by investments in related infrastructure around storage, networking, and connectivity. And, of course, to house all this tech equipment, there is also an investment in physical infrastructure, including data centers and related electricity generation.

Exhibit 3: AI data center systems TAM expected to reach ~USD 820 billion by 2030 (+26% CAGR between 2025 and 2030)



Source: BofA Global Research estimates, June 2025

As things stand today and according to prevailing wisdom, despite the dramatic increase in AI-related computing infrastructure over the past three years, more is still considered better in the search for improving the foundational model’s outputs. As a result, capex numbers have been significantly revised up, and on a regular basis. In August 2024, the estimated capex for 2025 from the leading 11 cloud infrastructure buyers was USD 288 billion.⁵ By early 2025, it increased to USD 320 billion, and as of the end of August 2025, it stood at USD 438 billion.⁶ Our view is that the actual number for 2025 will end up being even higher, constrained only by the supply of the relevant chips and the ability to manufacture and deploy the final hardware/racks.

Exhibit 4: Consensus CY25 capex estimates for top 11 hyperscalers have jumped >50% over the past 12 months

US\$ billion	CY25 Capex Estimate			Change from	
	A year ago	Beginning CY25	At end-Aug 25 (Post 2Q25)	A year ago	Beginning CY25
Top 4	228	257	345	117 <i>+51%</i>	88 <i>+34%</i>
Next 7	60	63	93	33 <i>+56%</i>	30 <i>+48%</i>
Top 11 hyperscalers	288	320	438	150 <i>+52%</i>	118 <i>+37%</i>

Source: Company data, FactSet, Morgan Stanley Research, August 2025

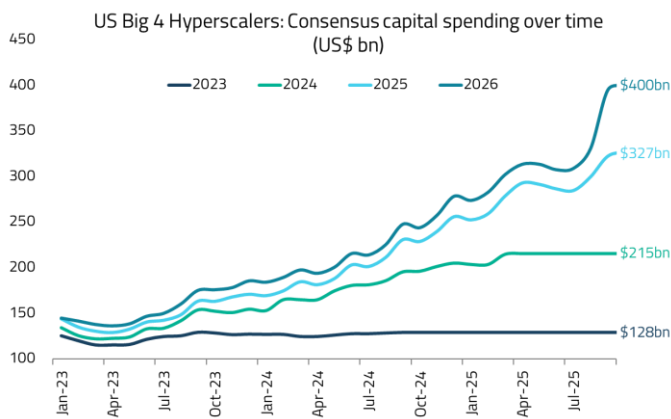
⁵ Source: Company data, FactSet, Morgan Stanley Research, August 2025

⁶ Source: Ibid.

We expect the 2026 capex estimate, which has been revised upward sharply in recent months, to continue rising. Firstly, note the sharper upgrade to capex from even the top 4 hyperscalers (Exhibit 5) following the June-quarter results – a function of increased visibility from the companies regarding their capex intentions for 2026, and, significantly, consistently stronger qualitative commentary around workload requirements and the pace of data center expansion plans. We expect 2026 capex estimates to move further upwards through the September-quarter and the December-quarter results as these hyperscalers provide more concrete plans for next year.

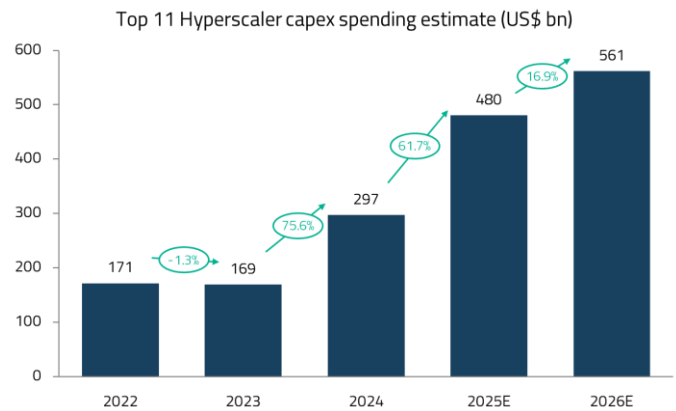
Secondly, Stargate – the OpenAI/Oracle/SoftBank project – wasn’t explicitly in the street’s numbers for spending earlier, but is likely to be a meaningful spender in 2026. Those numbers need to be added, as do several other plans initiated by OpenAI (which by themselves imply almost USD 1 trillion of capacity build over the next five years).⁷

Exhibit 5: US Big 4 hyperscalers are ramping up capital spending



Source: Company reports, UBS, VisibleAlpha, August 2025

Exhibit 6: The trend is global, with the top 11 hyperscalers globally also ramping up spends



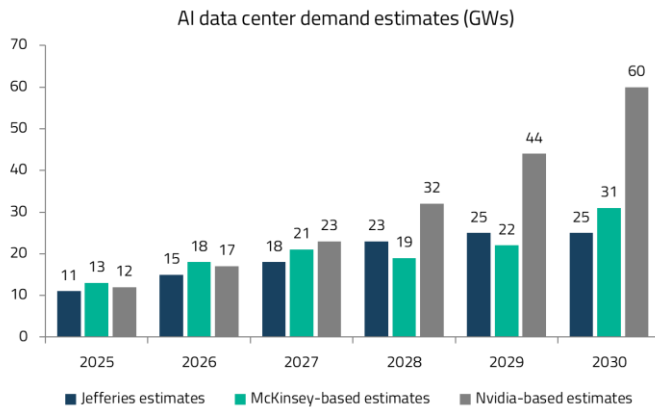
Source: Company data, Bloomberg consensus estimates, UBS estimates, September 2025

Indeed, if one goes by what Jensen Huang (CEO of Nvidia) has been saying – that AI-related infrastructure spending in 2030 will be around USD 3-4 trillion – then that will translate into approximately 60GW of compute power, roughly twice the size of capacity underlying the assumptions made in Exhibit 3 above, and more in line with OpenAI’s recently announced plans.⁸ If true, that would imply that instead of a 26% CAGR over 2025-30 for AI spending capex, the annual growth rate could be as high as 42%!

⁷ Source: Shikhara Investment Management analysis, September 2025

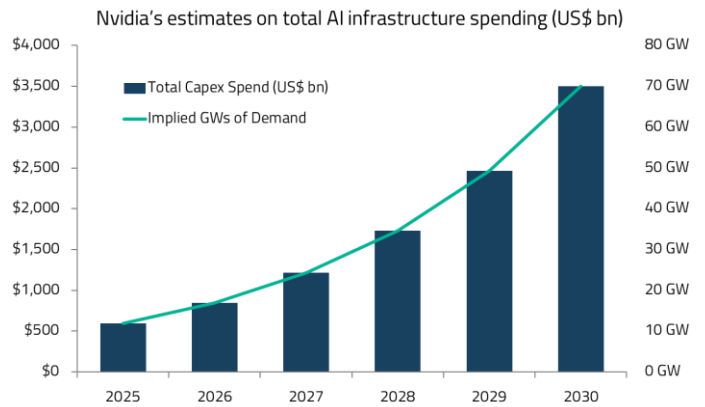
⁸ Source: Jefferies, September 2025

Exhibit 7: AI data center demand – Nvidia’s scenario tops estimates by industry specialists



Source: Jefferies assumptions, company reports, McKinsey estimates, September 2025

Exhibit 8: Nvidia projects USD 3-4 trillion in AI infrastructure spend by 2030



Source: NVDA earnings call, Jefferies, September 2025

While the current wave of spending for the infrastructure buildout is coming from the big tech hyperscalers / existing cloud service providers, the list of entities wanting and willing to spend money to build AI infrastructure is growing. The constraining factor, at this point in time, is the availability of chips and hardware. Prominent amongst this list of new ‘entities’ are the sovereign ones backed by national ambitions to build and secure a local AI infrastructure.

