AI Hardware Supply Chain: A Structural Map and Decision-Making Framework

Industry Architecture and Tracking Variables

Report Period: July 2026 · For Professional Investors

Published by

DH Wealth Management Pte Ltd

Singapore CMS Licence: CMS101242

Trader Insights | June 2026: AI Resilience Holds as Market Leadership Evolves

In June 2026, markets tested the resilience of the AI theme. Semiconductors experienced a mid-month pullback of nearly 10 per cent, driven by two key pressures: first, the SpaceX IPO absorbed a portion of risk capital, market attention and trading liquidity; second, investors began re-evaluating hyperscaler AI capital expenditure trajectories – whether this wave of CapEx can be sustained, and whether it will ultimately generate sufficient returns through AI revenue, efficiency gains, and commercialisation. Capital rotated into less crowded areas including financials, industrials, healthcare and small caps. Towards month-end, Micron's earnings revalidated AI memory demand and pricing power, and semiconductors staged a strong recovery in the final two trading days, with SMH ultimately gaining 9.5 per cent. This reinforced the most important market signal for June: AI remains the core theme, but markets are now demanding greater scrutiny on valuations, position crowding, and CapEx return profiles.

Index performance reflected this structural divergence. SPY fell 1.3 per cent, QQQ fell 0.3 per cent, while RSP rose 1.9 per cent and IWM gained 3.5 per cent. Equal-weighted indices and small caps outperformed, suggesting some broadening of market participation. However, the timing suggests this rotation was more about a tactical rebalancing from crowded AI positions than a decisive leadership change. What remains significant is that semiconductors, after a sharp pullback, were able to recover and finish the month higher – indicating that conviction in AI infrastructure demand remains intact.

At the sector level, industrials and healthcare led, with XLI gaining 7.0 per cent and XLV 6.2 per cent. Communication services and energy lagged, with XLC down 7.4 per cent and XLE off 5.7 per cent. Oil prices were a key macro variable in June, with USO falling 17.5 per cent, weighing on the energy sector while simultaneously alleviating inflation tail risks, providing a more favourable trading backdrop for industrials, financials, small caps and select cyclical consumer names. This environment allowed markets to look for catch-up trades beyond AI, but capital ultimately returned to the AI infrastructure value chain – the area with the strongest earnings validation and most visible demand.

Micron's earnings were a pivotal moment this month. It delivered a clear signal: AI demand is materialising through memory pricing, order volumes and revenue. For semiconductors, this matters more than thematic narratives, because it moves AI capital expenditure from "expectation" to "earnings delivery". But the same signal raises a larger question for the next phase: if upstream suppliers are already capturing stronger profitability, the massive CapEx commitments from hyperscalers will ultimately need to be validated by real applications and commercial returns. Going forward, market focus will gradually shift from "who benefits from the build-out" to "whether these investments generate adequate returns".

This is also why AI application layers – particularly physical AI, robotics and embodied intelligence – are attracting more attention. This direction could become an important extension of the AI theme, as it directly addresses the future usage scenarios and commercialisation pathways for AI infrastructure. However, the space remains in the early stages of theme formation and validation, and should not be equated with a mature earnings cycle. Markets will continue to differentiate companies with genuine orders, revenue visibility and earnings upgrades, while more rapidly penalising those driven primarily by narrative expansion.

Looking ahead to July, the near-term market environment appears favourable, particularly in the first half. New capital allocations, passive inflows, retirement account contributions and systematic rebalancing typically provide support at month start. The index approaching previous highs and VIX returning to low levels suggest markets are preparing for further upside. We see a continued opportunity for modest upside in early July, but

this window should not be extrapolated too far. Seasonals typically weaken from late July; August and September tend to see lower liquidity and higher volatility, with less stable return distributions. Accordingly, a constructive but disciplined approach appears appropriate: AI remains the central theme, diversification is a positive, and the most attractive opportunities lie where earnings, fund flows and positioning are mutually reinforcing.

