

## RESEARCH REPORT

## The AI Infrastructure Supercycle: Where the Money Goes, and What Comes Back

Nearly \$900 billion of cloud capex in 2026, a ~\$765B AI spend split that leaves power at just 5%, OpenAI at a \$40B run-rate against a ~\$14B projected loss — and a record \$11.6B of data centre investment into Asia Pacific. Chainfir Research reads the AI infrastructure supercycle through a profit-first lens.

Industry Reports · September 18, 2026 · Chainfir Research

### The largest private capital cycle in history

The numbers have moved beyond historical comparison. TrendForce estimates that the combined 2026 capital expenditure of the world's nine largest cloud service providers — Google, Amazon, Meta, Microsoft, Oracle, ByteDance, Tencent, Alibaba and Baidu — will exceed US\$886.7 billion, up roughly 90% year over year, before approaching US\$1.3 trillion in 2027. Morgan Stanley's estimates for the big-four US hyperscalers alone point to roughly US\$779 billion in 2026, up about 83%, with Alphabet and Meta both guiding to spend close to double their 2025 levels.

The long arc is even more striking. Deloitte's TMT Predictions 2026 puts global AI data center capex at US\$400–450 billion in 2026 — more than half of it chips — rising to around US\$1 trillion by 2028. PwC's Global Data Centre Outlook projects, in its baseline scenario, more than US\$31.6 trillion of cumulative capital spending on AI data centers between 2026 and 2050, of which US\$8.2 trillion lands in Asia Pacific. Goldman Sachs frames its own 2026–2031 estimate of roughly US\$7.6 trillion as about twelve times the inflation-adjusted cost of the entire US Interstate Highway System.

**Big-four hyperscaler capex, 2026E (US\$B, YoY: +65% / +109% / +61% / +101%)**

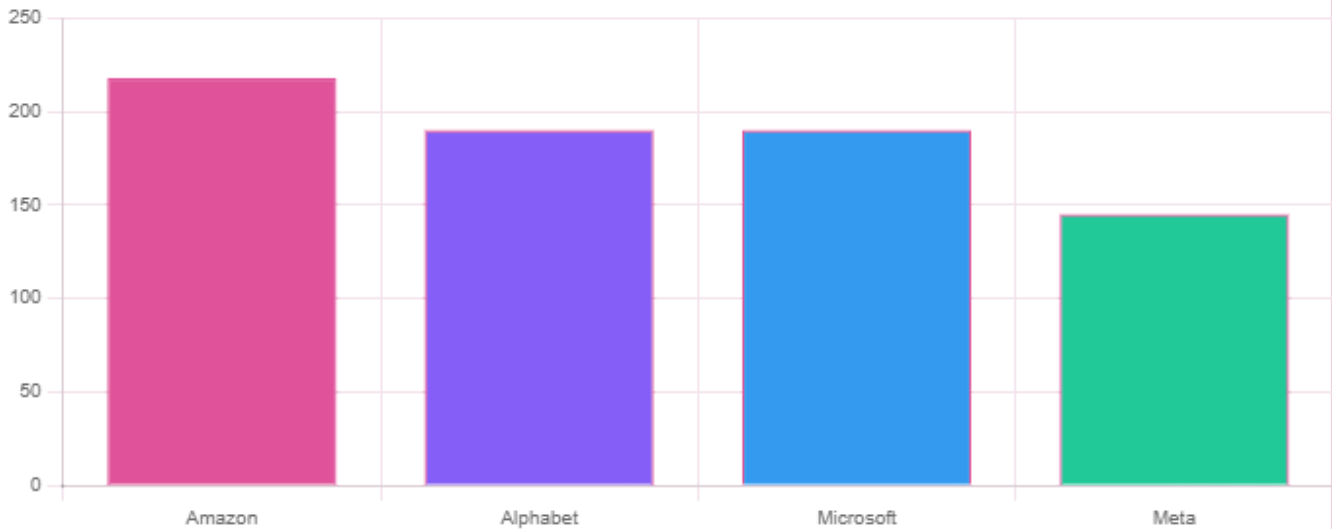


Chart 1: Big-four US hyperscaler capital expenditure estimates for 2026 (US\$ billions), with year-over-year growth in the chart title.

Source: Morgan Stanley estimates, September 2026 (cited via industry research); scope is total company capex.

## Where the money goes: chips first, power last

Goldman Sachs' baseline breaks the roughly US\$765 billion of 2026 AI capital expenditure into three layers: US\$494 billion of compute (chips and servers), US\$232 billion of data center shells, and just US\$39 billion of power infrastructure. Two-thirds of the money buys silicon that fully depreciates within four to six years.