Part 2: AI-accelerator roadmap remains aggressive

While there has been tremendous growth in per-GPU compute power in recent years, that pace of growth is expected to be maintained in the coming years, with the industry having a strong, visible roadmap ahead. Nvidia has laid out an aggressive annual cadence for its range of GPUs, with major upgrades every two years. In addition to delivering GPU computing power, Nvidia has also innovated at the system level by creating clusters of its GPUs within the same rack, while also advancing connectivity/networking technology.

Nvidia’s roadmap through 2028 confirms this cadence. The Blackwell Ultra (2025) is expected to be ~1.5 times faster than current Blackwell chips, while the 2026 Rubin architecture is anticipated to incorporate advanced 3nm manufacturing and next-generation HBM4 memory. At the data center level, Nvidia’s new rack systems will pack 576 GPUs, delivering massive computing power, with annual upgrades planned through 2028.

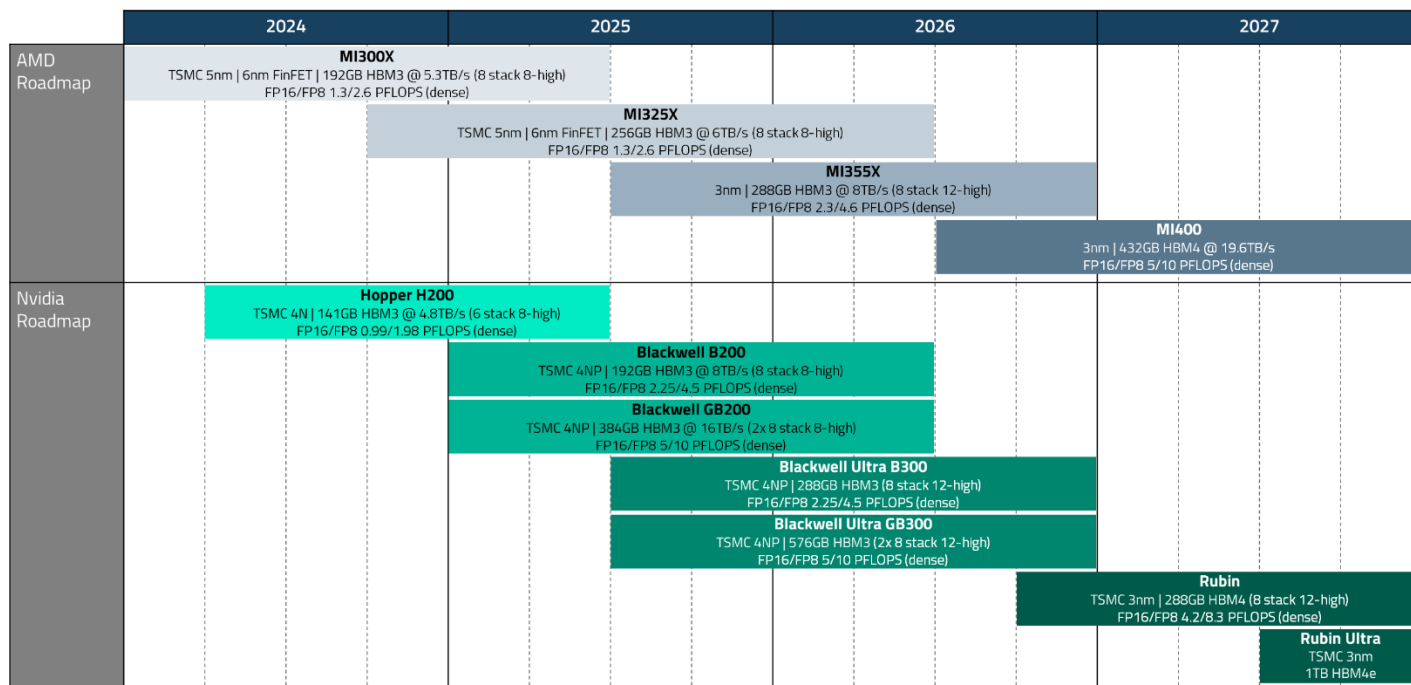
Exhibit 9: Nvidia GPU roadmap

Architecture	CPU	GPU	System	Year	Notes
Blackwell	Grace	Blackwell (GB200)	NVL72	2024-2025	Current generation, with Blackwell Ultra (B300) planned for late 2025.
Rubin	Vera	Rubin (R100)	NVL144	2026	Expected to launch in the second half of 2026. It will feature HBM4 memory and significant performance gains over Blackwell.
Rubin Ultra	Vera	Rubin Ultra	Kyber (NVL576)	2027	An upgraded version of the Rubin architecture. A full Kyber rack could deliver 14 times the inference performance of the current GB200 NVL72 system.
Feynman	Vera	Feynman	TBA	2028	The architecture following Rubin, expected in 2028 with next-generation HBM memory.

Source: Company data, Shikhara Investment Management analysis, September 2025

While AMD is currently a distant second in terms of market share in the merchant AI GPU market, its product offerings and roadmap remain quite strong as well. As the chart below shows, AMD is set to launch its MI355X suite of products on TSMC’s 3nm process, which, on specs, compares quite favorably with Nvidia’s Blackwell Ultra B300 chip. The difference in end market share comes from Nvidia’s strong existing share, its entrenched CUDA ecosystem, and its innovative rack and networking design. However, the strong performance benchmark specs mean that AMD will remain in the game, pushing the innovative frontier further for the industry. AMD’s GPU roadmap received a major nod of approval through the recent announcement of its multi-billion-dollar deal to provide OpenAI with its MI400 range starting late 2026.