About DH Global Investment:

A diversified, multi-asset fund designed to deliver robust, diversified returns through investment across a broad range of global asset classes. The fund's portfolio includes listed fixed income and equity securities, foreign exchange and private credit, combining traditional and alternative investment channels to capture growth and stability across global markets. The fund employs dynamic asset allocation, adjusting to market conditions and investment opportunities, to maximise returns while managing risk.

Executive Summary | Five-Minute Overview

- **I. The AI Industry Foundation** — The AI hardware supply chain can be distilled into four structural layers: compute chips, memory and advanced packaging, interconnect and thermal management, and systems and software. Each layer exhibits distinct supply-demand dynamics, technical barriers and capital intensity, driving different profit pools and investment characteristics.
- **II. Sources of Moat among Leading Players** — Incumbents across the value chain have established differentiated competitive advantages: architectural lock-in (Nvidia), process monopoly (TSMC), capacity pre-emption (SK hynix), ecosystem lock-in (Broadcom), and system integration (Vertiv). The nature of each moat informs its respective valuation framework.
- **III. The Evolution of Key Supply Constraints** — Capacity bottlenecks (HBM/CoWoS) are expected to ease from late 2026 into 2027; technological bottlenecks (advanced nodes) remain a structural constraint; ecosystem bottlenecks (software/network) are the most enduring to resolve; and geopolitical bottlenecks are the most uncertain.
- **IV. Capital Expenditure Flows and Structure** — Hyperscalers, sovereign AI initiatives and enterprise AI are the three key drivers of global AI infrastructure investment. In June 2026, South Korea announced a combined public-private investment programme exceeding KRW 1,461 trillion (approximately US$948 billion), adding a significant new variable to the global CapEx picture.
- **V. Profit Structures and Trends across Segments** — Gross margins, operating margins and ROIC show meaningful dispersion across the value chain. Hardware layers (GPU/HBM) currently command the highest margins, but the scope for further expansion is narrowing; equipment/materials are in early margin expansion; systems/software offer the highest long-term margin potential, albeit with longer payback periods.
- **VI. Contrasting Industrial Pathways** — The US model is driven by technological innovation and global ecosystems; the China model by application deployment and import substitution. The two pathways differ systematically in their drivers, constraints, and valuation logics.
- **VII. A Decision-Making Framework at Elevated Valuations** — Judgement at current levels depends on three key variables: CapEx growth trajectory, the pace of supply additions, and valuation levels. A framework-based approach should replace directional forecasting.

Introduction

The June 2026 half-year report cross-validated public commentary from Intel and Nvidia CEOs, pointing to a structural shift in AI hardware investment logic from a "GPU-centric" to a "CPU+GPU collaborative" paradigm. This month's report builds on that foundation, constructing a durable AI hardware supply chain map that translates industry structure into a framework suitable for ongoing tracking.

This report does not constitute investment advice, nor does it offer any individual security recommendations. Its purpose is to provide a structural framework to assist readers in systematically understanding the current configuration and evolving dynamics of the AI hardware supply chain.

The report unfolds across the following dimensions: the four-layer foundation of the AI industry and their interdependencies; the sources of moat among dominant players and their respective valuation logics; the evolution of key supply constraints; global capital expenditure flows and structure; segment-level profit structures and trends; a comparative framework of US and China industry pathways; and a decision-making framework for elevated valuation environments.

Together, these dimensions form a complete chain from industry foundation to decision variables, allowing readers to scrutinise their own assumptions against market consensus.

Chapter 1: The AI Industry Foundation

1.1 From Nine Layers to Four: Abstraction of the Industry Foundation

This month's report introduces a nine-layer map to provide a detailed structural view of the AI supply chain. The value of an anatomical map lies in diagnosing individual layers; however, to build a useful investment framework, these nine layers must be abstracted into four higher-order structural foundations.