The asymmetry matters. Power is the cheapest line on the capex table and the hardest problem in the room. JLL puts data center construction costs at US\$11.3 million per megawatt in 2026, up 47% from 2020; high-voltage transformers now carry 36-month lead times — roughly triple the norm of three years ago — while GPU racks require 18-to-24-month advance orders. The industry is discovering that the scarce asset is the megawatt, not the model.

Global AI capex split, 2026 (~\$765B total)

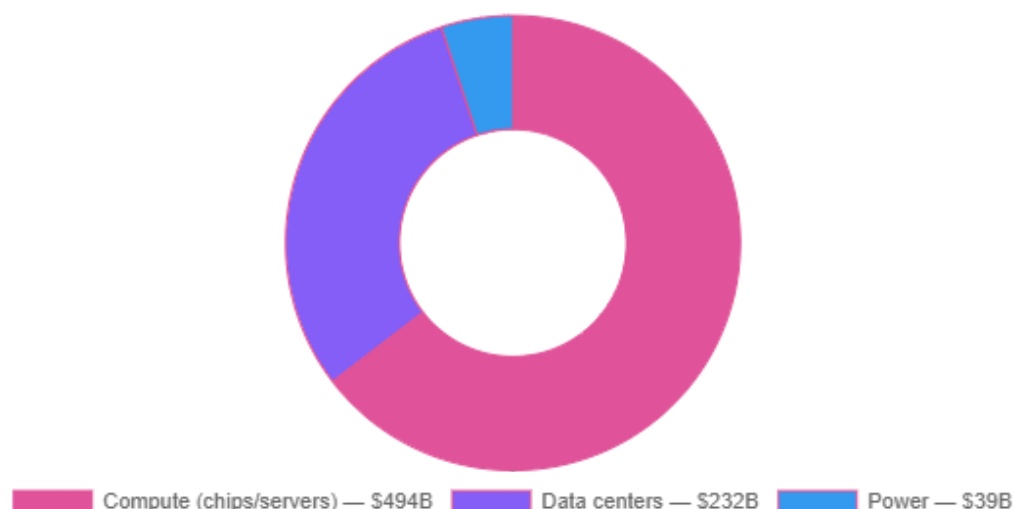


Chart 2: Goldman Sachs baseline decomposition of 2026 global AI capital expenditure (~US\$765B). Source: Goldman Sachs Global Institute / Global Investment Research, March 2026.

## The revenue gap nobody can wave away

On the other side of the ledger sits a much smaller number. UBS estimates that Amazon, Alphabet and Microsoft will spend about 102% of their cloud revenue on capital expenditure in 2026 — recycling essentially all of it, and then some, back into infrastructure. Sequoia's David Cahn has upgraded his famous "AI's \$200B Question" into a "\$1.5 trillion question".

The pure-plays are compounding impressively from small bases. OpenAI's annualized revenue run rate reached US\$40 billion by August 2026 — yet its leaked audited financials showed US\$13.07 billion of 2025 revenue against a US\$20.9 billion operating loss, with roughly US\$14 billion of losses projected for 2026. Anthropic's run rate reportedly climbed from about US\$9 billion at end-2025 to more than US\$30 billion by April 2026. Impressive, yes. Sufficient to amortize a trillion-dollar asset base? Not yet: US\$40 billion of run rate equals roughly 6% of projected 2026 hyperscaler capex.

Our reading is that this is not evidence that AI is useless — McKinsey estimates generative AI's steady-state economic value at US\$2.6–4.4 trillion per year — but that most of that value currently settles with the users of AI rather than its vendors. The gap is a value-capture problem, and value-capture problems are solved by business models, not by more GPUs.

Model-company revenue run-rates (US\$B)