Exhibit 10: AMD and Nvidia GPU roadmaps



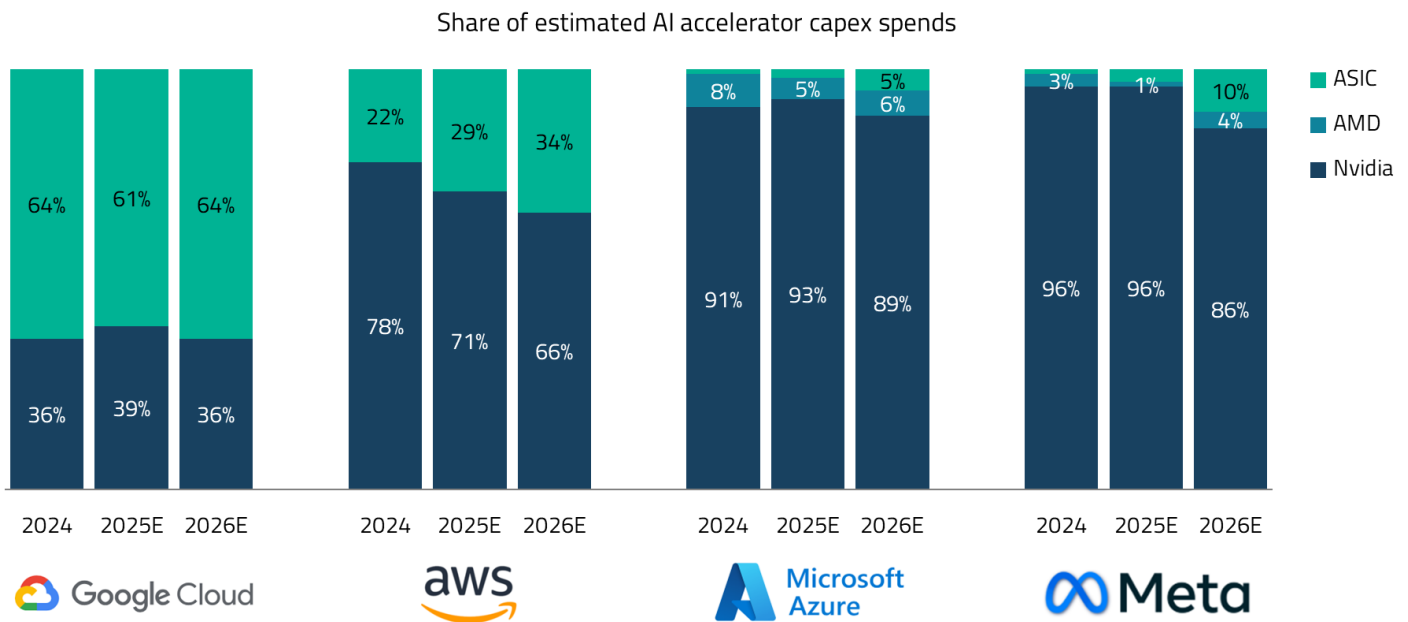
Source: Company data, UBS, September 2025

Custom AI accelerators

In addition to the merchant AI GPU market, where Nvidia and AMD have been the key players, there is also the custom AI GPU market – the ASICs.⁹ Google has been the biggest player in the space so far (with its TPU series of chips), while Amazon has also focused on custom AI GPUs (Trainium series, leveraging its Annapurna Labs). Both Google and Amazon made an early start to incorporating custom chips into their cloud data centers, from a time when annual capital expenditures on data centers were more modest compared to current levels. However, now with the top four hyperscalers forecast to spend USD 100 billion each in 2026 itself, and with a few others such as OpenAI on a similar (or even steeper) path, there is an added focus on custom AI chips – both from an economic angle as well as from a technological angle (custom workloads).¹⁰

Merchant chips from Nvidia account for almost 40% of total data center capex, and of that, almost three-quarters is Nvidia’s gross margin.¹¹ That leaves a lot of economic incentive for these capex spenders to develop in-house alternatives, particularly for certain workloads that are fine-tuned to internal requirements. It also likely helps with price negotiations and securing some supply. As the capex spend moves more towards inference compute from training compute, demand for such ASICs is likely to grow, and that is what we are seeing.

Exhibit 11: Estimated AI accelerator capex spends see in-house ASICs gaining share from merchant GPUs (Nvidia & AMD)



Source: Company data, Mizuho Securities, September 2025

In addition to the trend towards in-house chip manufacturing, there is a growing sub-trend within that ASIC strategy. Google’s ASIC strategy has so far largely been to work collaboratively with Broadcom to develop a custom chip. Google’s in-house team has been responsible for the overall architecture and specialized

⁹ ASIC (Application-Specific Integrated Circuit): Custom-designed chips built for a specific purpose, rather than general-purpose processors. In AI, these are chips designed specifically for machine learning workloads. Unlike merchant GPUs that are sold commercially off-the-shelf, ASICs are typically developed internally by companies like Google and Amazon for their own specific needs.

¹⁰ Source: Company reports, UBS, VisibleAlpha, August 2025

¹¹ Source: Shikhara Investment Management analysis, October 2025

components of the TPU (such as the core computing elements), and Broadcom has provided its advanced expertise to turn Google’s designs into physical products, including providing some critical technologies, such as high-speed serializer-deserializer (SerDes) interfaces. Amazon, on the other hand, has leveraged its in-house chip expertise (largely through its acquisition of Annapurna Labs several years ago) to do more of the frontend chip design work in-house and use external vendors for backend work. As the total pie of dollars spent on chips grows, there is a natural incentive to bring the higher-value-added frontend design work in-house. This is what it appears Google plans to do with its upcoming TPU v7 chips, where it is expected to proceed with two versions: the -7e will feature an in-house frontend design, while the -7p will continue with Broadcom.

We expect that as the total capex spend grows, more of the big AI capex spenders will likely start incorporating ASIC chips in their spend, starting off with full outsourcing and then, depending on the scale and expertise, eventually moving the frontend chip design in-house and only outsourcing the backend services. Such a move has specific implications for the Asian supply chain, as we discuss later in this paper.