Industry Foundation	Segments Covered	Core Characteristics
Compute Chips	GPU/CPU Die	Source of performance and power; highest market attention
Memory & Advanced Packaging	HBM, CoWoS, Substrates/PCB	Bandwidth and integration bottleneck; lowest supply elasticity
Interconnect & Thermal	Network chips, optical comms, liquid cooling, power	Physical constraint on scaling; penetration rates rising rapidly
Systems & Software	Server/rack assembly, software platforms	Ultimate value capture layer; deepest moat but longest payback

> **Structural logic of the four foundations:** The further upstream (compute chips), the higher the technical barriers and the lower the supply elasticity – but market consensus is also higher and the scope for surprise narrower. The further downstream (systems and software), the more fragmented the technical barriers and the greater the supply flexibility – but market consensus is lower and the potential for change larger. This pattern recurs throughout the report and directly informs valuation logic and timing.

1.2 Supply Constraint Characteristics of the Four Foundations

Each foundation exhibits materially different supply elasticity. Some layers require multi-year capacity expansion cycles, while others can scale more rapidly. These differences determine each layer's pricing power and earnings resilience in the face of changing demand conditions.

Industry Foundation	Current Supply Status	Supply Elasticity	Capacity Expansion Cycle
Compute Chips	Tight, capacity fully committed	Low	12-18 months
Memory & Advanced Packaging	Persistently tight, capacity booked through 2027	Very low	18-24 months
Interconnect & Thermal	Demand growing rapidly, supply scaling	Medium	6-12 months
Systems & Software	Demand visible, but monetisation takes longer	High	Depends on system integration capability

These differences in supply elasticity imply that when demand growth moderates, the layers with the lowest elasticity (memory and advanced packaging) can sustain pricing power for longer, while those with higher

elasticity (systems and software) are more dependent on continued demand expansion to support profitability.

Chapter 2: Sources of Moat among Leading Players

2.1 Competitive Landscape across Segments

Across the four industry foundations, certain players have established significant competitive advantages. Understanding these advantages requires two assessments: first, whether the player controls irreplaceable core assets (technology, capacity, customer relationships); and second, whether these assets exhibit network effects or scale economies that strengthen over time.

Player	Segment	Core Source of Competitive Advantage	Moat Type
Nvidia	Compute Chips	CUDA ecosystem locks in developers; product roadmap leadership	Architectural lock-in + ecosystem
TSMC	Advanced Process + Packaging	Process leadership + CoWoS capacity; high customer dependency	Process monopoly + capacity pre-emption
SK hynix	HBM Memory	HBM technology leadership; capacity booked long-term	Technology first-mover + capacity lock
Broadcom	Network Chips	Comprehensive interconnect silicon ecosystem; hyperscaler custom partner	Ecosystem lock-in + customisation
Vertiv	Liquid Cooling / Data Centre Infrastructure	Integrated solution capability; direct beneficiary of AI rack penetration	System integration + first-mover deployment

The moat types differ significantly – software ecosystems, process technology, capacity lock-in, and system integration each represent distinct competitive barriers. The importance of understanding these differences lies in their implications for resilience through industry cycles and the corresponding valuation frameworks. Ecosystem-based moats (Nvidia) have the highest switching costs, while capacity-first moats (SK hynix) are more dependent on sustained supply-demand tightness.

2.2 Moat Type and Valuation Logic

Moat Type	Representative Player	Core Valuation Variables	Source of Valuation Elasticity
Architectural lock-in + ecosystem	Nvidia	Software revenue mix, customer switching costs	Ecosystem expansion rate
Process monopoly + capacity pre-emption	TSMC	Advanced process revenue mix, capacity utilisation	Process leadership gap
Technology first-mover + capacity lock	SK hynix	HBM revenue mix, DRAM cycle position	Technology gap maintenance
Ecosystem lock-in + customisation	Broadcom	AI custom chip revenue, interconnect share	Hyperscaler self-design trend
System integration + first-mover deployment	Vertiv	Liquid cooling penetration, system value per rack	Rack power density growth

Understanding the relationship between moat type and valuation logic helps avoid comparing companies with different business models using a single framework. Nvidia's valuation reflects its expanding ecosystem value, while SK hynix's valuation reflects pricing power during a period of sustained supply tightness – their respective valuation criteria should differ accordingly.

Chapter 3: The Evolution of Key Supply Constraints

3.1 Four Types of Supply Constraints

This chapter shifts the focus from structure to dynamics: what are the key constraints currently limiting AI hardware supply chain expansion, and at what pace are these constraints expected to evolve? Understanding this evolution is a prerequisite for assessing earnings inflection points across segments.