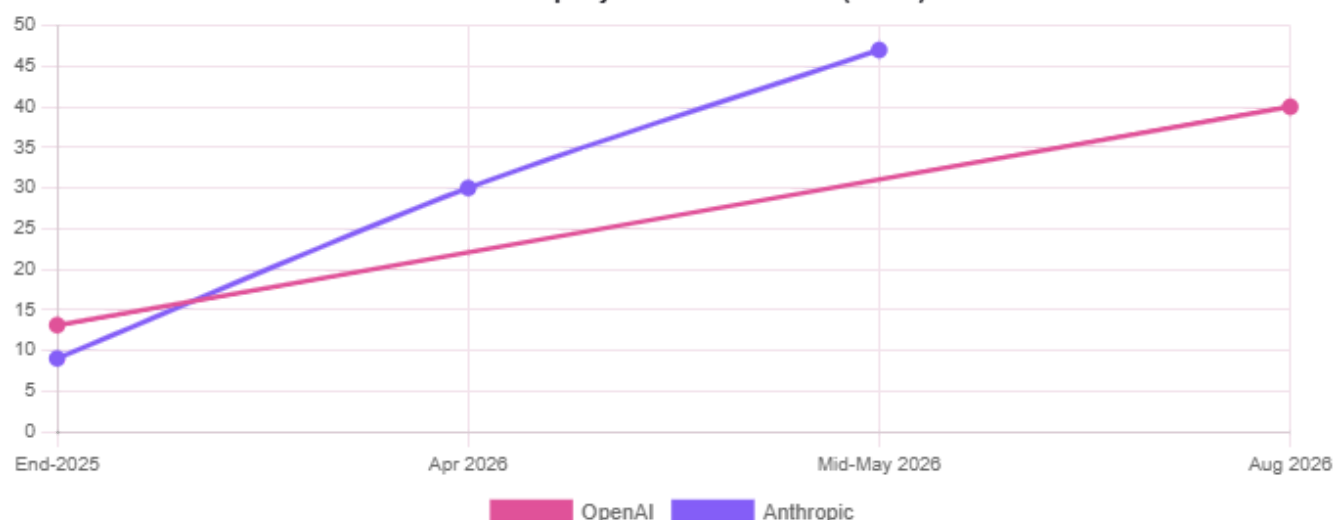


Chart 3: OpenAI end-2025 figure is audited booked revenue (US\$13.07B); Aug-2026 figure is annualized run-rate (~US\$40B). Anthropic figures are reported annualized run-rates (~\$9B → \$30B+ → \$47B+). Source: leaked audited financials and company disclosures, 2026 (compiled via NextFin, VentureBeat).

## Why we still take the buildout seriously

None of this reads to us as pure mania. IDC measured worldwide generative AI spending at US\$644 billion in 2025, up 76.4% year over year, even as per-token prices fell roughly a thousand-fold — a textbook Jevons effect. Agentic workloads consume on the order of a thousand times more tokens than chat for equivalent tasks, and usage, not training, is becoming the dominant cost driver.

On the infrastructure side, CBRE and Goldman Sachs data show AI still accounted for less than 15% of total data center workloads in 2025, with a projected rise to 29% by 2030 — a four-year demand CAGR of 98% since 2023. Supply, meanwhile, is tightening: across Asia Pacific, data center availability fell 43% year over year in Q1 2026, and regional vacancy sits at just 7%. Whatever the financing risks, the physical-constraint story is real.

The geography of demand is also broadening rather than concentrating. Deloitte's analysis points to compute shifting from centralized pre-training toward distributed inference — metro-adjacent inference sites, enterprise edge deployments and sovereign AI programs in which countries insist on domestically operated capacity. That widens the addressable map well beyond the handful of US mega-campuses that dominate headlines, and it is precisely the layer where Asia Pacific's licensing regimes and power surpluses become competitive advantages rather than afterthoughts.