Exhibit 12: Hyperscaler ASIC projects in 2026

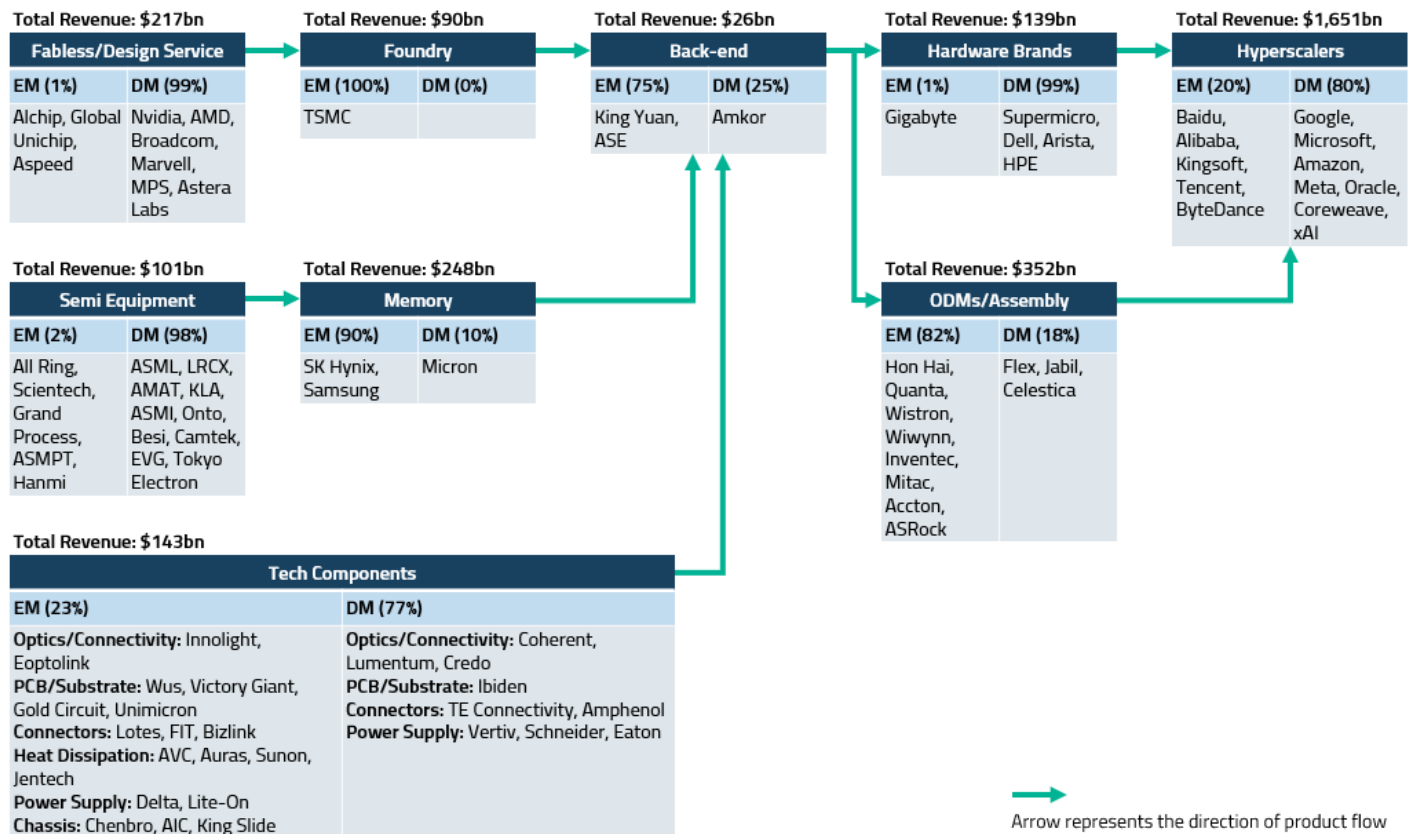
Company	Amazon						Alphabet					Microsoft			Meta	Intel/ Habana	Tesla	
High-performance computing ASICs	Graviton 3	Inferentia 1	Trainium 1/ Inferentia 2	Trainium 2/ Inferentia 3	Training 2.5/ Inferentia 3.5	Trainium 3/ Inferentia 4	TPU v5p/v6p	TPU v5e	TPU v6e (Trillium)	TPU v7	TPU v7e	Cobalt 100	Maia 100	Maia 200	MTIA3	Gaudi 3	Dojo	Dojo 2
Function	Data center CPU	AI inference	AI training & inference	AI training & inference	AI inference	AI training & inference	AI training & inference	AI training & inference	AI training & inference	AI training & inference	AI training & inference	Data center CPU	AI training & inference	AI training & inference	AI training & inference	AI training & inference	AI training & inference	AI training & inference
Production	2022	2021	2022	2024	2026	2026	2023	2023	2025	2026	2026	2024	2024	2025	2026	2024	2023	2026
Foundry	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC	TSMC
Process node	5nm	16nm	7nm	5nm	3nm	3nm	5nm	5nm	4nm	3nm	3nm	5nm	5nm	3nm	3nm	5nm	7nm	5nm
Foundry wafer price (US\$)	\$15,000	\$8,000	\$11,500	\$15,000	\$20,000	\$20,000	\$15,000	\$15,000	\$16,500	\$20,000	\$22,000	\$15,000	\$15,000	\$20,000	\$20,000	\$15,000	\$11,500	\$17,000
Total cost per die (US\$)	\$104	\$199	\$471	\$682	\$819	\$970	\$988	\$305	\$371	\$2,243	\$1,418	\$291	\$671	\$994	\$853	\$675	\$405	\$761
IC design margins	n/a	18%	13%	38%	38%	12%	60%	60%	60%	60%	30%	8%	8%	8%	60%	18%	18%	18%
IC design service/supply chain management partner	n/a	Alchip	Alchip	Marvell	Marvell	Alchip	Broadcom	Broadcom	Broadcom	Broadcom	MediaTek	GUC	GUC	GUC	Broadcom	Alchip	Alchip	GUC

Source: Company data, UBS, September 2025

Part 3: Global supply chain

The AI data center infrastructure buildout involves a range of companies across the technology supply chain and around the world. Using 2024 revenues as a base and considering only the listed names in the supply chain with reasonable exposure to this AI infrastructure spend leads to the accompanying chart (Exhibit 13). While some of the key innovations at the level of frontier models and chip design are happening in the US, the total value is spread across the world, with almost a third of the revenues being accounted for by listed technology companies in emerging markets (EMs), with a vast majority of those in Asia ex-Japan. In addition, China has been quite innovative across the AI chain – from foundational models to chip design and manufacturing, further adding to the EM weight for this theme.

Exhibit 13: Global AI supply chain for developed vs emerging markets in 2024 (US\$) – EM tech has 34% revenue weighting in the AI value chain