Constraint Type	Specific Manifestation	Current Status	Expected Evolution
Capacity	HBM, CoWoS, advanced nodes	Fully committed, demand exceeds supply	Gradual easing from H2 2026 into 2027
Technological	Node scaling, HBM stacking, CPO	Continuous evolution	Long-term constraint with periodic breakthroughs
Ecosystem	Software ecosystems, network standards	Nvidia CUDA dominant	Longest resolution horizon: 3-5 years
Geopolitical	Export controls, supply chain security	Most uncertain	Most unpredictable evolution path

The evolution of these four constraint types differs materially. Capacity constraints offer a relatively clear path and can be assessed against known expansion plans. Technology and ecosystem constraints are longer-duration, more suitable as tracking variables. Geopolitical constraints are the most uncertain and should be treated as scenario variables rather than baseline assumptions.

3.2 Investment Implications of Constraint Evolution

- **Capacity constraints (HBM/CoWoS):** Expected to ease gradually from H2 2026 into 2027. Pricing power may shift from "strongly supplier-favoured" towards "more balanced" as new supply comes on stream.
- **Technological constraints (advanced nodes):** A structural feature. Technology leaders' advantages will persist, but the gap may narrow over a 2-3 year horizon.
- **Ecosystem constraints (software/network):** The longest resolution horizon. Nvidia's CUDA ecosystem and network standards remain the strongest moats, with no clear near-term alternatives.
- **Geopolitical constraints (export controls):** Most uncertain. Supply chain regionalisation will continue to reshape the geographic distribution of the AI supply chain.

> **Phasing of constraint easing:** Before capacity constraints ease, the most consensus-driven segments can continue to be held for tracking. After capacity eases, the structural shift in pricing power will affect earnings visibility across segments. The interconnect and thermal layer's penetration-driven growth is less exposed to capacity constraints and offers a longer time horizon.

Chapter 4: Capital Expenditure Flows and Structure

4.1 Three Key Drivers

According to projections from Goldman Sachs, IDC and others, cumulative global AI infrastructure-related capital expenditure for 2026-2030 is estimated at US$7.6 trillion, driven by three key forces:

Source of Capital	Characteristics	2026-2027 Trend
Hyperscalers	Largest scale, most sustainable	Combined AI capex expected near US$600bn
Sovereign AI	Strategically driven, not purely economic	Multiple nations have announced programmes
Enterprise AI	Dispersed, longer decision cycles	Transitioning from exploration to deployment

Hyperscaler investment, driven by returns on capital and competitive dynamics, is the most stable source of demand for the AI supply chain, though changes in its growth rate have the most direct impact. Sovereign AI investment, more influenced by geopolitical strategy, is more volatile but has strong local impetus. Enterprise AI investment is more dispersed, with its penetration trajectory a key marginal variable over the next 1-2 years.

4.2 South Korea's "Three Mega Projects": A New Variable

On 29 June 2026, South Korean President Lee Jae-myung announced a national-level "Three Mega Projects" programme covering semiconductors, physical AI and AI data centres, with combined public-private investment exceeding KRW 1,461 trillion (approximately US$948 billion).

Investment Area	Investment Size	Core Content
Semiconductor Manufacturing Base	KRW 800 tn	Second semiconductor cluster in south-west Korea, 4 memory fabs
HBM Packaging Base	KRW 81 tn	HBM packaging cluster in Chungcheong region
AI Data Centres	KRW 550 tn	SK Group, GS Group, Naver; initial 8.4GW capacity cluster
Next-Generation Semiconductor R&D	KRW 30 tn	AI chips, physical AI, defence semiconductors

> **Structural implications of the Korean programme:** By defining semiconductors, physical AI and AI data centres as "three axes", the programme signals that Korea is elevating physical AI (robotics) and data centre infrastructure to the same strategic level as chip manufacturing. This provides longer-term regional demand support for interconnect and thermal management, server assembly, and related segments.