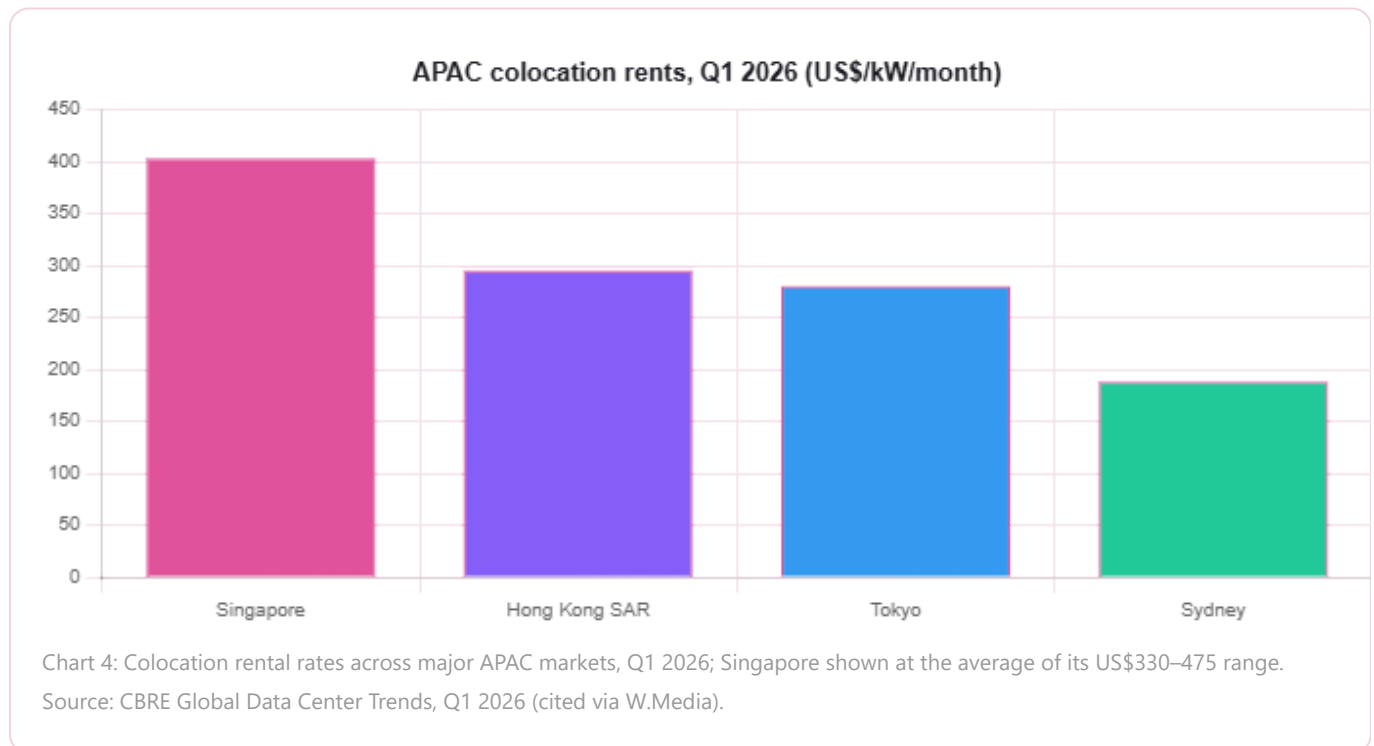
## APAC: power-rich markets are winning

The regional pattern mirrors the global one: capital follows megawatts. Direct data center investment in Asia Pacific hit a record US\$11.6 billion in 2025, according to CBRE. Johor led live-capacity growth at 53% year over year, followed by Melbourne above 30%; Greater Seoul, Sydney, Mumbai and Tokyo grew 15–25% from much

larger bases. Singapore and Hong Kong grew more slowly — constrained by policy and power allocation, not by demand.

The price signals are unambiguous. Singapore colocation rents average US\$403 per kW per month — among the world's highest — on just 2% vacancy. Hong Kong's vacancy fell from 28% to 18% within a year, rents rose to US\$295, and the city is repositioning as a high-density hub for AI inference and financial workloads, with the Sandy Ridge cluster due online in 2029. Mainland demand is compounding too: ByteDance has lifted its 2026 AI capex above RMB 200 billion (about US\$30 billion), and Alibaba's reached RMB 126.1 billion (US\$18.3 billion) in its latest fiscal year.

A new demand cohort is reinforcing the trend. CBRE highlights the rise of neocloud operators — GPU-as-a-service specialists from CoreWeave and Lambda globally to Yotta, E2E Networks and Neysa AI in India and Sustainable Metal Cloud in Singapore — while national frameworks such as Australia's National AI Plan and Korea's "AI for All" scheme underpin sovereign AI demand. The report is candid about the caveat: neoclouds carry shorter track records and weaker credit profiles, which is exactly why landlord covenants, contracted pre-leasing and power security — not headline gigawatts — are the underwriting that matters.



## The profit-first reading

We apply a single filter to this cycle — the same one we apply to every asset we hold:

- **Contracted revenue beats hoped-for utilization.** Tier-one US markets are effectively pre-leased through 2028; in APAC we look for the same contracted discipline before a single dollar of ours lands.
- **The moat is the megawatt, not the model.** Power access, grid interconnection and land are assets competitors cannot fork; chips depreciate, substations compound.

- **Cash discipline over narrative.** We pass on compute priced below cost and demand stimulated by credits — subsidized utilization is marketing, not revenue.
- **Buy the shovels with structural margin.** Cooling, power equipment and contracted colocation sit upstream of the hype and downstream of physics.

*"Scale is not an argument. We ask one question: can these machines keep selling compute at a positive margin next year, in five years, in eight? The last cycle rewarded storytellers; this one will reward operators with cash discipline." — Brian Tian, Founder & CEO, Chainfir Capital*

## Our stance

Chainfir Capital invests from its own balance sheet — roughly 90% of our deployed capital is proprietary — and we hold every portfolio company to the standard we hold ourselves to: at least eight consecutive quarters of positive operating contribution before growth capital is committed. In AI infrastructure, our attention is on Asia Pacific — Hong Kong, Singapore and New Zealand — where power-advantaged assets, contracted colocation and the picks-and-shovels layer offer durable economics.

The supercycle is real. So is the discipline it will demand. We deploy capital, talent and resources as a single allocation, backed by a 374+ advisor network — and we intend to be on the side of the trade that clears cash flow, not the side that clears headlines.