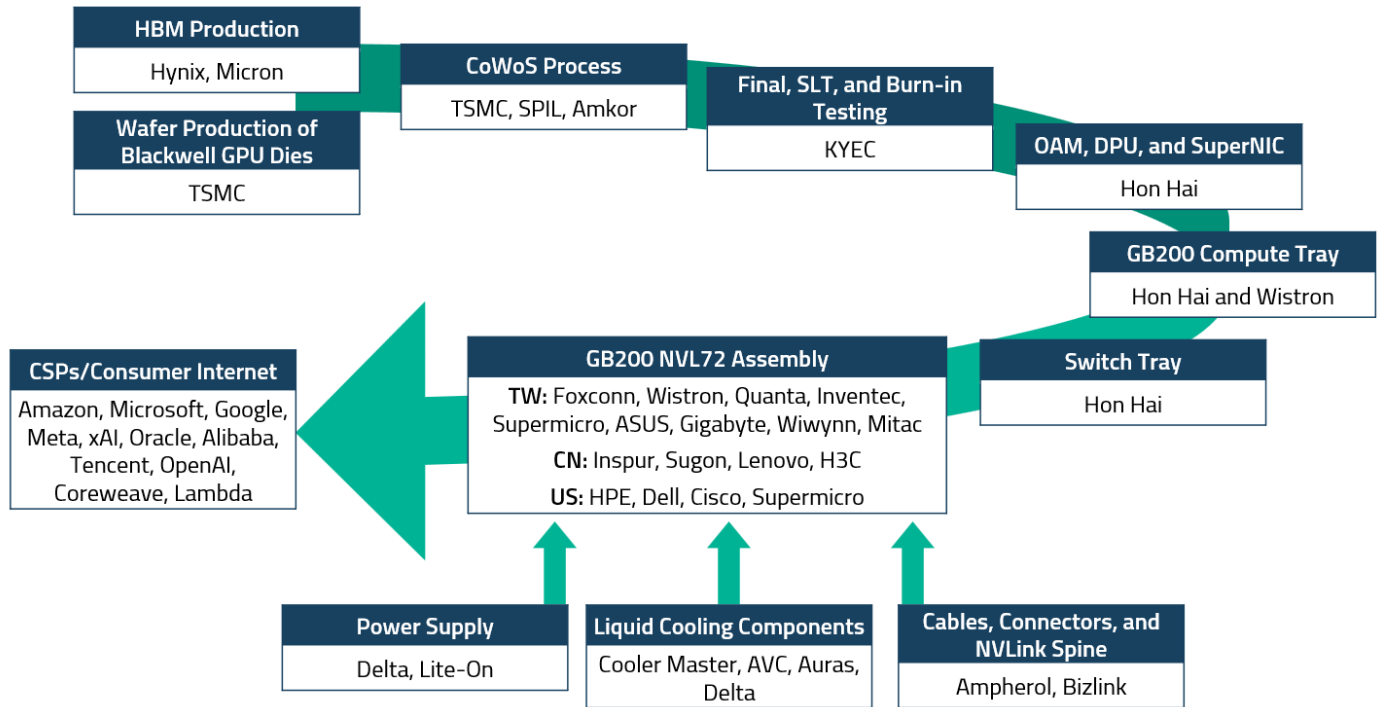


Source: Company data, UBS, September 2025

To illustrate how Asian companies fit in this chain, let us consider the supply chain for the current leading form of the AI data center hardware – Nvidia’s GB200 NVL72 rack. While the chips at the heart of the rack – the Blackwell GPU and the Grace CPU – are designed by Nvidia, they are fabricated at TSMC. The specialized memory accompanying the AI accelerator chips is made by SK Hynix (Samsung Electronics is in the process of qualification for the next generation). These chips then go through a packaging and testing process, largely dominated by Asian companies. A host of Asian companies are then involved in producing the boards and trays that house these chips, along with other key components, including passives, cables, and connectors. These boards and trays, along with power supplies and cooling components, are then assembled and connected into racks by key downstream vendors, where the Asian companies currently have about 90% share.¹²

¹² Source: Morgan Stanley, UBS, Shikhara Investment Management analysis, October 2025

Exhibit 14: An illustrative example of the Asia supply chain for a sample Nvidia GB200 rack



Acronyms: HBM (High Bandwidth Memory), GPU (Graphics Processing Unit), CoWoS (Chip on Wafer on Substrate), SLT (System Level Test), OAM (OCP Accelerator Module), OCP (Open Compute Project), DPU (Data Processing Unit), SuperNIC (Super Network Interface Card), NVLink (Nvidia Link), CSP (Cloud Service Providers). Source: Company data, UBS, September 2025

The buildout of new data centers requires spending on physical infrastructure (including buildings and power backups), storage, networking, and more. However, the biggest share of expenditure goes to the AI servers. While exact data isn't known, estimates and cross checks across companies in the supply chain and the ecosystems lead to an approximate split as follows: For every \$100 spent on the current generation of leading AI server rack (GB200 NVL72), approximately \$78-80 goes to Nvidia for its Blackwell chips, Grace CPUs, and NV Link switch.¹³ The remaining costs include: ~\$6-7 for major downstream content such as server boards and trays, ~\$3 for cooling components, and ~\$10-11 for various other semiconductors, cables, connectors, and assembly margins.¹⁴ Of the \$78-80 that goes to Nvidia, the company retains ~75% as its own gross margins, with the remaining ~25% covering production costs: 7-8% to TSMC for foundry and CoWoS packaging¹⁵, 8-10% to purchase HBM memory, another 2% to testing companies, and the rest on other items.¹⁶

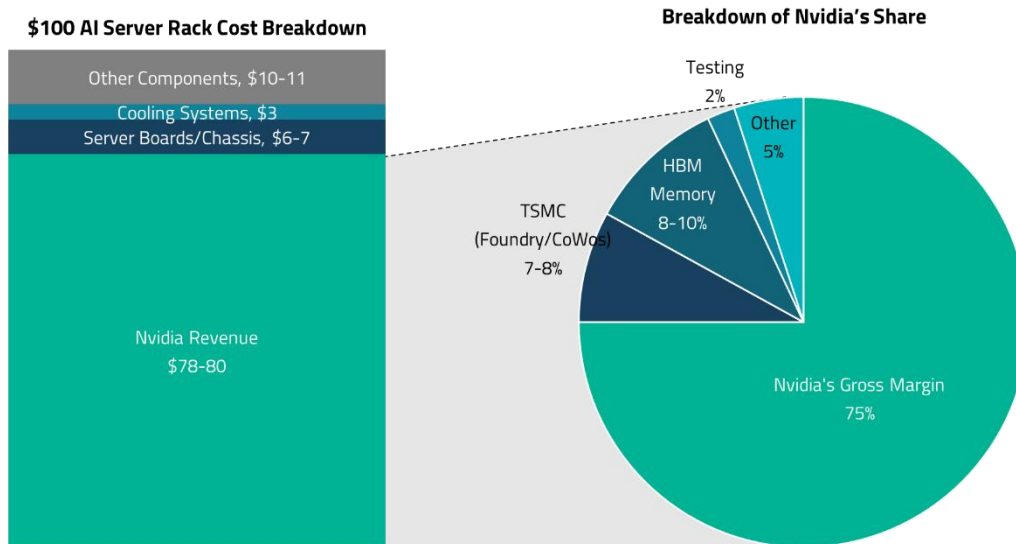
The above numbers are based on our estimates, speaking with companies, and going through various sell-side reports. It is worth noting that these estimates vary across different commentators and are therefore indicative in nature, although the actual numbers should be in the ballpark.

¹³ Source: Company data, Shikhara Investment Management analysis, September 2025

¹⁴ Source: Ibid.

¹⁵ CoWoS (Chip-on-Wafer-on-Substrate): An advanced packaging technology that integrates multiple chips (such as GPU cores and HBM memory) into a single package, enabling high-bandwidth connections and improved performance. TSMC is the leading provider of CoWoS packaging, which is essential for manufacturing advanced AI accelerators like Nvidia's latest chips.

¹⁶ Source: Company data, Shikhara Investment Management analysis, September 2025

Exhibit 15: Estimated breakdown of a GB200 NVL72 AI server rack cost – where does \$100 go?