4.3 The Multi-polar Trend in Global CapEx

The Korean announcement reinforces the multi-polar trend in global AI capital expenditure. US hyperscalers are expected to spend close to US$600 billion on AI capex in 2026 alone, while Korea has launched a long-

term framework exceeding US$900 billion. Japan, the EU, Singapore and other economies have also announced sovereign AI investment programmes.

This regional dispersion of capital expenditure carries several implications: the equipment and materials supply chain benefits from a broader demand base, extending the cycle for second-order beneficiaries; supply chain regionalisation accelerates, with each economy favouring local suppliers; and differences in regional spending cadence may reduce the impact of a slowdown in any single region.

Chapter 5: Profit Structures and Trends across Segments

5.1 Current Profitability Levels

Based on the latest company earnings data (as of Q1 2026), gross margins and operating margins show meaningful dispersion across the value chain. Understanding the structural reasons for this dispersion is essential for assessing segment-level valuation.

Segment	Representative Player	Gross Margin (approx.)	Operating Margin (approx.)	Margin Phase
Compute Chips	Nvidia (Q1 FY2026)	~71%	~56%	Stable at high levels; limited further expansion
HBM Memory	SK hynix (Q1 2026)	~45-50%	~35-40%	Scope for expansion with HBM4 upgrade
Advanced Packaging	TSMC (Q1 2026)	~66%	~50%	Advanced process + packaging mix driving margins
Interconnect & Thermal	Delta (Q1 2026), Vertiv (Q1 2026)	~30-37%	~10-15%	Expanding with penetration growth
Systems & Software	SMCI (Q3 FY2026) etc.	~10-15%	~5-8%	System value increase driving margin improvement

The core driver of margin dispersion is position within the industry foundation. Compute chips and memory/packaging capture the highest margins, but the scope for further improvement is constrained by the eventual easing of supply tightness. Interconnect and thermal margins are expanding, driven by technology upgrades and ASP increases. Systems and software margins are currently lower but have the greatest long-term upside.

5.2 Medium-Term Margin Outlook

- **Compute chips and memory/packaging:** Margins are likely to remain high, but the pace of improvement will moderate. As supply increases, pricing power may gradually normalise.
- **Interconnect and thermal:** Margins are expanding. As AI racks transition from "optional" to "standard", the ASP uplift story is still unfolding.
- **Systems and software:** Margin improvement will take longer. Product cycles and customer qualification timelines are extended, but the direction is clear.

Segments with margins still expanding (interconnect and thermal) offer greater margin of error in a high-valuation environment, while those with margins near peak require more disciplined valuation assessment.

Data sources: Company earnings reports (Nvidia Q1 FY2026 GAAP; SK hynix Q1 2026; TSMC Q1 2026; Delta Q1 2026; Vertiv Q1 2026; SMCI Q3 FY2026 GAAP). Valuation data are consensus estimates.

Chapter 6: Contrasting Industrial Pathways

6.1 The Importance of Comparison

The US and China are the two largest AI markets globally, but their industry structures and development trajectories differ fundamentally. Understanding these differences is important to avoid applying a single valuation framework across markets. US AI companies are valued against a global addressable market, reflecting expectations of global AI spending; Chinese AI companies are valued against the import substitution opportunity, reflecting "maximum feasible share under constrained conditions".

6.2 Key Differences

Dimension	US Pathway	China Pathway
Core Driver	Technology innovation + global ecosystem	Application deployment + import substitution
Profit Source	IP, system integration, software subscriptions	Scale, scenario adaptation, policy support
Key Constraints	Advanced node capacity, power, geopolitics	Advanced nodes, EDA tools, equipment supply
Valuation Logic	Global pricing, technology premium	Import substitution premium, policy-driven
Time Horizon	2-3 year product roadmaps	5-10 year import substitution cycle
Geopolitical Factor	Export control uncertainty	Supply chain disruption risk, technology封锁

6.3 Investment Implications of the Differences

- **Different valuation anchors:** US AI companies' valuation ceilings depend on global market penetration; Chinese AI companies' valuation ceilings depend on import substitution progress. The former is driven by global CapEx, the latter by policy and technology milestones.
- **Geopolitical impacts may operate in opposite directions:** When export controls tighten, the global addressable market (TAM) for US AI companies may contract, while the import substitution opportunity for Chinese AI companies may expand. The same event can have opposite valuation effects.
- **Different time horizons:** US AI companies are tracked on product roadmaps (2-3 years); Chinese AI companies are tracked on import substitution milestones (5-10 years). The former focuses on quarterly data, the latter on policy and technology milestones.