Source: Company data, Shikhara Investment Management analysis, September 2025

Part 4: Our approach to investing in the Asian supply chain

As highlighted in the previous section, there is no dearth of Asian companies that are plays on this AI infrastructure rollout theme. However, given where we are in the cycle – both the rollout of AI as a theme and in stock markets – we take a multi-pronged approach to building our portfolio across what we see as two distinct categories: AI enablers and AI adopters. AI enablers are companies that provide the infrastructure, components, and services that make AI possible. AI adopters are companies that use AI to improve their own business operations and customer offerings.

Firstly, within AI enablers, we recognize that the biggest TAM/revenue pool currently comes from the Nvidia supply chain and that a lot of bullishness on the street across various Asian names is tied to that one driver. Hence, we want to avoid duplication and instead only pick a selective group of stocks on that one underlying driver that, in our analysis, provides the best risk-adjusted returns.

Secondly, also within AI enablers, the technology spending theme should broaden to include some non-Nvidia AI GPU expenditure, including ASICs, general purpose servers, and eventually front-end/edge devices. We want to have stocks that are leveraged to this expanding tech spending footprint in the coming years.

Thirdly, we focus on AI adopters – companies that are early beneficiaries of using AI applications. These range from Chinese and ASEAN internet platforms leveraging AI for better user engagement and operational efficiency to IT services companies helping enterprises integrate AI tools into their operations.

AI Enablers: Semiconductors

In semis, one of our high conviction ideas is TSMC. It is probably a high-conviction consensus buy for the street as well, and for good reasons. Almost all roads seem to lead to TSMC. As the current near-monopoly provider of

leading-edge semiconductor fabrication services and, importantly, advanced packaging solutions, it matters little whether tech spending growth is captured by Nvidia, AMD, or ASICs, or if the spending growth broadens out to non-GPU markets (outside of Intel CPUs). Within the current spending dominated by Nvidia solutions, other than Nvidia, TSMC is likely the company capturing the most dollars out of that TAM, both because it does higher value-added work and because it currently has a 100% share of Nvidia's GPU business.

Exhibit 16: Estimated dollar content captured by TSMC in every GB200 NVL72 rack

TSMC's GB200 Content	GB200	Interposer	CPUs	Connect X7	NV Link Switch	Bluefield 3	Spectrum 52-Por	Aspeed BMC
Tech Node	4nm	65nm	4nm	4nm	4nm	4nm	4nm	28nm
Transistors (bn)	208	n/a	n/a	8	50	22	100	n/a
Density (1mn trans/mm ²)	127.8	n/a	n/a	76.1	122.9	122.9	122.9	n/a
Die size (mm ²)	1628	2832	809	105	407	179	814	n/a
Wafer area (mm ²)	70,686	70,686	70,686	70,686	70,686	70,686	70,686	70,686
Usable area (Edge/size)	78%	50%	78%	82%	82%	85%	78%	90%
Usable area (mm ²)	55,347	35,343	55,135	57,962	57,609	60,351	55,347	63,547
Total Die per wafer	34.0	12.5	68.2	551.6	141.5	337.0	68.0	1323.9
Yield %	40%	95%	45%	80%	60%	80%	40%	95%
Good die per wafer	13.6	11.9	30.7	441.2	84.9	269.6	27.2	1257.7
Wafer price (US\$)	\$15,500	\$10,500	\$15,500	\$15,500	\$15,500	\$15,500	\$15,500	\$3,500
Cost per Die (US\$)	\$1,140	\$886	\$505	\$35	\$183	\$57	\$570	\$2.78
# of Die per GB200 System	72	72	36	72	18	18	2	63
TSMC Content per System	\$82,066	\$63,775	\$18,195	\$2,529	\$3,285	\$1,035	\$1,140	\$175
TSMC content per GB200 System	\$172,200							

Source: Company data, UBS, September 2025

Despite it being a consensus buy idea, we believe that the upside for the stock will potentially come from the continued upgrade in capex plans for AI infrastructure, as has been the case for the past two years (as noted earlier in this paper). In addition, if tech spending ex-AI infrastructure sees an upside in the coming years, then TSMC stands to capture a fair share of that spending too. While valuations for the stocks are richer than their own history, they are at a discount relative to some relevant global peers. As per FactSet, TSMC at ~19x P/E (24x on ADR pricing) is trading at about 10 turns discount to Nvidia and the Philadelphia Semiconductor index on 2026E earnings (both trading at ~30x); comparables such as Broadcom are at 39x 2026E EPS, and ASML is at 35x.¹⁷

The memory industry, which is globally dominated by Asian companies, has also undergone a structural shift for the better. In DRAM¹⁸, the industry is now significantly consolidated – effectively a three-player market outside of China – and demand is being driven aggressively by high-bandwidth memory (HBM) that is making the product more custom than commodity. With the reduction in the number of players, the industry has also become much more capital-expenditure disciplined. Consequently, margins and profitability for the industry have increased sharply, but so have valuations – particularly for SK Hynix and Micron.

¹⁷ Source: FactSet, data as of 28 October 2025

¹⁸ DRAM (Dynamic Random Access Memory): High-speed volatile memory used in computers and servers for temporary data storage. Essential for system performance but requires constant power to retain data.

Our preferred way to play the theme is now through Samsung Electronics, the traditional industry leader, which, for various reasons, has been lagging in the newer generation of HBM products for the past three years. We see multiple catalysts for the stock: attractive valuations, an expected earnings upgrade cycle over the next six months driven by strength in its traditional areas of dominance (traditional DRAM (DDR5, DDR7) and NAND¹⁹), it is controlling losses in its foundry business and may even see a major turnaround in this division if it executes well on its recent order wins from Tesla, and we are confident that it will be a supplier of HBM4 in 2026 (how much market share it ends up gaining remains a matter of debate). Importantly, despite Samsung's recent stock price run, the street remains divided in its opinion about the stock, creating an opportunity for investors willing to back the company's execution.

Staying with semiconductors, and as we argued in an earlier section, we remain bullish about the prospects of ASICs – custom-designed AI GPUs for hyperscalers and large users. Google and Amazon have been the two major users of ASICs, and supply chain checks and industry commentary suggest that other large spenders, such as Meta, Microsoft, and OpenAI, are also expanding their use of ASICs to complement and augment their spending on Nvidia chips. In the previous section, we also argued that we see the ASIC development trend evolving toward a co-development model with ASIC partners (the frontend in-house, the backend outsourced) rather than being a largely outsourced turnkey model.

This evolution opens the door for Asian semiconductor design houses such as Alchip, Global Unichip, and MediaTek, among others. Alchip has established itself as a leading backend design partner for a few years now, most notably for Amazon's Trainium ASIC chips, while MediaTek represents a new entrant to the field, with Google expected to commercialize a version of its TPU 7 chip in partnership with them next year.

We have liked MediaTek for its strong position in a consolidated market for smartphone processor chips, executing well to gain share of the premium segment of the market over the past two years, and its recent launch of the competitive Dimensity 9500 chip. If it executes well with Google TPU, then it opens up the TAM for a material new market that can add 15-20% revenues to an already large base, and importantly, add to the company's valuation. Comparable names such as Broadcom, Alchip, and even Marvell trade at substantially higher forward P/E multiples than MediaTek, which, as per FactSet consensus estimates, is currently trading at 17x 2026 P/E.²⁰

We have recently turned constructive on Alchip, which has been one of Asia's ASIC success stories. We expect it to rapidly ramp up the next version of Amazon's Trainium chip through 2026 and make progress in further expanding its footprint across marquee global clients and Chinese automakers (ADAS chips²¹).

AI Enablers: Hardware / downstream tech

While there is no shortage of companies in Asia in downstream/hardware tech that are suppliers to the Nvidia supply chain, we find that several of them are largely a beta play on the theme and hence difficult to differentiate in terms of investment rationale. We also find entry barriers in downstream tech to be lower than in upstream

¹⁹ NAND: A type of non-volatile flash memory used for data storage in devices like smartphones, SSDs, and data centers. Unlike DRAM, NAND retains data even when power is turned off, making it ideal for permanent storage of files, applications, and operating systems.

²⁰ Source: FactSet, data as of 28 October 2025

²¹ ADAS (Advanced Driver Assistance Systems) chips: Specialized automotive semiconductors that process sensor data to enable safety features like automatic braking and lane-keeping. It represents a growing market as automakers integrate more AI-powered driver assistance capabilities.

tech, thereby raising the risk of market share shifts and medium-term margin dilution. Moreover, several of these stocks are already richly valued.

Quanta Computers is our preferred pick here. Quanta derives its revenues from notebook assembly (it is the global number one in notebook ODM) and data center servers. Traditionally, it provided only general-purpose servers for cloud data centers, but has now evolved to be one of the top two global providers of Nvidia servers and racks (the other one being Hon Hai/Foxconn Group), and is the leading provider of ASIC-based servers to hyperscalers. For various reasons, the stock has lagged behind the broader AI supply chain rally and, until last month, traded at levels it had in July 2023, when it first started shipping its initial Nvidia servers based on Hopper chips. The consensus estimate for Quanta's 2025 earnings was under NT\$10 in 1H2023, and that got lifted to NT\$14 by August of 2023 as the street started to understand the full impact of Nvidia's server business.²² However, the estimate for 2025E now stands at NT\$18 with a two-year (FY27) EPS estimate of NT\$24.²³ We believe that as the company accelerates its production of GB300-based racks in Q4 2025, it will capture the market's attention, some of which we have already begun to see in recent weeks. The company also provides a decent 4.3% dividend yield, so it is partly paying for patience. Additionally, if there is an uptick in the PC replacement cycle driven by an aging installed base and/or the need for more AI-edge computing, Quanta's notebook ODM business is likely to benefit.

AI Adopters: AI users and beneficiaries

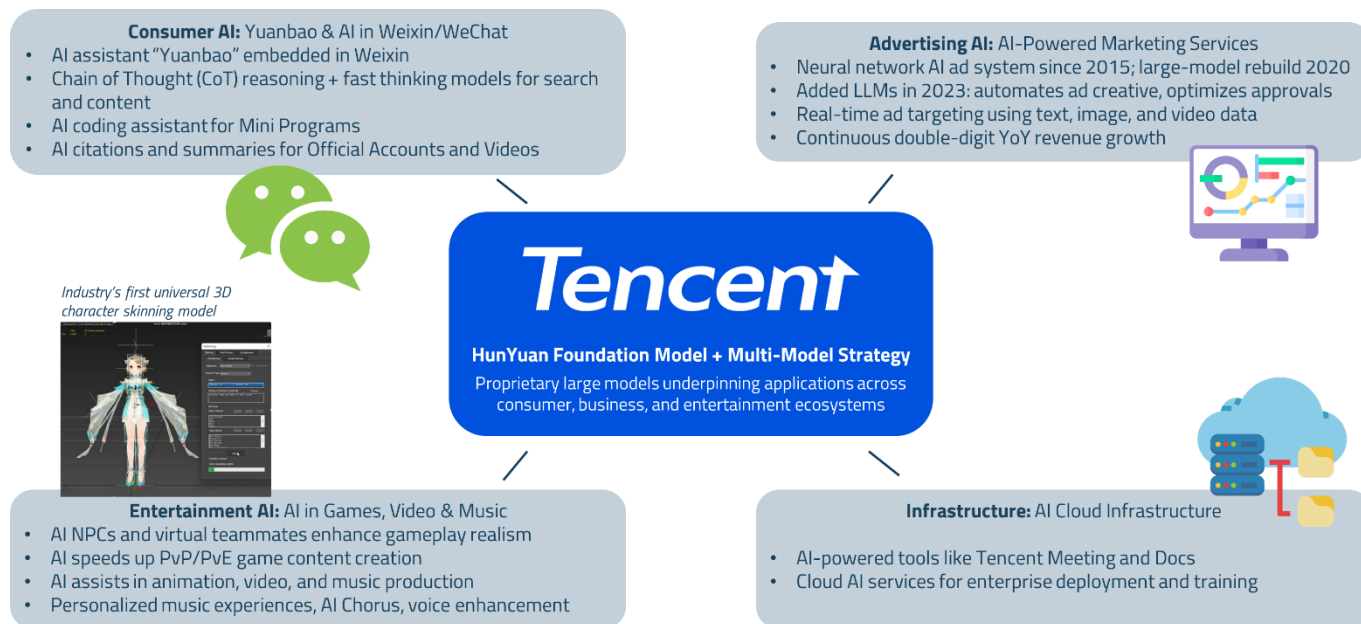
At this stage, not surprisingly, some of the biggest beneficiaries of the progress in AI computing have been the 'internal' users rather than commercially available applications. Software companies have used it extensively to reduce the basic cost of coding, and consumer tech companies have utilized it to better target their customers, whether in terms of delivering more relevant content or advertising.

In Asia, some of the biggest early adopters and beneficiaries of AI-led technologies are the Chinese companies. Alibaba has been a leader in building out the Chinese AI cloud infrastructure and has also launched one of the leading foundational/large language models, Qwen. Alibaba has leveraged AI to make its platforms, such as Taobao and Tmall, more intuitive and efficient, driving higher engagement and sales conversion rates. It has also applied AI learnings to optimize its vast logistics operations. Tencent has been another heavy investor in AI usage. According to the company, AI is now integrated into nearly 70% of its internal operations, helping create better products and driving up user engagement (both gaming and video accounts), and thus driving revenues.²⁴

²² Source: FactSet, October 2025

²³ Source: Ibid.

²⁴ Source: Tencent, September 2025

Exhibit 17: Tencent – leveraging AI

Source: Tencent, Shikhara Investment Management analysis, September 2025

An interesting name that we’ve been monitoring for a while is Tencent Music (TME). While our initial position in the stock was driven by its dominant market position in a very consolidated and closed market (effectively a two-player market in China) that, in our opinion, was ripe for a significant average selling price (ASP) and margin-led growth over the next five years, the introduction of AI-led tools has accelerated that process. The company is now better placed in suggesting content to its users, is better able to monetize its non-subscribing customers via targeted advertising and, equally importantly and less well understood by the market, is able to drive down costs.