> **Framework application note:** Investors should avoid applying US AI valuation logic directly to Chinese AI companies, and vice versa. Identifying whether a company's pricing anchor is the global market or import substitution is the starting point for valuation assessment.

Chapter 7: A Decision-Making Framework at Elevated Valuations

7.1 Three Dimensions for Assessing Valuation Levels

Dimension	Current Observation	Signals to Watch
CapEx growth rate	Double-digit growth expected for 2026; Korea adds incremental support	If growth slows from 30% to 10%, earnings expectations should be re-assessed
Supply release pace	HBM/CoWoS capacity expected to increase from H2 2026	If supply comes online earlier than expected, pricing power may normalise sooner
Valuation levels	Tier 1 P/E in upper half of historical range	If valuations already discount 2-3 years of earnings growth, margin for error is thin

These three dimensions are interconnected: CapEx growth defines the demand ceiling, supply release defines the supply pressure, and valuation levels define how much expectation is already reflected in price. When all three point in the same direction, the margin for error narrows significantly.

> **Current reference points:** Tier 1 (compute chips/memory packaging) reflects high consensus; Tier 2 (interconnect and thermal) remains in the cognitive formation phase; Tier 3 (systems and software) requires a longer time horizon.

7.2 Framework: Scenario-Based Decision-Making

Scenario	Trigger	Framework Reference
CapEx continues to expand	Capex maintains double-digit growth	Full chain benefits, but elasticity depends on expectation gaps across segments
CapEx growth moderates	Capex growth slows to single digits	Tier 1 faces valuation compression; Tier 2 penetration story still provides support
Geopolitical tensions rise	Export controls tighten	Global supply chain uncertainty rises; import substitution chains gain attention

The value of scenario planning lies not in predicting which scenario will occur, but in clarifying in advance: "If a given scenario occurs, how should I respond?"

7.3 Three Sustainable Tracking Themes

- **Capacity release pace:** HBM/CoWoS expansion progress is the key variable for assessing whether Tier 1 valuations are reasonable.
- **Penetration inflection points:** The penetration curves for liquid cooling, 800V DC power and 1.6T optical modules are core indicators for Tier 2 earnings inflection points.
- **Geopolitical evolution:** Changes in export control policy will re-set valuation boundaries for AI companies across regions.

> **The enduring value of the framework:** The analytical chain presented here—from industry foundation to decision variables—can be carried forward across future monthly reports. As supply chain dynamics evolve, the

> data and judgments within the framework can be refreshed, building a cumulative research asset.

Risk Warnings

- AI sector valuations are at historically elevated levels; any earnings disappointment could trigger significant drawdowns.
- Geopolitical risks (US-China relations, Middle East tensions) could have a more pronounced impact on global asset pricing than currently anticipated.
- Supply chain disruptions, technology pathway shifts, and export controls could affect the earnings outlook for specific segments.
- The execution of large-scale capital expenditure programmes (including South Korea's recently announced US$900 billion-plus programme) faces implementation uncertainty, including infrastructure, labour availability, and power supply constraints.

Appendix: Data Sources

- Company earnings reports (Nvidia Q1 FY2026 GAAP; SK hynix Q1 2026; TSMC Q1 2026; Delta Q1 2026; Vertiv Q1 2026; SMCI Q3 FY2026 GAAP)
- Goldman Sachs, IDC, Counterpoint, Reuters, Bloomberg
- South Korean government "Three Mega Projects" official announcement (29 June 2026)
- Company investor presentations and earnings call transcripts

This report is for use **within Singapore** and under applicable cross-border wealth management frameworks.

8. Enquiries and Complaints
Email: compliance@dhwealth.sg

DH Wealth Management Pte Ltd | Singapore CMS Licence: CMS101242