Outside of China, major platforms in Southeast Asia, such as Sea and Grab, are also effectively utilizing AI tools, further augmenting our original investment thesis in these names. Sea, in particular, has deployed a comprehensive AI strategy across its three main business segments through partnerships with OpenAI, Google, and WIZ.AI, combined with internal research from its Sea AI Lab. In e-commerce, Shopee utilizes AI agents for autonomous, personalized search and customer support, while augmented reality features, such as Beauty Cam, enhance user engagement for beauty and personal care products. Sea’s gaming vertical, Garena, leverages AI to segment lapsed gaming players and deliver tailored incentives, improving retention and reducing customer acquisition costs. Most impressively, Monee, Sea’s fintech business, utilizes generative AI agents to handle multilingual customer outreach for financial services, achieving 40-50% higher activation rates and helping to grow the user base from 1 million to 15 million users in 2023.²⁵

Elsewhere, Indian IT services names are an interesting area of debate – there is a deflationary impact of generative AI tools (coding becomes cheaper, so those benefits need to be passed on to the customers, deflating revenues), but there is a significant revenue opportunity on the horizon too. Enterprises are likely to be the major users of AI tools and applications, and integrating these tools with business transformation needs for the AI age will inevitably require software and technical manpower, which is where Indian IT companies excel. Our base case

²⁵ Source: WIZ.AI, accessed August 2025

view is that, eventually, the upside opportunity is much higher than the deflationary risks; however, in the near term, the outlook remains hazy. Stock valuations for most firms don't seem to fully discount that near-term uncertainty, despite the past year's underperformance.

Cognizant stands out in that respect. In our opinion, it is a turnaround story – after a few years of mis-execution by the previous management, the new management is executing well to reverse the damage, and that is now showing up in numbers with 2025E and 2026E estimated EPS being revised up by the street over the past six months. Meanwhile, estimates have been revised down or have been flat for all its major peers over the same period. For now, Cognizant still trades at a substantial discount on forward P/E multiples vis-à-vis its peers as well as its own previous history.

Part 5: Risks and how we navigate them

We believe that AI-led transformation will be one of the defining features of the coming few decades and that it will impact almost all aspects of economic activity. However, as in almost all cases of technological transformation, the process is not linear; it will not be without pauses, and identifying winners and losers with precision will be difficult. There will also be over-investments in capacity from time to time, and, inevitably, there will be unreasonable valuations for some stocks or sectors. Investing, therefore, requires significant diligence and flexibility; investment theses will need to evolve and adapt to the incoming data and real-world experience.

Over the next two years, from a fundamentals perspective, the biggest realistic risk to this theme we see comes from a 'pause' – that while the three-to-five year capex spending CAGR holds, we get, say, a 12-month period of subdued capex as spenders pause to digest the significant built-in capacity or they are forced to pause given constraints elsewhere. Such constraints are currently part of the discussion framework and relate to the physical infrastructure build-out, including power generation infrastructure, as well as the shortage of trained technical manpower in the US. Our approach to mitigating this risk is to be selective in our stock picks and avoid duplicating the same underlying revenue driver across stocks. We are valuation-sensitive and continue to expand our portfolio by acquiring beneficiaries of AI usage, rather than just holding beneficiaries of AI capital expenditures.

There is also a 'macro' risk – if the US were to get into a recession in the coming year, the big spenders on AI capex are one way or the other linked to consumer spending and hence may choose to flatten out their capex for a year or so, while retaining the high absolute levels. While this is an ever-present risk, our working hypothesis is that if the economic impact of a recession is relatively modest on the hyperscalers, their behavior regarding capex may not change much, as most of these companies seem to consider this AI-wave as an existential threat – they would rather sacrifice near-term free cash flows than put the longer-term survival at risk. However, as we said, this is a working hypothesis, and we are willing to modify it based on incoming information.

The other risk is the usual 'cyclical' risk – that eventually there will be oversupply in parts of the supply chain or an influx of new competitors in other parts, thereby suppressing margins. We consider this risk as a regular part of any investment thesis, and we look to navigate it accordingly.

While these risks are real and require careful navigation, we believe the structural nature of the AI transformation and our selective approach position us well to capitalize on what we see as one of the most significant investment opportunities of our generation.

Disclaimer

For sophisticated investors only. For informational purposes only. The information presented in the material is not and may not be relied on in any manner as legal, tax, investment, accounting or other advice or as an offer to sell or a solicitation of an offer to buy an interest in any investment product or any other entity sponsored or managed by Shikharā Investment Management. This material doesn't constitute and should not be considered as any form of financial opinion or recommendation.

Shikharā Investment Management LP ("Shikharā") is currently an Exempt Reporting Adviser that is exempt from registration as an investment adviser with the U.S. Securities and Exchange Commission. This material does not constitute an offer to sell or the solicitation of an offer to buy in any state of the United States or other U.S. or non-U.S. jurisdiction to any person to whom it is unlawful to make such offer or solicitation in such state or jurisdiction.

Investment involves risk. Past performance is not indicative of future performance. It cannot be guaranteed that the performance of the investment product will generate a return and there may be circumstances where no return is generated. Investors could lose all or a substantial portion of any investment made. Before making any investment decision, investors should read the Prospectus for details and the risk factors. Investors should ensure they fully understand the risks associated with the investment product and should also consider their own investment objective and risk tolerance level. Investors are advised to seek independent professional advice before making any investment.

Shikharā's investment products are suitable only for sophisticated investors and require the financial ability and willingness to accept the high risks and lack of liquidity inherent in Shikharā's investment products. Prospective investors must be prepared to bear such risks for an indefinite period of time. No assurance can be given that the investment objectives of any given investment product will be achieved or that investors will receive a return of their investment.

Certain of the information contained in this material are statements of future expectations and other forward-looking statements. Views, opinions and estimates may change without notice and are based on a number of assumptions which may or may not eventuate or prove to be accurate. Actual results, performance or events may differ materially from those in such statements.

Certain information contained in this material is compiled from third-party sources. The information and any opinions contained in this document have been obtained from sources that Shikharā considers reliable, but Shikharā does not represent such information and opinions are accurate or complete, and thus should not be relied upon as such. Furthermore, all opinions are current only as of the date of distribution are subject to change without notice. Shikharā does not have any obligation to provide revised opinions in the event of changed circumstances. Whereas Shikharā has, to the best of its endeavor, ensured that such, information is accurate, complete and up-to-date, and has taken care in accurately reproducing the information, Shikharā takes no responsibility for the accidental publication of incorrect information, nor for investment decisions taken based on this material. Neither Shikharā nor any of its affiliates makes any representation or warranty, express or implied, as to the accuracy or completeness of the information contained herein, and nothing contained herein should be relied upon as a promise or representation as to past or future performance of any investment product or any other entity.

The contents of this material are prepared and maintained by Shikharā and has not been reviewed by the Securities and Exchange Commission of the United States.